# Evaluation of ST Math in the Los Angeles Unified School District 

Submitted to:

Andrew Coulson

MIND Research Institute
3631 S. Harbor Blvd.
Suite 200
Santa Ana, CA 92704

Submitted by:
Staci Wendt
John Rice

WestEd
4665 Lampson Avenue
Los Alamitos, CA 90720

Date: April 2013
© 2013 WestEd. All rights reserved.
Requests for permission to reproduce any parts of this report should be directed to WestEd Publications Center, 730 Harrison Street, San Francisco, CA 94107-1242, 888-293-7833, fax 415-512-2024, permissions@WestEd.org, or http://www.WestEd.org/permissions.

Suggested Citation: Wendt, S. \& Rice, J. (2013). Evaluation of ST Math in the Los Angeles Unified School District. San Francisco: WestEd.

## About the Evaluation Research Program at WestEd

The Evaluation Research Program (ERP) provides policymakers and practitioners with impartial, evidence-based information to improve education and other services.

ERP staff have expertise in evaluating everything including migrant education programs, teacher evaluation systems, school reform and turnaround efforts, online learning environments, and data systems. Staff apply rigorous social science methodology to produce accurate and reliable studies, then communicate actionable findings on the implementation and effectiveness of programs and policies.

## About WestEd

WestEd, a nonprofit research, development, and service agency, works with education and other communities to promote excellence, achieve equity, and improve learning for children, youth, and adults. While WestEd serves the states of Arizona, California, Nevada, and Utah as one of the nation's Regional Educational Laboratories, our agency's work extends throughout the United States and abroad. It has 17 offices nationwide, from Washington and Boston to Arizona, Southern California, and its headquarters in San Francisco. For more information about WestEd, visit our website: WestEd.org; call 415.565 .3000 or, toll-free, (877) 4-WestEd; or write: WestEd / 730 Harrison Street / San Francisco, CA 94107-1242.

## Table of Contents

Executive Summary ..... i
Background ..... 1
Analyses Using Comparison Groups Selected by the MIND Research Institute ..... 2
Method ..... 2
Results for ITT Analyses ..... 5
Results for ToT Analyses ..... 7
Analyses Using Comparison Schools Selected by WestEd ..... 10
Method ..... 10
Identification of Comparison Grades ..... 10
Data and Analyses. ..... 11
Results for the ITT Analyses ..... 13
Results for the ToT Analyses. ..... 17
Discussion and Next Steps ..... 22
References ..... 24
Appendix A. Baseline Comparisons with Comparison Groups Selected by the MIND Research Institute ..... 26
ITT Sample ..... 26
ToT Sample ..... 28
Appendix B. Example of the Meta-analytic Technique Used to Pool Effect Sizes Across Grades ..... 31
Appendix C. Baseline Comparisons Using Comparison Groups Selected by WestEd ..... 32

## Executive Summary

The MIND Research Institute contracted with the Evaluation Research Group at WestEd to conduct an independent assessment of the effects of ST Math in the Los Angeles Unified School District (LAUSD) in grades 2 through 5 using two sets of analyses. The first set used comparison grades selected by the MIND Research Institute to examine differences in grade-level California Standards Test (CST) Math mean scale scores as well as the proportions of students who were Proficient, or Proficient or Advanced in math. The second set of analyses used comparison grades selected by WestEd. The outcomes for these analyses were grade-level CST Math and CST Reading mean scales scores as well as the proportions of students in each grade who were Proficient, or Proficient or Advanced in math, and the proportions of students who were Proficient, or Proficient or Advanced in reading. Both sets of analyses estimated program effects using intent-to-treat (ITT) and treatment-on-treated (ToT) analyses.

