Evaluation of the MIND Research Institute's Spatial-Temporal Math (ST Math) Program in California

Submitted to:

Andrew Coulson
MIND Research Institute
111 Academy, Suite 100
Irvine, CA 92617

Submitted by:

Staci Wendt
John Rice
Jonathan Nakamoto

WestEd
4665 Lampson Avenue
Los Alamitos, CA 90720

Date: October 2014

WestEd

Table of Contents

List of Exhibits

Executive Summary

The MIND Research Institute contracted with the Evaluation Research Program at WestEd to conduct an independent assessment of mathematics outcomes in elementary school grades across California that were provided with the ST Math program. The outcomes examined were grade-level California Standards Test (CST) scale scores in mathematics as well as the proportions of students who were proficient or advanced in mathematics based on their CST scores. These outcomes were examined one year after grades were first provided with the ST Math program.

The unit of analysis for the evaluation was "grade" as opposed to classroom or school. A "grade" included all the classes in a school that taught content for a specific grade level. For example, the data from an elementary school with three grade 4 classes were included in the evaluation as a single "grade." Data from 463 grades ranging from 2 through 5 and provided with ST Math were included in the evaluation (found in columns 1 and 2 of Exhibit ES1). These grades were nested in 212 schools.

In addition, outcomes were examined only for grades that fully implemented the program in the first year. For the purposes of this study, full implementation was considered to have occurred for a grade when, at a particular school, at least 85 percent of the students enrolled in the grade had logged into ST Math during the academic year, and where at least 50 percent of the grade-level material in ST Math was covered by those students. Of the 463 grades in the evaluation that were provided with ST Math, 209 met the criteria for the full implementation analysis (found in columns 3 and 4 of Exhibit ES1). These grades were nested in 129 schools.

Exhibit ES1. Number of Grades Provided with ST Math Included in the Evaluation and Number of Grades Fully Implementing ST Math, by Grade Level

		Grades Fully Implementing ST Math	
Grade	Grades Provided with ST Math	Number of Grades	As a % of Grades Provided with ST Math
2	108	45	41.67
3	120	63	52.50
4	116	44	37.93
5	119	57	47.90
Total	463	209	45.14

Exhibit reads: For grade 2, 108 grades were provided with ST Math. Of these 108 grades, 45 (or 41.67 percent) fully implemented the program.

The evaluation utilized a quasi-experimental design that compared outcomes for grades that were provided with ST Math with outcomes for matched grades that were not provided with ST Math. Outcomes were compared separately for each grade level using analysis of covariance (ANCOVA) in order to account for differences in several school characteristics as

well as differences in grade-level mathematics performance prior to the provision of ST Math. In addition, in order to account for the nesting of grades within schools, differences in outcomes between the two groups were examined across all grade levels using hierarchical linear modeling.

RESULTS FOR GRADES PROVIDED WITH ST MATH

Students in second grades that were provided with the ST Math program had significantly higher mean mathematics scale scores on the CST compared to the CST scores of students in matched second grades that were not provided with the program. In addition, second grades that were provided with ST Math had a significantly larger proportion of students at the advanced level in mathematics on the CST compared to second grades that were not provided with the program. Also, second grades that were provided with the program had a significantly larger proportion of students at the proficient or advanced level in mathematics on the CST compared to second grades that were not provided with the program. The difference for each outcome was statistically significant after correcting for the examination of multiple outcomes. No statistically significant differences were found in any other grades.

For each of the three outcomes, the pooled difference across grades was statistically significant. Specifically, students in grades that were provided with the ST Math program had significantly higher mean mathematics scale scores compared to those of students in grades that were not provided with the program. The difference had an effect size of 0.16 (i.e., a difference of 0.16 of a standard deviation). For example, if the average comparison grade's scale score were at the 50th percentile, an effect size of 0.16 would mean that the average ST Math grade's scale score would be at the 56th percentile, for a difference of 6 percentile points.

In addition, when pooling differences across grade levels, grades that were provided with ST Math had a significantly larger proportion of students who scored at the advanced level in mathematics on the CST (35.01 percent) compared to grades that were not provided with the program (32.54 percent). The effect size for this difference was 0.17. Also, grades that were provided with ST Math had a significantly larger proportion of students that scored at either the proficient or advanced level of mathematics on the CST (64.88 percent) compared to grades that were not provided with the program (62.58 percent). The effect size for this difference was 0.16. All the differences across grade levels were statistically significant after correcting for the examination of multiple outcomes (Exhibit ES2).

Exhibit ES2. Grades Provided with ST Math and Comparison Grades: Adjusted Average Percentages of Students Advanced, or Proficient or Advanced, on the Mathematics CST

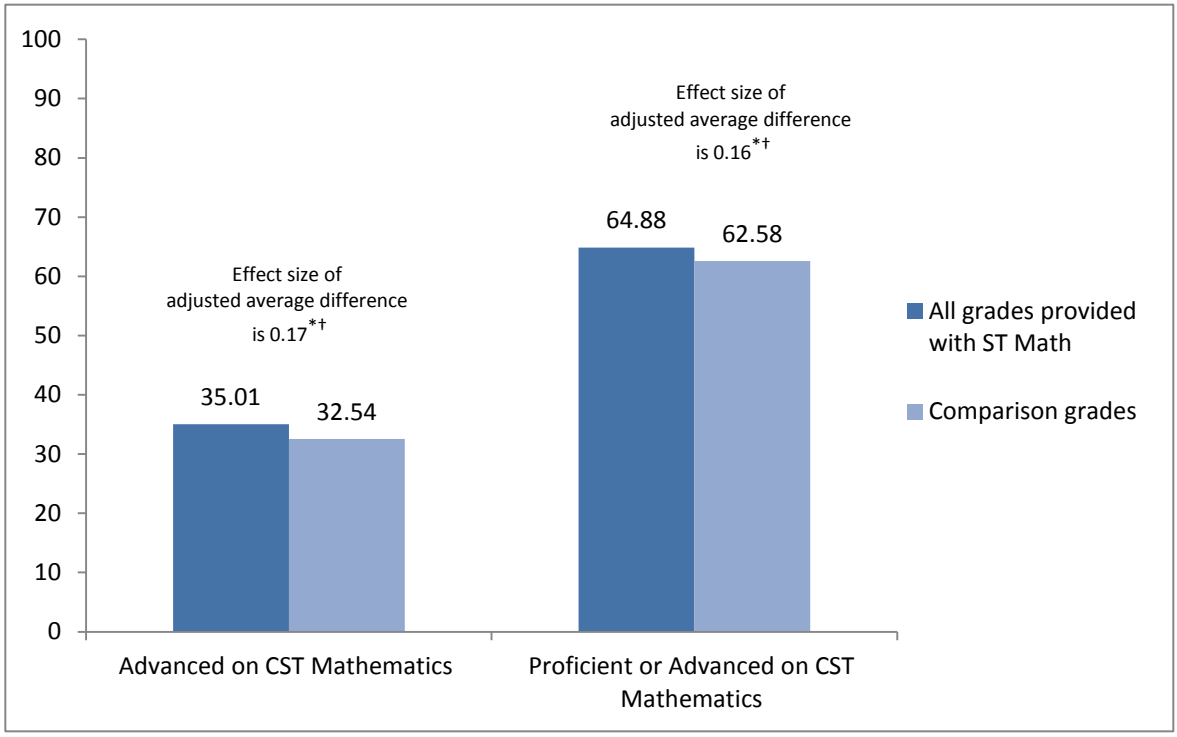


Exhibit reads: 35.01 percent of students in grades that were provided with ST Math were advanced on the mathematics CST compared to 32.54 percent of students in grades that were not provided with the program.
* statistically significant at *p*-value < .05, two-tailed test
† statistically significant at < BH critical value, correcting for the false discovery rate for multiple comparison tests
Note: All outcomes adjusted for grade-level 2010 percentage of students proficient or advanced, and for school-level percentage of Latino, Native American, and African American students, number of students enrolled, and number of students eligible for free or reduced-price lunch. Hierarchical linear modeling was used to account for the nesting of grades (n = 926) within schools (n = 544).

RESULTS FOR GRADES THAT FULLY IMPLEMENTED ST MATH

Students in grades 2, 3, and 5 that fully implemented the ST Math program had significantly higher mean mathematics scale scores on the CST compared to the CST scores of students in matched grades that were not provided with the program. In addition, a significantly higher proportion of students in grades 2, 3, and 5 that fully implemented the ST Math program scored at the advanced level in mathematics on the CST, and a significantly greater proportion of these students scored at either the proficient or advanced level in mathematics on the CST than students in comparable grades that were not provided with the program. The differences were statistically significant after correcting for the examination of multiple outcomes. No statistically significant differences were found on any of the outcomes for grade 4.

