

This study looked at grade-level average California Standards Test (CST) Math scores at schools in Los Angeles Unified School District that used the ST Math program, and compared them to schools in the same district that did not use ST Math.

Who was included in the study?

The treatment group included all 45 of the [LA Math Initiative elementary schools](#) whose first year of using ST Math was the 2010-2011 school year. These schools were from the bottom 3 deciles of California state API (Academic Performance Index), and the grades using ST Math included more than 9,000 2nd through 5th grade students. The demographics were on average 95% Low Income, 85% Hispanic; 8% African American, and 45% English Language Learners.

What data was used?

Thanks to grade-wide implementations of ST Math, the study used grade-average 2010 and 2011 CST Math scores and proficiency level percentages reported on the California Department of Education [website](#).

What's the biggest take-away?

When averaged across all grades, the report found a statistically significant higher proportion of students who scored at the Advanced level in schools using ST Math (p. 15, Exhibit 17), and the proportion of students who scored above the No Child Left Behind (NCLB) requirement as either Proficient or Advanced had a 0.41 effect size (p. 15, Exhibit 17). This effect size is well beyond the federal What Works Clearinghouse criteria of 0.25 for “substantively important” effect. Effect size is the difference between the mean values of two sets of data—one treatment and one control—and is measured in units of standard deviation. Specifically, after one year, schools using ST Math saw a 10.4 point increase in the weighted mean percent of students scoring Proficient or Advanced on the CST, compared to 6.5 points at the comparison schools.

Did the study look at individual grade levels?

In addition to aggregating across all grade-levels 2-5, the study also looked at individual grade levels. While the effect sizes were often the same or higher than those of the overall grade-aggregated effect, due to the reduced sample sizes at individual grades levels, most grades did not show statistical significance on their own (p. 22, Discussion).

Could anything else have caused the improvement?

The study also looked at English Language Arts CST scores in order to determine whether the positive finding on math outcomes at the treatment schools may have been due to something other than ST Math (e.g., another general school-level program). No similar relationship in ELA was found (p. 17, Exhibit 22 and p. 22, Discussion).

How were the comparison schools chosen?

Two separate techniques for selecting comparison groups were used. One used the Mahalanobis distance matching to identify comparison grades similar in demographics, as well as in reading and math performance (p. 11). The other used a random sampling based on math performance characteristics only. The results from both techniques were consistent (p. 7, Exhibit 6 compared to p. 15, Exhibit 17).



What kinds of analyses were done for the study?

This report performed Intent-to-Treat (ITT) as well as Treatment-on-Treated (ToT) analyses:

- **ITT**, considered a more conservative estimate of impact, looked at differences between all grades that were provided ST Math, regardless of the extent to which they implemented the program (p. 2).
- **ToT** analyses included only grades that implemented ST Math to a minimally adequate level of coverage of math concepts.. That is, at least 85% of students in each grade used the program and covered at least 50% of the material.

Both analysis methods produced consistent results with the ToT analysis generally showing slightly higher effect sizes (p. 15, Exhibit 17; p. 19, Exhibit 26). However, the ToT had a reduced number of grades (e.g. only 10 grade 2's met ToT requirements) and so most ToT effects seen did not pass tests of statistical significance.

What outcomes did the study examine?

- Grade-level 2011 CST-Math mean scale scores. (No statistically significant impact found.)
- The proportion of students in each grade who were Advanced in math. (Statistically significant impact found.)
- The proportion of students in each grade who were either Proficient or Advanced in math. (Statistically significant impact found when aggregated across grades.)

How were the effect sizes determined?

Effect sizes at the grade level show how much the ST Math grades outgrew the comparison grades in terms of School-level standard deviations. Effect sizes for mean scale scores used a student level estimate method from the What Works Clearinghouse Handbook 2.1 (p. 12).

What modeling was used?

ANCOVA models were used to examine the effects of ST Math. Demographic as well as baseline math and reading performance covariates were evaluated (p. 12).

Where can I find more details in the report?

Section	Subject	Type	Comparison Technique	Main Finding
Pages 5-7	Math	Intent to Treat (ITT)	Math Performance	Exhibit 6
Pages 7-9	Math	Treatment on Treatment (ToT)	Math Performance	Exhibit 11
Pages 13-15	Math	ITT	Math, Reading, and Demographics	Exhibit 17
Pages 15-17	Reading	ITT	Math, Reading, and Demographics	Exhibit 22
Pages 17-19	Math	ToT	Math, Reading, and Demographics	Exhibit 26