## RESULTS FROM ANALYSES USING COMPARISON GROUPS SELECTED BY THE MIND RESEARCH INSTITUTE

The ITT analyses found that second grades that were provided with the program had significantly higher math mean scale scores compared to second grades that were not provided with the ST Math program, as well as a significantly greater proportion of students scoring at the Advanced level, and a significantly greater proportion of students scoring at the Proficient or Advanced levels in math. In addition, third grades that were provided with the program had significantly greater proportions of students scoring at the Proficient or Advanced levels in math compared to third grades that were not provided with the ST Math program All these differences met the threshold for statistical significance after correcting for multiple-comparison testing. No statistically significant differences were found in any other grades. In addition, when averaging the ITT effect sizes across grades, those that were provided with ST Math had a significantly greater proportion of students who scored at the Advanced level, and a significantly greater proportion of students at either the Proficient or Advanced level compared to grades that were not provided with ST Math. These differences were statistically significant after the correction for multiple comparisons.

The To T analyses found that fifth grades that were provided with the program had significantly higher math mean scale scores compared to fifth grades that were not provided with the ST Math program, as well as a significantly greater proportion of students scoring at the Proficient or Advanced levels in math. These differences met the threshold for statistical significance after correcting for multiple-comparison testing. No statistically significant differences were found in any other grades and there were no statistically significant differences when averaging the ToT effect sizes across grades.

## RESULTS FROM ANALYSES OF MATH OUTCOMES USING COMPARISON GROUPS SELECTED BY WESTED

The ITT analyses of math outcomes found that fourth grades that were provided with the program had significantly higher math mean scale scores compared to fourth grades that were not provided with the ST Math program, but this difference was no longer statistically significant after correcting for multiple-comparison testing. No statistically significant differences were found in any other grades. When averaging the ITT effect sizes across grades, those that were provided with ST Math had a significantly greater proportion of students who scored at the Advanced level, and a significantly greater proportion of students at either the Proficient or Advanced level compared to grades that were not provided with ST Math. These differences were statistically significant after the correction for multiple comparisons.

The ToT analyses of math outcomes found that fourth grades that were provided with the program had significantly greater proportions of students who scored at the Advanced level compared to grades that were not provided with ST Math. This difference was statistically significant after correcting for multiple comparisons. No statistically significant differences were found in any other grades. When averaging the ToT effect sizes across grades, those that were provided with ST Math had significantly greater proportions of students who scored at the Proficient or Advanced levels compared to grades that were not provided with ST Math. This difference was no longer statistically significant after correcting for multiple comparisons.

## RESULTS FROM ANALYSES OF READING OUTCOMES USING COMPARISON GROUPS SELECTED BY WESTED

Both the ITT and ToT analyses of reading outcomes found that fifth grades that were provided with the program had significantly greater proportions of students who scored at the Proficient or Advanced levels in reading compared to grades that were not provided with ST Math. For the ITT analysis, but not the ToT analysis, the difference was statistically significant after correcting for multiple comparisons. However, the ToT analyses found that third grades that were provided with the program had significantly smaller proportions of students who scored at the Proficient or Advanced levels compared to grades that were not provided with ST Math. There were no statistically significant differences when averaging the ITT or ToT effect sizes for reading outcomes across grades. Therefore, it is inconclusive whether the findings for English Language Learners (ELA) outcomes indicate that the positive findings for ST Math are due to some other factor associated with treatment schools.

If a larger sample of grades had been included in the analyses, it is likely that a greater number of statistically significant differences would have been found for individual grade levels, particularly for math outcomes. For example, when using the WestEd-selected comparison group, the effect size (d) for a number of math outcomes in grades 2,4 , and 5 was over 0.40 with the largest being 0.76 . Assuming similar effect sizes as those found in the current study, pooling effects across a larger
number of schools (hence, grades) would likely reveal statistically significant differences in math outcomes for more grades.

Future research on ST Math could be strengthened in several ways, including analyzing grade-level outcomes across multiple school years using student-level outcomes, and estimating program effects through a randomized-control trial.

## Background

In 2012, the MIND Research Institute conducted an analysis of its ST Math software program by comparing grade-level math data from elementary schools in the Los Angeles Unified School District (LAUSD) that implemented the program with grade-level data from matched grades in LAUSD elementary schools that had not implemented the program. The study outcomes were grade-level California Standards Test (CST) math scale scores and math proficiency rates.