When pooling differences across grade levels, students in grades that fully implemented the ST Math program had significantly higher mean mathematics scale scores compared to those of students in grades that were not provided with the program. The effect size of this

difference was 0.42. For example, if the average comparison grade's scale score were at the 50th percentile, an effect size of 0.42 would mean that the average ST Math grade's scale score would be at the 66th percentile, for a difference of 16 percentile points.

In addition, when averaging differences across grade levels, grades that fully implemented ST Math had a significantly larger proportion of students who scored at the advanced level in mathematics on the CST (37.15 percent) compared to grades that were not provided with the program (31.57 percent). The effect size for this difference was 0.40. Also, grades that fully implemented ST Math had a significantly larger proportion of students at either the proficient or advanced level in mathematics on the CST (67.86 percent) compared to grades that were not provided with the program (61.54 percent). The effect size for this difference was 0.47. All the differences across grade levels were statistically significant after correcting for the examination of multiple outcomes (Exhibit ES3).

Exhibit ES3. Grades that Fully Implemented ST Math and Comparison Grades: Adjusted Average Percentages of Students Advanced, or Proficient or Advanced, on the Mathematics CST

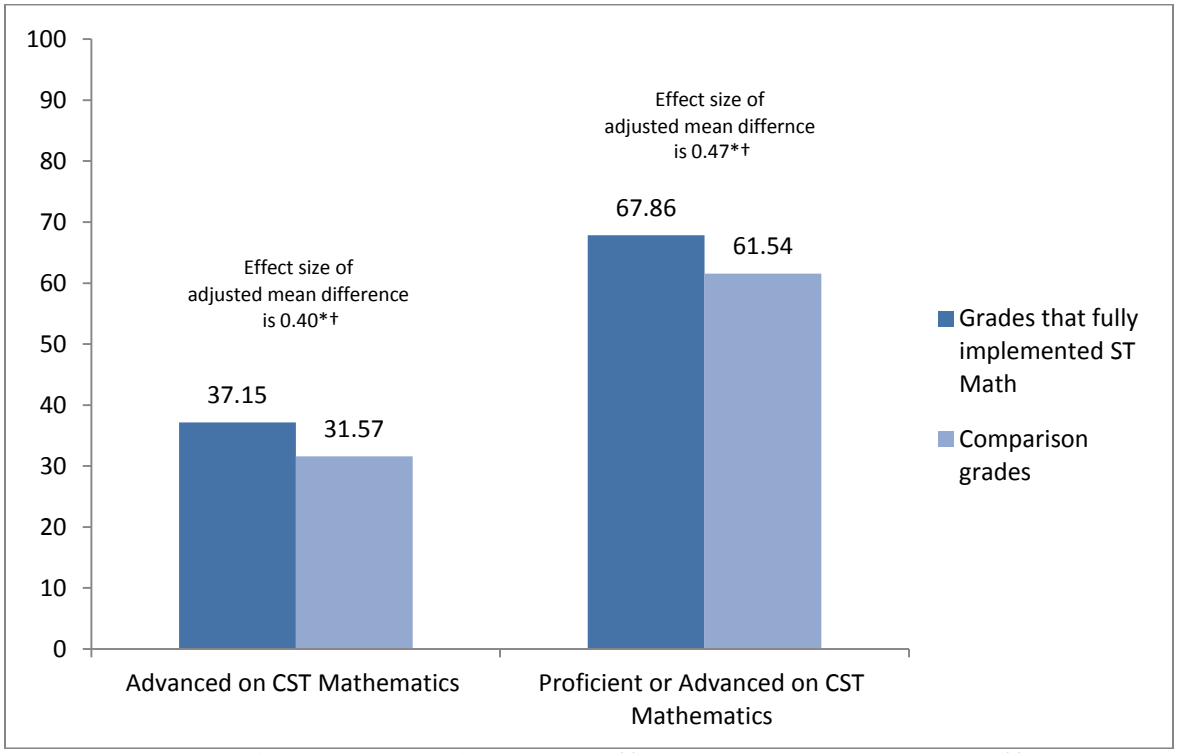


Exhibit reads: 37.15 percent of students in grades that fully implemented ST Math were advanced on the mathematics CST compared to 31.57 percent of students in grades that were not provided with the program.
* statistically significant at *p*-value < .05, two-tailed test
† statistically significant at < BH critical value, correcting for the false discovery rate for multiple comparison tests
Note: All outcomes adjusted for grade-level 2010 percent of students proficient or advanced, and for school-level percentage of Latino, Native American, and African American students, number of students enrolled, and number of students eligible for free or reduced-price lunch. Hierarchical linear modeling was used to account for the nesting of grades (n = 418) within schools (n = 306).

EVALUATION LIMITATIONS

The primary limitation to this ST Math evaluation is that, even though it compared grades that were similar on several known characteristics (e.g., grade-level mathematics proficiency from the year prior to ST Math being provided to treatment grades), the treatment and comparison groups may have differed in ways that were not measured, especially because the principals at schools with treatment grades volunteered to implement the ST Math program. In other words, because grades that were provided with ST Math elected to participate in the program, there may have been factors other than participating in ST Math (such as a greater focus on mathematics achievement) that contributed to improvements in mathematics outcomes in these grades. This is even more of an issue for the sample of grades that fully implemented ST Math because these schools not only volunteered to participate in ST Math, but these grades also chose (or were able) to fully implement the program.

An additional limitation of the evaluation is that, even though the samples of treatment and comparison grades were limited to those grades that had no previous exposure to ST Math, it is possible that ST Math had been implemented in other grades in these schools. So, it is possible that students in the treatment or comparison grades could have been exposed to ST Math in previous grades.

Evaluation of the MIND Research Institute's Spatial-Temporal Math (ST Math) Program in California

BACKGROUND

The stability of the U.S. economy and the productivity of its workforce depend on having a K–12 education system that produces students who possess strong mathematics skills (National Council of Teachers of Mathematics, 2009). Occupations in fields such as health care and science that have accounted for an increasing proportion of the workforce require more mathematics skills and higher executive functioning than occupations in fields that have been declining (Executive Office of the President Council of Economic Advisers, 2009). Unfortunately, the low mathematics achievement of many students in the U.S. poses a threat to their future academic and employment prospects as well as to the future competitiveness of the U.S. economy.

Although the 2013 National Assessment of Educational Progress (NAEP) scores in mathematics reached a new high for fourth-grade students, the rapid rate of improvement witnessed in past decades has slowed dramatically in recent years. From 1990 to 2011, the average scale score increased by 28 points. However, between 2005 and 2013, the average scale score increased by only 4 points, and between 2011 and 2013, it increased by only 1 point. In addition, 58 percent of fourth-grade students in 2013 scored below proficient on the NAEP mathematics assessment. Students in the fourth grade who scored below proficient did not demonstrate competency with skills such as computation with whole numbers, common fractions, and decimals as well as how to read and interpret data representations (e.g., bar graphs). Scoring below proficient also meant that students did not have a solid understanding of units of measurement for attributes such as length, area, and volume (National Center for Education Statistics, 2013). In California, the average scale score of 234 on the NAEP mathematics assessment for fourth graders was below the national average in 2013. Only 32 percent of fourth graders in California scored proficient or advanced on the NAEP mathematics assessment. Additionally, 26 percent of fourth graders scored below basic in 2013 (National Center for Education Statistics, 2013).

The Trends in International Mathematics and Science Study (TIMSS) allows for the comparison of the mathematics performance of U.S. students in fourth grade with their peers from countries across Africa, Asia, Europe, and Latin America. The TIMSS data from 2011 showed that U.S. fourth graders scored higher than the overall average on the mathematics assessment. However, fourth graders in eight industrialized countries, such as England, Japan, and Singapore, outperformed U.S. students on the mathematics assessment by margins that

reached statistical significance. Overall, the results from the TIMSS and NAEP indicate that effective mathematics interventions are needed in U.S. schools.

OVERVIEW OF THE ST MATH PROGRAM AND THE EVALUATION

Spatial-Temporal Math (ST Math) is game-based, instructional software for K–12 students created by the MIND Research Institute. The purpose of the program is to boost math comprehension through visual learning. ST Math is integrated into classroom instruction but can also be used in a computer lab or at home. The ST Math software games follow JiJi, a penguin. Students help JiJi pass obstacles by solving spatial math puzzles.

The MIND Research Institute contracted with the Evaluation Research Program at WestEd to conduct an independent assessment of ST Math in California schools. The evaluation compared the outcomes for grades that were provided with ST Math and matched grades in the same district that were not provided with ST Math. The three outcomes examined were: mathematics scale scores on the California Standards Test (CST),[1] the proportion of students who were advanced in mathematics based on CST scores, and the proportion of students who were either proficient or advanced in mathematics based on CST scores. One set of analyses examined outcomes for all grades that were provided with ST Math; additional analyses examined outcomes only for grades that fully implemented the program.

METHOD

The current study utilized a matched-comparison, quasi-experimental design that matched grades that were provided with ST Math to grades that were not provided with ST Math. The unit of analysis for the evaluation was "grade" as opposed to classroom or school. A "grade" included all the classrooms in a school that taught content for a specific grade level. For example, the data from an elementary school with three grade 4 classes were included in the evaluation as a single "grade."

WestEd examined three outcomes of interest: (1) grade-level 2011 CST mathematics scale scores; (2) the proportion of students in each grade who were advanced in mathematics based on these CST scores; and (3) the proportion of students in each grade who were either proficient or advanced in mathematics based on these CST scores. Analysis of covariance (ANCOVA) was used to adjust the outcomes for the influence of school-level demographic factors and grade-level proficiency rates from the year before ST Math was provided. The difference in adjusted outcomes between ST Math grades and comparison grades was calculated separately for all grades that were provided with ST Math and for only grades that fully implemented ST Math. For each of the three outcomes, the average difference across

[1] CST was the state standardized test that California used during the 2009/10 and 2010/11 school years.

grade level was examined, accounting for the relationship between grades and schools (i.e., using hierarchical linear modeling, HLM) because some schools contained multiple grade levels used in the evaluation.

SELECTION OF TREATMENT AND COMPARISON GRADES

WestEd conducted two series of analyses. This first set of analyses examined outcomes for all grades that were provided with ST Math, regardless of the extent to which the program was implemented (i.e., an intent-to-treat analysis). The second set of analyses included only a subset of these grades that fully implemented ST Math in that year (i.e., a treatment-on-treated analysis). Appendix A provides a flowchart of how the samples of treatment grades and the pool of comparison grades were identified. Samples of treatment grades and the pool of comparison grades were identified using CST mathematics data provided by the MIND Research Institute and school-level demographic data from the National Center for Education Statistics Elementary/Secondary Information system (http://nces.ed.gov/ccd/elsi/).

IDENTIFICATION OF THE SAMPLE OF GRADES PROVIDED WITH ST MATH

The treatment grades in the current evaluation were grades 2 through 5 that were provided with ST Math beginning in the 2010/11 school year. The treatment pool began with all 556 grades in California that were provided with ST Math beginning in 2010/11. Next, grades were eliminated from the treatment pool if the grades were missing CST mathematics data from 2010 (i.e., the pre-intervention year) or 2011 (i.e., the first intervention year). The CST data were provided by the MIND Research Institute. These CST data included scale scores and, based on the scale scores, percentages of students who were proficient in mathematics or advanced in mathematics. There were 36 grades missing either the 2010 or 2011 CST data, reducing the treatment pool to 520 grades.

In addition, treatment grades were excluded if their schools were missing the following data from the 2009/10 school year: school-level percentages of African American, Latino, Native American, and White students, students eligible for free or reduced-price lunch, and the number of students enrolled in the school. These data were not available for four grades in the pool, further reducing the treatment pool to 516 grades.

The purpose of the evaluation was to focus on grades other than those with the highest-performing students, thus 53 additional grades were eliminated from the pool of treatment grades because the percentage of students considered basic, proficient, or advanced in mathematics on the 2010 CST was one standard deviation above the mean of the percentage of students in the entire pool of treatment grades who were considered basic, proficient, or advanced in mathematics on the 2010 CST. This narrowed the final number of treatment grades to 463. These grades were in 212 different schools, and 159 of these schools contained multiple grade levels provided with ST Math beginning in 2010/11. This included 108 schools

with grade 2, 120 schools with grade 3, 116 schools with grade 4, and 119 schools with grade 5 (columns 1 and 2 of Exhibit 1).

IDENTIFICATION OF THE SAMPLE OF GRADES THAT FULLY IMPLEMENTED ST MATH

The second series of analyses included only grades that fully implemented ST Math in 2010/11. For the purposes of this study, full implementation was considered to have occurred for a grade when, at a particular school, at least 85 percent of the students enrolled in the grade had logged into ST Math during the academic year, and at least 50 percent of the grade-level material in ST Math was covered by those students.[2] Of the 463 grades in the treatment group, 209 grades (45 percent) fully implemented ST Math in 2010/11. The 209 grades were in 129 schools, and included 45 schools with grade 2, 63 schools with grade 3, 44 schools with grade 4, and 57 schools with grade 5 (columns 3 and 4 of Exhibit 1).

Exhibit 1. Number of Grades Provided with ST Math Included in the Evaluation, and Number of Grades Fully Implementing ST Math, by Grade Level

		Grades Fully Implementing ST Math	
Grade	Grades Provided with ST Math	Number of Grades	As a % of Grades Provided with ST Math
2	108	45	41.67
3	120	63	52.50
4	116	44	37.93
5	119	57	47.90
Total	463	209	45.14

Exhibit reads: For grade 2, 108 grades were provided with ST Math. Of these 108 grades, 45 (or 41.67 percent) fully implemented the program.

IDENTIFICATION OF THE POOL OF POTENTIAL COMPARISON GRADES

The comparison grades in the current evaluation were grades 2 through 5 that had not been provided with the ST Math program prior to, or during, the 2010/11 school year. There were 21,566 such grades in the pool of potential comparison grades. Next, grades were excluded if they were missing grade-level CST mathematics data from 2010 or 2011. Excluding grades that were missing these data reduced the comparison pool to 19,494 potential comparison grades. In addition, grades were excluded if they were missing data on any of the school-level characteristics necessary for matching and for conducting the outcomes analyses. Excluding grades with missing school-level demographic data further reduced the comparison pool of

[2] Enrollment was calculated by dividing the number of students who were enrolled in ST Math during 2010/11 by the number of students who took the mathematics CST in 2011. The data to calculate enrollment in ST Math were obtained from the MIND Research Institute, as were data on students' coverage of the material.

grades to 19,061 grades. Finally, because grades with high-performing students were not the focus of the evaluation, 3,180 additional grades were eliminated from the pool of comparison grades because the percentage of students considered basic, proficient, or advanced in mathematics on the 2010 CST was one standard deviation above the mean of the percentage of students in the entire pool of possible comparison grades who were considered basic, proficient, or advanced in mathematics on the 2010 CST.

The research team then selected only grades that were in districts where ST Math had first been offered in 2010/11. This last step occurred in order to identify potential comparison grades that shared geographically similar characteristics with the treatment grades (Cook, Shadish, & Wong, 2008). This step narrowed the number of grades in the pool of potential comparison grades to 3,196 grades in 1,802 schools. Of the 1,802 schools, 881 included multiple grade levels. This included 852 schools with grade 2,795 schools with grade 3,716 schools with grade 4, and 833 schools with grade 5.

MATCHING

For each treatment grade, WestEd identified a comparison grade from the pool of potential comparison grades. The purpose of matching was to create two groups that would be essentially equal on observable characteristics known to be related to mathematics achievement.[3] Several types of matching strategies exist (Guo & Fraser, 2010), and propensity score matching is one such technique. However, propensity score matching requires a larger sample size than possible in the current evaluation (Luellen, Shadish, & Clark, 2005). As an alternative, WestEd used Mahalanobis distance matching (specifically, Stata macro "mahascores" and a "greedy matching technique") to identify comparison grades (Stuart, 2009). The following characteristics were used to match ST Math grades with comparison grades: grade-level percentage of students proficient or advanced on the 2010 CST; and school-level percentages of African American, Latino, Native American, and White students, students eligible for free or reduced-price lunch, and the number of students enrolled in the school.

For the analyses of grades that were provided with ST Math, WestEd identified a different comparison grade to match to each grade that was provided with ST Math. Similarly, for the analyses of grades that fully implemented ST Math, WestEd identified a different comparison grade to match to each grade that fully implemented ST Math. The two groups of treatment grades were matched separately in order to maximize the closeness of the matches for each

[3] Matching is a quasi-experimental alternative to a randomized-control trial. When conducted with large samples, randomization makes the treatment and control groups equal on all characteristics other than the treatment condition, allowing for any differences between groups seen after the treatment or program to be causally determined as a result of exposure to the treatment or program. Without randomization, the possibility that two groups differ on other characteristics besides exposure to the treatment or program is a threat to causal conclusions (Shadish, Cook, & Campbell, 2002).

analysis. Thus, grades that were included in both the analyses of the larger sample and of the full implementation sample could have been matched to different grades for each of the analyses.

To examine the reliability of the matching technique, treatment and comparison grades were compared on the matching characteristics. The comparison and treatment grades for both analyses did not significantly differ on any of the matching characteristics (Exhibits B1–B8 in Appendix B). The difference for all comparisons between treatment and respective comparison grades was less than a quarter of a standard deviation, which is considered an acceptable level for minimizing bias between matched groups (Ho, Imai, King, & Stuart, 2007).