The MIND Research Institute contracted with the Evaluation Research Group at WestEd to conduct an independent assessment of the effects of ST Math in LAUSD. WestEd examined the effects of ST Math in LAUSD through two sets of analyses. The first set used the comparison groups selected by the MIND Research Institute. The outcomes examined in these analyses were grade-level CST Math mean scale scores as well as the proportions of students who were Proficient, or Proficient or Advanced in math. The second set of analyses used comparison groups selected by WestEd. The outcomes for these analyses were CST Math and CST Reading mean scales scores as well as the proportions of students who were Proficient, or Proficient or Advanced in math, and the proportions of students who were Proficient, or Proficient or Advanced in reading. Both sets of analyses estimated program effects using intent-to-treat (ITT) and treatment-on-treated (ToT) analyses. This report discusses the methods and results for each set of analyses separately. It also briefly discusses the implications of the findings and suggested next steps for further research on the ST Math software program.

## Analyses Using Comparison Groups Selected by the MIND Research Institute

## METHOD

In order to assess the effects of the ST Math software program, the first group of analyses utilized the comparison groups previously identified by the MIND Research Institute. WestEd examined three outcomes of interest in these analyses: (1) grade-level 2011 CST Math mean scale scores; (2) the proportion of students in each grade who were Advanced in math; (3) and the proportion of students in each grade who were either Proficient or Advanced in math.

WestEd conducted two types of analyses: ITT and ToT. The ITT analyses examined differences between all grades that were provided ST Math (regardless of the extent to which they implemented it) and the comparison grades. The ITT analyses provide an estimate of program effects when implementation is imperfect and is considered a conservative estimate of treatment effects (Gupta, 2011). The ToT analyses provide estimates of program effects in situations only when high levels of implementation occurred. In this case, the treatment group used in the ToT analyses was comprised of grades that received ST Math, and where at least 85 percent of the students in these grades had logged into the software program and covered at least 50 percent of the material.

For each type of analysis, WestEd used the comparison grades selected by the MIND Research Institute. The comparison grades used in the ITT analyses were randomly selected from a pool of schools in LAUSD that never used the ST Math program. ${ }^{1}$ Grades from these schools were matched to treatment grades based on grade-level 2010 CST Math mean scale scores and the proportion of students at each proficiency level (Advanced, Proficient, Basic, Below Basic, Far Below Basic) within each grade. The MIND Research Institute selected a separate group of comparison grades for the ToT analyses using the same procedure as the one for selecting the ITT comparison groups. Twelve schools that were included as comparison schools in the ITT analyses were also included as comparison schools in the ToT analyses. Exhibit 1 displays the sample sizes for the ITT and ToT analyses.

[^0]Exhibit 1. Sample Sizes - Analyses Using the Comparison Grades Selected by the MIND Research Institute

| Grade Level | Intent-to-Treat (ITT) Analyses |  | Treatment-on-Treated (ToT) Analyses |  |
| :---: | :---: | :---: | :---: | :---: |
|  | \# of Treatment <br> Grades | \# of Comparison <br> Grades | \# of Treatment <br> Grades | \# of Comparison <br> Grades |
| 2 | 32 | 55 | 10 | 19 |
| 3 | 40 | 65 | 20 | 39 |
| 4 | 26 | 49 | 9 | 17 |
| 5 | 14 | 25 | 8 | 16 |

Note: One school had been inadvertently included as both a treatment and comparison school in the original analysis performed by the MIND Research Institute. This school was considered only a treatment school in WestEd's analysis.

WestEd conducted t-tests between the ITT treatment grades and its set of comparison grades and between the ToT treatment grades and its set of comparison grades. The t-tests examined differences between the following baseline measures: grade-level 2010 CST Math mean scale scores, and school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and schoollevel student enrollment size. Some significant differences were found (Appendix A). For grade 2, the comparison group of schools had significantly higher CST Reading mean scale scores compared to the treatment schools, whereas the treatment schools had a significantly higher percentage of Latino students, English language learner students, and students eligible for free or reduced-price lunch. Similarly, for grade 3, the treatment schools had a higher percentage of Latino students, English language learners, and students eligible for free or reduced-price lunch. The grade 3 comparison schools had a significantly higher percentage of White students. For grade 4, the treatment schools had a significantly higher percentage of English language learner students and students eligible for free or reduced-price lunch. For grade 5 the treatment schools had a significantly higher percentage of English language learners. No other significant differences were found.