ANALYSES

ANCOVA models were used to examine the differences between ST Math grades and comparison grades on each of the three outcomes: 2011 CST mathematics scale scores, the proportion of students labeled proficient based on the 2011 CST mathematics scores, and the proportion of students labeled proficient or advanced based on the 2011 CST mathematics scores. Separate ANCOVAs were conducted for each grade level, 2 through 5. For each grade level, the ANCOVA models included the following covariates: grade-level percentage of students proficient or advanced based on 2010 CST mathematics scores; and school-level percentages of Latino, Native American, African American students, and students eligible for free or reduced-price lunch, and the number of students enrolled. When averaging ST Math and comparison differences across grade levels, WestEd used the same covariates in an HLM (Raudenbush & Bryk, 2002). HLM was used to account for nesting of grades within schools.[4] In addition, because the risk of Type-I error increases as the number of outcome comparisons increases, WestEd used the Benjamini-Hochberg (BH) correction for each set of three outcome analyses—both for the analyses at each grade level and for the analyses across grade levels (Benjamin & Hochberg, 1995). The results from each grade-level analysis, and for the analyses across grades, are presented with and without the BH correction.

RESULTS FOR GRADES PROVIDED WITH ST MATH

For grades that were provided with ST Math and their comparison grades, Exhibit C1 in Appendix C contains the unadjusted mean CST mathematics scale scores and standard deviations from the year before ST Math was provided to the treatment schools and from the first year of implementation. Appendix C also includes the mean percentage and standard

[4] For the HLM analyses: Level 1: $\text{Outcome}_{ij} = \beta_{0j} + \beta_{1j}(\text{Baseline Percent Proficient/Advanced})_{ij} + \beta_{2j}(\text{Treatment})_{ij} + \beta_{3j}(\text{Demographic Covariate 1})_{ij} + \ldots + \beta_{4j}(\text{Demographic Covariate 9})_{ij} + r_{ij}$.
Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Treatment Status})_j + u_{0j}$

deviations of students who were proficient, or advanced or proficient, based on the CST mathematics scores for both groups from the same two years.

The analyses of grades that were provided ST Math in 2010/11 revealed statistically significant differences for grade 2, but not for grades 3, 4, or 5 (Exhibit 2). Specifically, after accounting for several school-level characteristics and second-grade math proficiency rates from the year before ST Math was provided, second grades that were provided with the ST Math program had students with CST mathematics scale scores that were, on average, 8.09 points higher than the CST scores of students in second grades that were not provided with the ST Math program. In addition, second grades that were provided with the ST Math program had students who were considered advanced in mathematics at a rate that was, on average, 4.20 percentage points higher than for students in second grades that were not provided with the ST Math program. Finally, second grades that were provided with the ST Math program had students who were considered proficient or advanced in mathematics at a rate that was, on average, 4.44 percentage points higher than for students in second grades that were not provided with the program. The findings for all three outcomes were statistically significant after accounting for the fact that comparisons were made on multiple outcomes.

Exhibit 2. Differences on CST Mathematics Performance for All Grades Provided with ST Math, by Grade Level

Grade 2	Adjusted Mean		Adjusted Mean			
Outcome	ST Math (N = 108)	Comparison (N = 108)	Difference	*F*-test	Effect Size	*p* value
Scale score	375.19	367.10	8.09	5.86	0.26	.016*†
% advanced	32.58	28.38	4.20	6.94	0.29	.009*†
% proficient or advanced	63.01	58.57	4.44	7.44	0.29	.007*†
Grade 3	**Adjusted Mean**		**Adjusted Mean**			
Outcome	**ST Math (N = 120)**	**Comparison (N = 120)**	**Difference**	***F*-test**	**Effect Size**	***p* value**
Scale score	394.00	390.00	4.00	2.43	0.14	.121
% advanced	37.92	35.92	200	2.56	0.15	.111
% proficient or advanced	65.68	63.63	2.05	2.84	0.16	.093
Grade 4	**Adjusted Mean**		**Adjusted Mean**			
Outcome	**ST Math (N = 116)**	**Comparison (N = 116)**	**Difference**	***F*-test**	**Effect Size**	***p* value**
Scale score	384.79	383.33	1.46	0.33	0.06	.565
% advanced	41.04	40.11	0.93	0.40	0.00	.526
% proficient or advanced	67.55	67.36	0.19	0.02	0.01	.892
Grade 5	**Adjusted Mean**		**Adjusted Mean**			
Outcome	**ST Math (N = 119)**	**Comparison (N = 119)**	**Difference**	***F*-test**	**Effect Size**	***p* value**
Scale score	392.04	385.92	6.12	3.56	0.19	.064
% advanced	33.73	31.18	2.55	3.05	0.18	.082
% proficient or advanced	63.64	60.73	2.91	3.33	0.19	.069

Exhibit reads: The adjusted mean scale score of students in second grades provided with ST Math was 375.19 and the adjusted mean scale score of students in comparison second grades was 367.10, for an adjusted mean difference of 8.09, indicating a higher adjusted mean scale score for second grades provided with ST Math.
* statistically significant at *p*-value < .05, two-tailed test
† statistically significant at < BH critical value correcting for the false discovery rate under multiple testing within each grade
Note: All outcomes adjusted for grade-level 2010 percent proficient or advanced. All outcomes adjusted for the following school-level factors: percentages of Latino, Native American, and African American students; number of students enrolled, and the number of students eligible for free or reduced-price lunch. A positive adjusted mean difference indicates a higher mean for the ST Math group.

The pooled analyses for grades 2 through 5 revealed statistically significant differences between the treatment and comparison grades for all three outcomes, after accounting for the nesting of grades within schools, grade-level percent proficient or advanced on the 2010 CST, and several school-level characteristics (Exhibit 3). Specifically, grades that were provided with the ST Math program had students with average standardized CST mathematics scale scores that were higher than those of students in grades that were not provided with the ST Math program.

Exhibit 3. Differences on CST Mathematics Performance for All Grades Provided with ST Math, Across Grade Levels

Outcome	Adjusted Mean		Adjusted Mean Difference	*z*-test	Effect Size	*p* value
	ST Math	Comparison				
Scale score[a]	0.01	-0.15	0.16	2.86	0.16	.004*†
% advanced	35.01	32.54	2.47	2.84	0.17	.002*†
% advanced or proficient	64.88	62.58	2.30	3.16	0.16	.005*†

Exhibit reads: The average standardized adjusted mean scale score of students in grades provided with ST Math was 0.01 and the standardized adjusted mean scale score of students in comparison grades was -0.15, for an adjusted mean difference of 0.16, indicating a higher standardized adjusted mean scale score for grades provided with ST Math.
[a]Because CST scale scores are not vertically aligned across grades, standardized scores (i.e., z-scores) were used for the scale score analysis.
* statistically significant at *p*-value < .05, two-tailed test
† statistically significant at < BH critical value correcting for the false discovery rate under multiple testing
Note: All outcomes adjusted for grade-level 2010 percent proficient or advanced. All outcomes adjusted for the following school-level factors: percentages of Latino, Native American, and African American students; number of students enrolled, and the number of students eligible for free or reduced-price lunch. A positive adjusted mean difference indicates a higher mean for the ST Math group. Hierarchical linear modeling was used to account for nesting of grades within schools.
n = 926 grades in 544 schools.

The effect size of this difference between the groups on standardized CST mathematics scale scores was 0.16. The difference in percentile points that correspond to an effect size of 0.16 along a normal distribution of scale scores can be found in Exhibit 4. In this case, if the average comparison grade's scale score were at the 5th percentile in a ranking of all scale scores statewide, an effect size of 0.16 would mean that the average treatment grade's scale score is at the 7th percentile in statewide scale score ranking, for a difference of 2 percentile points. However, if the average comparison grade's scale score were at the 50th percentile, an effect size of 0.16 would mean that the average treatment grade's scale score is at the 56th percentile, for a difference of 6 percentile points.

Exhibit 4. Average Percentile Point Differences for Grades Provided with ST Math When Effect Size = 0.16.

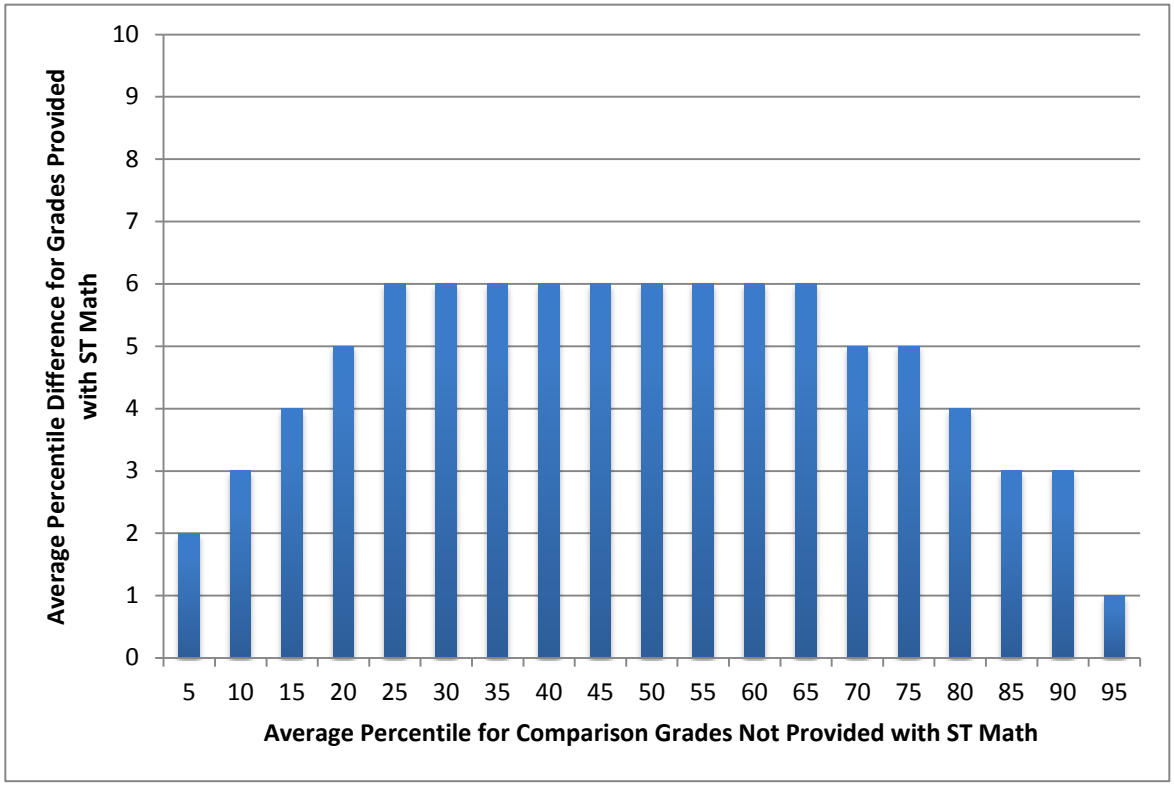


Exhibit reads: If the average comparison grade's student scale score were at the 5th percentile rank, an effect size of 0.16 would mean that the average ST Math grade's student scale score increased by 2 percentile points.

In addition, grades that were provided with the ST Math program had students who were considered advanced in mathematics on the CST at a rate that was, on average, 2.47 percentage points higher than for students in comparison grades that were not provided with the ST Math program. Finally, grades that were provided with the ST Math program had students who were considered proficient or advanced in mathematics on the CST at a rate that was, on average, 2.30 percentage points higher than for students in comparison grades that were not provided with the ST Math program. The findings for all three outcomes across grades were statistically significant after accounting for the fact that comparisons were made on multiple outcomes.

RESULTS FOR GRADES THAT FULLY IMPLEMENTED ST MATH

For grades that fully implemented ST Math and their comparison grades, Exhibit C2 in Appendix C contains the unadjusted mean CST mathematics scale scores and standard deviations from the year before ST Math was provided to the treatment schools and from the first year of implementation. Appendix C also includes the mean percentage and standard deviations of students who were proficient, or advanced or proficient, based on the CST mathematics scores for both groups from the same two years.

The analyses of grades that fully implemented ST Math in 2010/11 revealed statistically significant differences for grades 2, 3, and 5, but not for grade 4 (Exhibit 5). Specifically, after accounting for several school-level factors and second-grade math proficiency rates from the year before ST Math was provided, second grades that fully implemented ST Math had students with CST mathematics scale scores that were, on average, 15.48 points higher than the CST scores of students in matched second grades that were not provided with the program. In addition, second grades that fully implemented ST Math had students who were considered advanced in mathematics on the CST at a rate that was, on average, 7.64 percentage points higher than for second grades that were not provided with ST Math. Finally, second grades that fully implemented ST Math had students who were considered proficient or advanced in mathematics on the CST at a rate that was, on average, 8.21 percentage points higher than for second grades that were not provided with ST Math. The findings for all three outcomes in grade 2 were statistically significant after accounting for the fact that comparisons were made on multiple outcomes.

After adjusting for several school-level factors and third-grade math proficiency rates from the year before ST Math was provided, third grades that fully implemented ST Math program had students with CST mathematics scale scores that were, on average, 7.89 points higher than those of students in matched third grades that were not provided with the program. In addition, third grades that fully implemented ST Math had students who were considered advanced in mathematics on the CST at a rate that was, on average, 3.13 percentage points higher than for third grades that were not provided with ST Math. Finally, for third grades that fully implemented ST Math, the rate of students who were considered proficient or advanced in mathematics on the CST was, on average, 4.37 percentage points higher than for third grades that were not provided with ST Math. The findings for all three outcomes in grade 3 were statistically significant after accounting for the fact that comparisons were made on multiple outcomes.

After adjusting for several school-level factors and fifth-grade math proficiency rates from the year before ST Math was provided, fifth grades that fully implemented ST Math program had students with CST mathematics scale scores that were, on average, 17.52 points higher than those of students in matched fifth grades that were not provided with the program. In addition, fifth grades that fully implemented ST Math had students who were considered advanced in mathematics on the CST at a rate that was, on average, 6.55 percentage points higher than for fifth grades that were not provided with ST Math. Finally, for fifth grades that fully implemented ST Math, the rate of students who were considered proficient or advanced in mathematics on the CST was, on average, 8.52 percentage points higher than for fifth grades that were not provided with the program. The findings for all three outcomes in grade 5 were statistically significant after accounting for the fact that comparisons were made on multiple outcomes.

Exhibit 5. Differences on CST Mathematics Performance for Grades that Fully Implemented ST Math, by Grade Level

Grade 2	Adjusted Mean		Adjusted Mean			
Outcome	ST Math (N = 45)	Comparison (N = 45)	Difference	*F*-test	Effect Size	*p* value
Scale score	384.56	369.08	15.48	9.32	0.49	.003*†
% advanced	37.09	29.45	7.64	9.29	0.51	.003*†
% proficient or advanced	67.45	59.24	8.21	12.35	0.56	.001*†
Grade 3	**Adjusted Mean**		**Adjusted Mean**			
Outcome	ST Math (N = 63)	Comparison (N = 63)	Difference	*F*-test	Effect Size	*p* value
Scale score	396.01	388.12	7.89	5.21	0.28	.024*†
% advanced	38.71	35.58	3.13	4.00	0.25	.033*†
% proficient or advanced	66.61	62.24	4.37	7.21	0.35	.008*†
Grade 4	**Adjusted Mean**		**Adjusted Mean**			
Outcome	ST Math (N = 44)	Comparison (N = 44)	Difference	*F*-test	Effect Size	*p* value
Scale score	390.24	382.24	8.00	3.41	0.31	.068
% advanced	43.92	39.28	4.64	3.88	0.31	.052
% proficient or advanced	70.65	66.69	3.96	3.01	0.31	.087
Grade 5	**Adjusted Mean**		**Adjusted Mean**			
Outcome	ST Math (N = 57)	Comparison (N = 57)	Difference	*F*-test	Effect Size	*p* value
Scale score	399.93	382.41	17.52	14.63	0.56	.001*†
% advanced	36.18	29.63	6.55	10.88	0.49	.001*†
% proficient or advanced	67.47	58.95	8.52	14.52	0.59	.001*†

Exhibit reads: The adjusted mean scale score of students in second grades that fully implemented ST Math was 384.56 and the adjusted mean scale score of students in comparison second grades was 369.08, for an adjusted mean difference of 15.48, indicating a higher adjusted mean scale score for second grades that fully implemented ST Math.
* statistically significant at *p*-value < .05, two-tailed test
† statistically significant at < BH critical value correcting for the false discovery rate under multiple testing
Note: All outcomes adjusted for grade-level 2010 percent proficient or advanced. All outcomes adjusted for the following school-level factors: percentages of Latino, Native American, and African American students; number of students enrolled, and the number of students eligible for free or reduced-price lunch. A positive adjusted mean difference indicates a higher mean for the ST Math group.

The pooled analyses for grades 2 through 5 revealed statistically significant differences between the treatment and comparison grades for all three outcomes, after accounting for the nesting of grades within schools as well as grade-level percent proficient or advanced in 2010, and school-level characteristics (Exhibit 6). Specifically, grades that fully implemented the ST Math program had students with average standardized CST mathematics scale scores that were higher than the CST scores of students in grades that were not provided with the ST Math program.