For both the ITT and ToT analyses, WestEd conducted outcome analyses separately for each grade using an analysis of covariance (ANCOVA) for each of the three outcomes. ANCOVA adjusts the outcome variable means as if both groups were equal on the baseline measure and allows for more precise estimate of the effect of the program (Tabachnik \& Fidell, 2006). For each of the three 2011 math outcomes examined, an ANCOVA used the corresponding baseline math outcome in 2010 as the covariate. Grade-level 2010 CST-math mean scale scores and grade-level 2010 CST Reading mean scale scores were also included in all analyses. In addition, school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students were included as school-level covariates in all the analyses as was school-level student enrollment size.

In order to calculate the effect size for CST mean scale scores, WestEd used the grade-level standard deviations for the CST Math exam in 2011 in LAUSD and divided the standard deviations by the square root of 0.20 . The denominator in this equation includes the intraclass correlation (ICC) of .20, which accounts for clustering of students within grades and is an accepted norm (What Works

Clearinghouse Handbook 2.1, 2011). This provides a student-level estimate. The use of the studentlevel standard deviations provides a more accurate estimate of the effect size than grade- or schoollevel standards deviations (Lipsey et al., 2012), and the effect sizes calculated using the student-level standard deviations were comparable to effect sizes from other electronic mathematics programs (Slavin \& Lake, 2008; Cheung \& Slavin 2011). Student-level or grade-level standard deviations were not available for the other two outcomes: the proportion of students scoring at the Advanced level in math and the proportion of students scoring Proficient or Advanced in math. Therefore, schoollevel standard deviations were used in the calculation of the effect sizes for these two outcomes.

WestEd used a meta-analytic technique (Lipsey \& Wilson, 2001) to determine the effect size of the program across grades. ${ }^{2}$. The meta-analytic technique calculates the average effect size, weighting each individual effect size by a term that represents the precision of the estimate: the inverse variance. The precision of the effect size estimate is impacted by the sample size; the larger the sample size, the smaller the sampling error (and the more precise the effect size estimate). Following Lipsey and Wilson (2001), the inverse variance weight was calculated with the following equation:

$$
\begin{equation*}
\omega=\frac{1}{S E^{2}} \tag{1}
\end{equation*}
$$

where $\omega$ is the inverse variance weight and standard error (SE) is equal to:

$$
\begin{equation*}
S E=\sqrt{\frac{n_{G 1}+n_{G 2}}{n_{G 1} n_{G 2}}+\frac{E S^{2}}{2\left(n_{G 1}+n_{G 2}\right)}} \tag{2}
\end{equation*}
$$

Where $n_{G 1}$ is the sample size of group 1 (the comparison group) and $n_{G 2}$ is the sample size of group 2 (the treatment group). Using the inverse variance weights, the average effect size is calculated using equation 3 :

$$
\begin{equation*}
\overline{E S}=\frac{\sum\left(\omega_{i} E S_{i}\right)}{\sum \omega_{i}} \tag{3}
\end{equation*}
$$

Where $\overline{E S}$ is the average effect size, $\omega_{\mathrm{i}}$ is the inverse variance for group $i$ and $E S_{i}$ is the effect size for group $i$.

Finally, WestEd examined multiple outcomes in each grade and across grades. As the number of outcome comparisons increases, the likelihood of Type-I error increases (i.e., identifying a statistically significant difference when one does not actually exist). To address this issue, WestEd used the Benjamini-Hochberg (BH) correction for comparison of each group of three outcomes

[^1]within a grade-level (Benjamini \& Hochberg, 1995; Schochet, 2008). ${ }^{3}$ The results of the analyses are presented both with and without the BH corrections.

## RESULTS FOR ITT ANALYSES

The results for the ITT analyses of 2011 CST Math outcomes, by grade level, appear in Exhibits 2 through 5 . There were statistically significant differences for grades 2 and 3 , and when pooling effect sizes across grades. Specifically, controlling for grade-level baseline achievement scores and schoollevel demographic variables, second grades that were provided the ST Math program had significantly higher 2011 CST Math mean scale scores compared to second grades that were not provided the ST Math program. In addition, second grades that were provided the ST Math program had a significantly greater percentage of students at the Advanced level compared to second grades that were not provided the ST Math program, after controlling for grade-level baseline achievement scores and school-level demographic variables. Finally, when controlling for grade-level baseline achievement scores and school-level demographic variables, second grades provided with ST Math had a significantly greater proportion of students who scored at Proficient or Advanced levels compared to second grades that were not provided with ST Math. All these differences were statistically significant after correcting for multiple comparisons.

When controlling for grade-level baseline achievement scores and school-level demographic variables, third grades provided with ST Math had a significantly greater proportion of students that scored at Proficient or Advanced levels compared to third grades that were not provided with ST Math. No other statistically significant differences were found at the other grade levels.

The meta-analysis of the combined grade-level results for each outcome (Exhibit 6) revealed statistically significant differences for the proportion of students who scored at the Advanced level and for the proportion of students who scored at either the Proficient or Advanced level. These differences were statistically significant both with and without the correction for multiple comparisons. There were no statistically significant differences between groups on the 2011 CST Math mean scale scores.

[^2]Exhibit 2. ITT Analysis - Differences on 2011 CST Math in Grade 2

| Outcome | $\frac{\text { Comparison Group }}{(n=55)}$ |  | $\frac{\text { Treatment Group }}{(\mathrm{n}=32)}$ |  | F | df | Partial$\eta^{2}$ | d | $p$ | BHcorrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Adjusted Mean or Mean \% | SE | Adjusted Mean or Mean \% | SE |  |  |  |  |  |  |
| Scale Score | 356.30 | 3.11 | 370.22 | 4.17 | 6.60 | 1,77 | . 079 | 0.25 | .012* | .017† |
| \% Advanced | 23.56 | 1.51 | 30.10 | 2.03 | 6.13 | 1,77 | . 075 | 0.66 | .016* | .033 ${ }^{+}$ |
| \% Proficient or Advanced | 53.44 | 1.46 | 59.56 | 1.96 | 5.75 | 1,77 | . 070 | 0.57 | .019* | .050† |

* statistically significant at $p$-value $<0.05$, two-tailed test
$\dagger$ statistically significant at $<\mathrm{BH}$ critical value correcting for the false discovery rate under multiple testing
Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 3. ITT Analysis - Differences on 2011 CST Math in Grade 3

| Outcome | $\frac{\text { Comparison Group }}{(n=66)}$ |  | $\frac{\text { Treatment Group }}{(\mathrm{n}=40)}$ |  | F | df | Partial $\eta^{2}$ | d | $p$ | BHcorrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Adjusted <br> Mean or <br> Mean \% | SE | Adjusted <br> Mean or <br> Mean \% | SE |  |  |  |  |  |  |
| Scale Score | 379.03 | 2.63 | 387.57 | 3.40 | 3.74 | 1,95 | . 038 | 0.15 | . 056 | . 033 |
| \% Advanced | 31.88 | 1.10 | 34.95 | 1.42 | 2.75 | 1,94 | . 028 | 0.35 | . 100 | . 050 |
| \% Proficient or Advanced | 58.13 | 1.33 | 63.62 | 1.73 | 5.96 | 1,94 | . 060 | 0.51 | .017* | .017† |

* statistically significant at $p$-value $<0.05$, two-tailed test
$\dagger$ statistically significant at $<\mathrm{BH}$ critical value correcting for the false discovery rate under multiple testing
Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 4. ITT Analysis - Differences on 2011 CST Math in Grade 4