Exhibit 6. Differences on CST Mathematics Performance for Grades that Fully Implemented ST Math, Across Grade Levels

Outcome	Adjusted Mean		Adjusted Mean Difference	*t*-test	Effect Size	*p* value
	ST Math	Comparison				
Scale score[a]	0.21	-0.21	0.42	5.68	0.42	.001*†
% advanced	37.15	31.57	5.58	5.84	0.40	.001*†
% proficient or advanced	67.86	61.54	6.32	5.41	0.47	.001*†

Exhibit reads: The standardized adjusted mean scale score of students in grades that fully implemented ST Math was 0.21 and the standardized adjusted mean scale score of students in comparison grades was -0.21, for an adjusted mean difference of 0.42, indicating a higher standardized adjusted mean scale score for grades that fully implemented ST Math.
[a]Because CST scale scores are not vertically aligned across grades, standardized scores (i.e., z-scores) were used for the scale score analysis.
* statistically significant at *p*-value < .05, two-tailed test
† statistically significant at < BH critical value correcting for the false discovery rate under multiple testing
Note: All outcomes adjusted for grade-level 2010 percent proficient or advanced. All outcomes adjusted for the following school-level factors: percentages of Latino, Native American, and African American students; number of students enrolled, and the number of students eligible for free or reduced-price lunch. A positive adjusted mean difference indicates a higher mean for the ST Math group. Hierarchical linear modeling was used to account for nesting of grades within schools.
n = 418 grades in 306 schools.

The effect size of this difference between the groups on CST mathematics scale scores was 0.42. The difference in percentile points that correspond to an effect size of 0.42 along a normal distribution of scale scores can be found in Exhibit 7. In this case, if the average comparison grade's scale score were at the 5th percentile in a ranking of all scale scores statewide, an effect size of 0.42 would mean that the average treatment grade's scale score is at the 11th percentile in statewide scale score ranking, for a difference of 6 percentile points. However, if the average comparison grade's scale score were at the 50th percentile, an effect size of 0.42 would mean that the average treatment grade's scale score is at the 66th percentile for a difference of 16 percentile points.

Exhibit 7. Average Percentile Point Differences for Grades that Fully Implemented ST Math When Effect Size = 0.42.

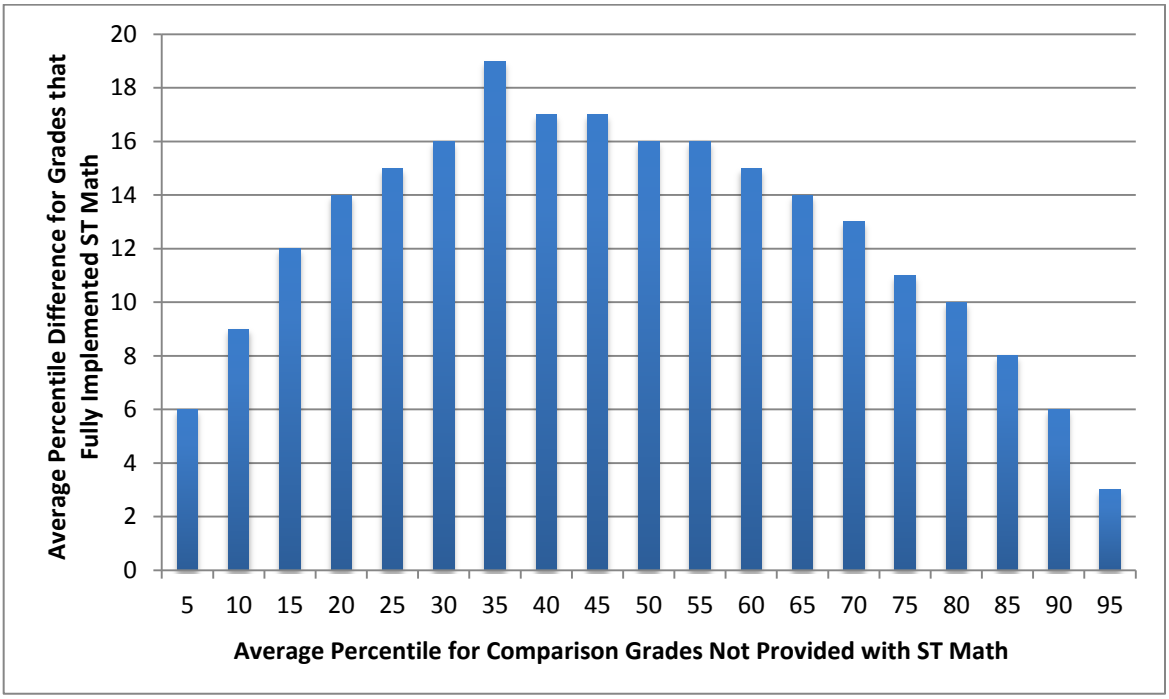


Exhibit reads: If the average comparison grade's student scale score were at the 5th percentile rank, an effect size of 0.42 would mean that the average ST Math grade's student scale score increased by 6 percentile points.

In addition, grades that fully implemented the ST Math program had students who were considered advanced in mathematics on the CST at a rate that was, on average, 5.58 percentage points higher than for comparison grades that were not provided with the ST Math program. Finally, grades that fully implemented the ST Math program had students who were considered proficient or advanced in mathematics on the CST at a rate that was, on average, 6.32 percentage points higher than for comparison grades that were not provided with the program. The findings for all three outcomes across grades were statistically significant after accounting for the fact that comparisons were made on multiple outcomes.

SUMMARY AND EVALUATION LIMITATIONS

The evaluation of ST Math in California used a quasi-experimental design that compared outcomes of students in grades that were provided with ST Math with outcomes of students in matched grades that were not provided with ST Math. Outcomes were compared separately for each grade level using analysis of covariance (ANCOVA) in order to account for differences in several school characteristics as well as grade-level mathematics performance prior to the provision of ST Math. In addition, differences in outcomes between the two groups were examined across all grade levels using HLM in order to account for the nesting of grades within schools.

RESULTS FOR GRADES PROVIDED WITH ST MATH

Students in second grades that were provided with the ST Math program had significantly higher mean mathematics scale scores on the CST compared to the CST scores of students in matched second grades that were not provided with the program. In addition, a significantly higher proportion of students in second grades that were provided with ST Math scored at the advanced level in mathematics on the CST than did students in second grades that were not provided with the program. Also, a significantly higher proportion of students in second grades that were provided with the program scored at the proficient or advanced levels in mathematics on the CST compared to the proportion of students in matched second grades that were not provided with the ST Math program. The difference for each of the three outcomes was statistically significant after correcting for the examination of multiple outcomes. No statistically significant differences were found in any other grades.

The pooled differences across grades were statistically significant for all three mathematics outcomes. Students in grades that were provided with ST Math had significantly higher mean mathematics scale scores on the CST compared to the CST scores of students in grade levels that were not provided with the program, for an effect size of 0.16. In addition, when pooling differences across grade levels, those that were provided with ST Math had a significantly larger proportion of students who scored at the advanced level on the CST (35.01 percent) compared to the proportion of students in matched grades that were not provided with the program (32.54 percent), for an effect size of 0.17. Also, grades that were provided with ST Math had a significantly larger proportion of students at either the proficient or advanced level on the CST (64.88 percent) compared to the proportion for grades that were not provided with the program (62.58 percent), for an effect size of 0.16. The difference for each of the three outcomes was statistically significant after correcting for the examination of multiple outcomes.

RESULTS FOR GRADES THAT FULLY IMPLEMENTED ST MATH

Students in grades 2, 3, and 5 that fully implemented the ST Math program had significantly higher mean mathematics CST scale scores compared to the CST scores of students in matched grades that were not provided with the program. In addition, a significantly larger proportion of students in grades 2, 3, and 5 that fully implemented the program scored at the advanced level in mathematics based on CST scores, and a significantly larger proportion of students scored at the proficient or advanced levels in mathematics based on CST scores, than the proportion of students in matched grades that were not provided with the program. The differences were statistically significant after correcting for the examination of multiple outcomes. No statistically significant differences were found in any of the outcomes for grade 4.

When pooling differences across grades, students in grades that fully implemented ST Math had significantly higher mean mathematics scale scores on the CST compared to the CST

scores of students in grades that were not provided with the program. In addition, grades that fully implemented ST Math had a significantly larger proportion of students who scored at the advanced level (37.15 percent) compared to the proportion for grades that were not provided with the program (31.57 percent). Also, grades that fully implemented ST Math had a significantly larger proportion of students at either the proficient or advanced level (67.86 percent) compared to the proportion for grades that were not provided with the program (61.54 percent). The difference for each of these outcomes was statistically significant after correcting for the examination of multiple outcomes.

EVALUATION LIMITATIONS AND POSSIBLE DESIGN IMPROVEMENTS

The primary limitation to this ST Math evaluation is that, even though it compared grades that were similar on several known characteristics (e.g., grade-level mathematics proficiency from the year prior to ST Math being provided to treatment schools), the treatment and comparison groups may have differed in ways that were not measured, especially because the principals at schools with treatment grades volunteered to implement the ST Math program. In other words, because grades that were provided with ST Math elected to participate in the program, there may have been factors other than participating in ST Math (such as a greater focus on mathematics achievement) that contributed to improvements in mathematics outcomes in these grades. This is even more of an issue for the sample of grades that fully implemented ST Math because these schools not only volunteered to participate in ST Math but also chose (or were able) to fully implement the program.

Future research could be strengthened in several ways. One way is to obtain grade-level outcome data for more than a single school year after treatment. Assuming that schools continue to implement ST Math beyond the first year, analyzing data from more than a single year would allow researchers to determine whether differences between ST Math grades and comparison grades increase with each year of exposure. In addition, obtaining student-level math outcomes would allow for a more precise estimate of standard errors and allow researchers to assess any potential impacts of the program on individual students over time, due to multiple years of exposure or long-term effects after exposure ends. Finally, as previously discussed, despite the careful matching of treatment and comparison grades on observable characteristics, it is possible that unobserved differences existed between the two sets of grades, and that these differences contributed (in whole or part) to the positive findings for ST Math. Without randomization, the possibility that groups differed on other characteristics besides exposure to ST Math impedes any casual conclusion (Shadish, Cook, & Campbell, 2002). However, a future randomized-control trial of ST Math, that is carefully executed, would allow for such a conclusion.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B, 57*, 289–300.

Cook, T. D., Shadish, W., & Wong, V. C. (2008). Three conditions under which observational studies produce the same results as experiments. *Journal of Policy Analysis and Management, 274,* 724–50.

Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: Sage Publications.

Ho, D. E., Imai, K., King, G. & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*, 199–236.

Luellen, J. K., Shadish, W. R., & Clark, M. H. (2005). Propensity scores: An introduction and experimental test. *Evaluation Review, 29*, 530–558.

Mullis, I. V. S., Martin, M .O., Foy, P., & Arora, A. (2012). *TIMSS 2011 results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

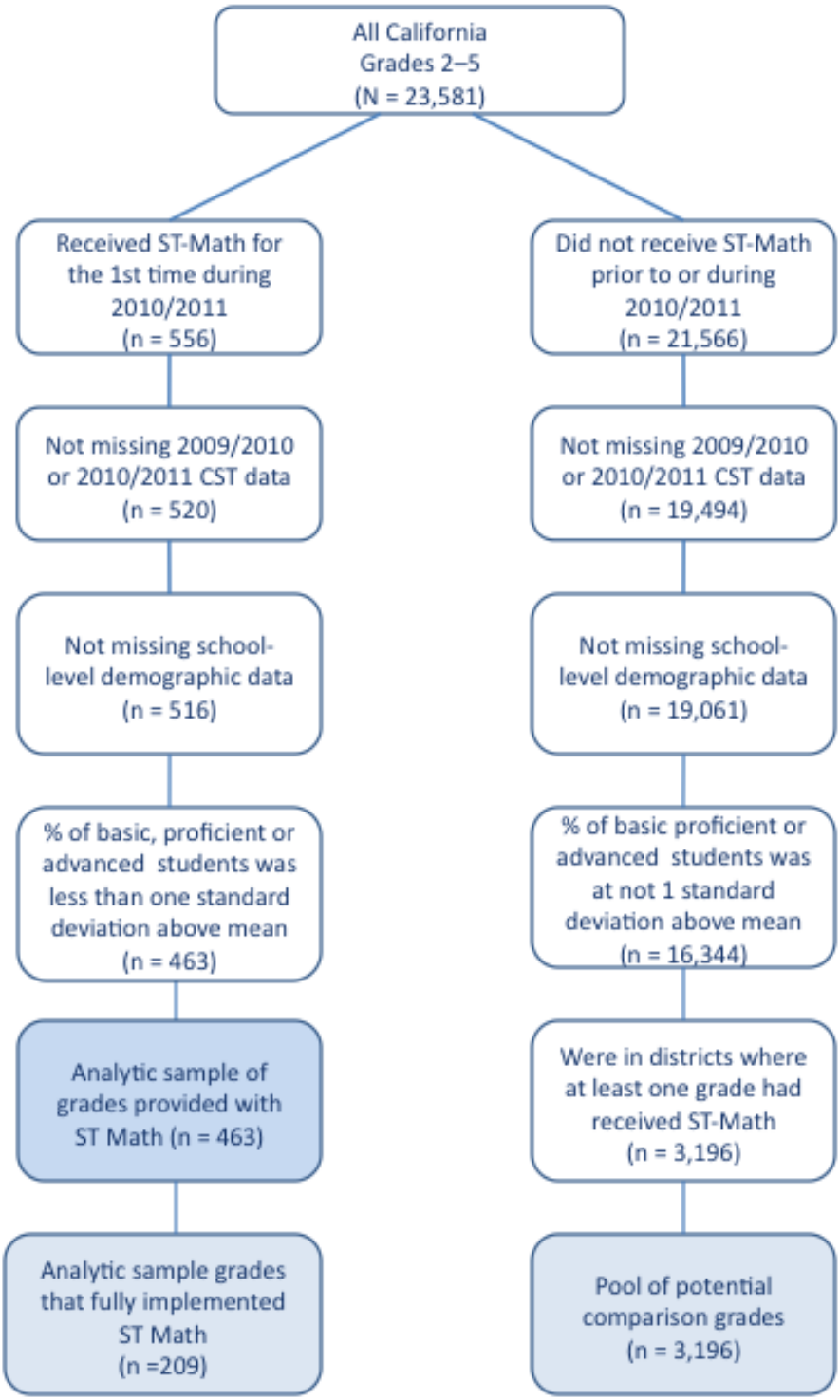
National Center for Education Statistics. (2013). *A first look: 2013 mathematics and reading*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inferences*. Boston, MA: Houghton Mifflin.

Stuart, E. A. (2009). *Matching methods for causal inference: A review and a look forward*. Baltimore, MD: Johns Hopkins Bloomberg School of Public Health.

Appendix A. Flow Chart of Sample Selection



Appendix B. Baseline Differences Between Treatment and Comparison Grades

GRADES PROVIDED WITH ST MATH AND COMPARISON GRADES

Exhibit B1. Grade 2

Outcome	Comparison M	Comparison SD	Treatment M	Treatment SD	*t* value	*p* value	effect size
% proficient or advanced on the mathematics CST	54.85	14.42	54.85	14.20	-0.17	.868	0.00
% Latino	67.26	28.38	65.20	29.22	0.53	.600	0.07
% Native American	0.37	0.49	0.54	1.74	-0.96	.337	-0.15
% African American	5.8	7.75	6.21	8.99	-0.33	.742	-0.05
% White	14.23	18.91	14.42	18.78	-0.08	.941	-0.01
Student enrollment	617.68	244.23	629.16	303.31	-0.31	.760	-0.04
% Free/reduced-price lunch	73.73	24.62	72.60	25.47	0.33	.739	0.05

Note: All factors were at the school level, except % proficient or advanced on the mathematics CST, which was at the grade level; M = mean and SD = standard deviation.

Exhibit B2. Grade 3

Outcome	Comparison M	Comparison SD	Treatment M	Treatment SD	*t* value	*p* value	effect size
% proficient or advanced on the mathematics CST	59.18	14.37	58.49	14.37	0.37	.713	0.05
% Latino	64.77	29.97	63.71	29.48	0.28	.783	0.04
% Native American	0.34	0.47	0.54	1.65	-1.27	.207	-0.19
% African American	6.20	9.37	6.93	9.84	-0.59	.555	-0.08
% White	14.76	20.07	14.24	18.98	0.21	.836	0.03
Student enrollment	617.99	253.35	637.91	291.97	-0.56	.573	-0.07
% Free/reduced-price lunch	70.83	27.90	71.23	26.42	-0.11	.910	-0.01

Note: All factors were at the school level, except % proficient or advanced on the mathematics CST, which was at the grade level; M = mean and SD = standard deviation.

Exhibit B3. Grade 4

Outcome	Comparison		Treatment		*t* value	*p* value	effect size
	M	SD	M	SD			
% proficient or advanced on the mathematics CST	62.12	12.56	62.91	13.47	-0.46	.647	-0.06
% Latino	66.80	28.81	64.31	29.80	0.65	.517	0.08
% Native American	0.37	0.39	0.40	0.42	-0.58	.564	-0.07
% African American	8.64	15.04	8.31	15.94	0.17	.869	0.02
% White	10.84	15.15	11.31	16.59	-0.23	.821	-0.03
Student enrollment	605.20	243.99	616.09	302.89	-0.30	.763	-0.04
% Free/reduced-price lunch	75.67	23.55	73.65	23.83	0.65	.518	0.09

Note: All factors were at the school level, except % proficient or advanced on the mathematics CST, which was at the grade level; M = mean and SD = standard deviation.

Exhibit B4. Grade 5

Outcome	Comparison		Treatment		*t* value	*p* value	effect size
	M	SD	M	SD			
% proficient or advanced on the mathematics CST	54.55	14.61	55.01	14.33	-0.25	.806	-0.03
% Latino	64.26	30.80	62.21	30.77	0.51	.609	0.07
% Native American	0.38	0.48	0.51	1.67	-0.82	.411	-0.12
% African American	7.90	14.87	8.19	16.10	-0.14	.885	-0.02
% White	11.99	17.40	12.49	17.21	-0.22	.824	-0.03
Student enrollment	609.64	243.00	625.38	300.01	-0.45	.657	-0.06
% Free/reduced-price lunch	72.91	27.02	71.67	26.21	0.36	.719	0.05

Note: All factors were at the school level, except % proficient or advanced on the mathematics CST, which was at the grade level; M = mean and SD = standard deviation.

Exhibit B5. Grade 2

Outcome	Comparison M	Comparison SD	Treatment M	Treatment SD	*t* value	*p* value	effect size
% proficient or advanced on the mathematics CST	56.78	13.43	56.04	14.31	0.25	.803	0.05
% Latino	69.19	25.48	66.96	27.56	0.40	.691	0.08
% Native American	0.33	0.32	0.38	0.38	-0.34	.734	-0.14
% African American	4.67	7.37	4.96	9.19	-0.17	.869	-0.04
% White	13.09	18.48	13.72	19.21	-0.16	.875	-0.03
Student enrollment	664.44	305.17	680.38	401.95	-0.21	.833	-0.05
% Free/reduced-price lunch	72.87	24.45	71.58	24.75	0.25	.805	0.05

Note: All factors were at the school level, except % proficient or advanced on the mathematics CST, which was at the grade level; M = mean and SD = standard deviation.

Exhibit B6. Grade 3

Outcome	Comparison M	Comparison SD	Treatment M	Treatment SD	*t* value	*p* value	effect size
% proficient or advanced on the mathematics CST	58.19	13.86	57.86	13.91	0.14	.893	0.02
% Latino	66.67	28.08	67.10	27.15	-0.09	.931	-0.02
% Native American	0.33	0.33	0.36	0.35	-0.48	.632	-0.09
% African American	6.72	10.24	6.20	9.32	0.30	.765	0.05
% White	13.36	18.10	12.98	16.71	0.12	.903	0.02
Student enrollment	639.35	278.54	672.67	346.95	-0.59	.553	-0.11
% Free/reduced-price lunch	72.89	25.08	72.07	24.85	0.18	.854	0.03

Note: All factors were at the school level, except % proficient or advanced on the mathematics CST, which was at the grade level; M = mean and SD = standard deviation.

Exhibit B7. Grade 4

Outcome	Comparison M	Comparison SD	Treatment M	Treatment SD	*t* value	*p* value	effect size
% proficient or advanced on the mathematics CST	61.59	13.46	63.57	13.57	-0.69	.494	-0.15
% Latino	67.34	30.77	65.84	30.35	0.23	.819	0.05
% Native American	0.35	0.34	0.38	0.38	-0.34	.734	-0.08
% African American	7.17	17.94	6.0	17.28	0.31	.756	0.07
% White	9.22	13.20	9.09	12.12	0.05	.962	0.01
Student enrollment	571.84	191.09	578.00	205.38	-0.15	.885	-0.03
% Free/reduced-price lunch	74.52	24.62	72.75	23.10	0.35	.729	0.07

Note: All factors were at the school level, except % proficient or advanced on the mathematics CST, which was at the grade level; M =mean and SD = standard deviation.

Exhibit B8. Grade 5

Outcome	Comparison M	Comparison SD	Treatment M	Treatment SD	*t* value	*p* value	effect size
% proficient or advanced on the mathematics CST	54.39	14.63	55.40	15.15	-0.37	.716	-0.07
% Latino	67.84	29.78	66.04	29.74	0.32	.748	0.06
% Native American	0.31	0.37	0.35	0.42	-0.50	.617	-0.10
% African American	7.61	16.52	7.09	17.81	0.61	.872	0.03
% White	9.46	15.27	10.69	15.70	-0.42	.672	-0.08
Student enrollment	621.91	273.74	631.56	347.36	-0.17	.869	-0.03
% Free/reduced-price lunch	74.37	26.22	72.50	24.84	0.393	.695	0.07

Note: All factors were at the school level, except % proficient or advanced on the mathematics CST, which was at the grade level; M = mean and SD = standard deviation.

Appendix C. Unadjusted Baseline and Follow-up Outcomes on CST Mathematics Performance

Exhibit C1. Grades in the Evaluation Provided with ST Math and Comparison Grades: Unadjusted CST Mathematics Outcomes at Baseline and After One Year, by Grade

		Baseline				Follow-up			
		Treatment		Comparison		Treatment		Comparison	
Outcome	Grade	M	SD	M	SD	M	SD	M	SD
Scale score	2	**366.14**	28.24	**365.84**	29.82	**375.47**	29.57	**366.82**	31.76
% advanced	2	**28.31**	13.02	**28.67**	14.10	**32.70**	14.14	**28.25**	14.27
% proficient or advanced	2	**55.18**	14.20	**54.85**	14.42	**63.09**	14.43	**58.49**	15.88
Scale score	3	**378.90**	30.20	**381.99**	33.34	**393.42**	27.93	**390.61**	30.54
% advanced	3	**31.21**	13.15	**32.29**	15.27	**37.67**	12.98	**36.18**	13.94
% proficient or advanced	3	**58.49**	14.37	**59.18**	14.37	**65.41**	13.18	**63.91**	12.91
Scale score	4	**378.36**	24.15	**376.95**	24.00	**385.48**	25.99	**382.64**	25.72
% advanced	4	**36.90**	12.91	**35.59**	12.80	**41.44**	14.76	**39.72**	14.51
% proficient or advanced	4	**62.77**	13.45	**62.12**	12.56	**67.83**	14.30	**67.08**	12.50
Scale score	5	**369.19**	27.58	**369.70**	30.56	**392.42**	34.45	**385.54**	30.78
% proficient	5	**23.15**	10.98	**23.75**	11.98	**33.90**	14.92	**31.01**	12.80
% proficient or advanced	5	**55.01**	14.33	**54.55**	14.61	**63.76**	15.90	**60.61**	13.88

Notes: M = mean; SD = standard deviation.

Exhibit C2. Grades that Fully Implemented ST Math and Comparison Grades: Unadjusted CST Mathematics Outcomes at Baseline and After One Year, by Grade

		Baseline				Follow-up			
		Treatment		Comparison		Treatment		Comparison	
Outcome	Grade	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Scale score	2	**370.53**	27.87	**368.44**	29.49	**384.52**	27.23	**369.12**	33.66
% advanced	2	**30.56**	13.24	**30.10**	12.96	**37.11**	13.56	**29.42**	15.37
% proficient or advanced	2	**56.04**	14.31	**56.45**	14.17	**67.33**	12.52	**59.36**	15.85
Scale score	3	**377.98**	28.57	**377.69**	30.70	**396.25**	24.61	**387.87**	30.74
% advanced	3	**30.81**	12.17	**30.39**	13.84	**38.83**	11.51	**35.46**	13.49
% proficient or advanced	3	**57.86**	13.91	**57.98**	13.87	**66.67**	11.60	**62.17**	12.91
Scale score	4	**379.95**	25.53	**375.26**	26.55	**391.33**	27.15	**381.05**	22.87
% advanced	4	**36.41**	13.08	**35.12**	13.78	**44.70**	15.73	**38.50**	14.03
% proficient or advanced	4	**63.57**	13.57	**61.05**	13.53	**71.20**	14.24	**66.14**	11.08
Scale score	5	**369.87**	28.21	**369.04**	30.51	**400.83**	31.65	**381.51**	28.26
% proficient	5	**23.72**	11.10	**23.71**	12.27	**36.60**	14.45	**29.21**	11.25
% proficient or advanced	5	**55.40**	15.15	**54.45**	14.75	**67.81**	14.19	**58.61**	13.25

Note: M = mean; SD = standard deviation