| Outcome | $\frac{\text { Comparison Group }}{(n=49)}$ |  | $\frac{\text { Treatment Group }}{(\mathrm{n}=26)}$ |  | F | df | Partial$\eta^{2}$ | d | $p$ | BHcorrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Adjusted Mean or Mean \% | SE | Adjusted Mean or Mean \% | SE |  |  |  |  |  |  |
| Scale Score | 374.87 | 3.05 | 385.73 | 4.30 | 3.94 | 1, 65 | . 057 | 0.20 | . 051 | . 017 |
| \% Advanced | 36.68 | 1.76 | 42.06 | 2.48 | 2.91 | 1,64 | . 043 | 0.44 | . 093 | . 033 |
| \% Proficient or Advanced | 62.33 | 1.63 | 66.68 | 2.31 | 2.15 | 1, 64 | . 033 | 0.38 | . 147 | . 050 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 5. ITT Analysis - Differences on 2011 CST Math in Grade 5

| Outcome | $\begin{gathered} \text { Comparison Group } \\ (n=25) \end{gathered}$ |  | $\frac{\text { Treatment Group }}{(\mathrm{n}=13)}$ |  | F | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BHcorrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Adjusted <br> Mean or <br> Mean \% | SE | Adjusted <br> Mean or <br> Mean \% | SE |  |  |  |  |  |  |
| Scale Score | 367.02 | 5.17 | 377.90 | 7.19 | 1.34 | 1,29 | . 044 | 0.16 | . 256 | . 033 |
| \% Advanced | 26.04 | 2.12 | 27.14 | 2.95 | 0.08 | 1,28 | . 003 | 0.11 | . 777 | . 050 |
| \% Proficient or Advanced | 51.36 | 2.62 | 58.43 | 3.65 | 2.18 | 1,28 | . 072 | 0.55 | . 151 | . 017 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 6. ITT Analysis - Effect Sizes of 2011 CST Math Outcomes Across Grade Levels

| 2011 CST Math Outcome | Mean <br> Effect Size | $95 \%$ Cl | Z-score | p | BH- <br> corrected <br> critical <br> value |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Mean Scale Score | 0.19 | $-0.04,0.40$ | 1.61 | .107 | .050 |
| Mean Percent Advanced | 0.40 | $0.19,0.58$ | 3.53 | $.000^{*}$ | $.033^{\dagger}$ |
| Mean Percent Proficient or Advanced | 0.47 | $0.28,0.62$ | 4.55 | $.000^{*}$ | $.017+$ |

* statistically significant at $p$-value $<0.05$, two-tailed test
$\dagger$ statistically significant at $<\mathrm{BH}$ critical value correcting for the false discovery rate under multiple testing


## RESULTS FOR TOT ANALYSES

The ToT analyses examined math outcomes for grades that were provided with the ST Math software program, and where at least 85 percent of the students logged into ST Math and completed at least 50 percent of the material. Grades that were provided with the ST Math program but that did not meet these conditions were excluded from the ToT analyses. No significant differences were found between the ToT treatment and comparison grades with regards to the baseline measures (Appendix A).

The results for the ANCOVAs of 2011 CST Math outcomes, by grade level, using the ToT sample appear in Exhibits 7 through 10. Controlling for grade-level baseline achievement scores and schoollevel demographic variables, fifth grades that were provided the ST Math program had significantly higher 2011 CST Math mean scale scores compared to fifth grades that were not provided the ST Math program. Likewise, when controlling for grade-level baseline achievement scores and school-level demographic variables, fifth grades provided with ST Math had a significantly greater proportion of students who scores at Proficient or Advanced levels compared to fifth grades that were not provided with ST Math. All these differences were statistically significant both with and without the correction for multiple comparisons. No other statistically significant differences were found for grades 2, 3, or 4 .

Similar to the ITT analysis, the grade-level results were included in a meta-analysis to examine the effect of ST Math across grades. None of the mean effect sizes were statistically significant (Exhibit 11).

Exhibit 7. ToT Analysis - Differences on 2011 CST Math in Grade 2

| Outcome | Comparison <br> Group ( $\mathrm{n}=19$ ) |  | $\frac{\text { Treatment Group }}{(n=10)}$ |  | F | df | Partial $\eta^{2}$ | d | $p$ | BHcorrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Adjusted <br> Mean or <br> Mean \% | SE | Adjusted <br> Mean or <br> Mean \% | SE |  |  |  |  |  |  |
| Scale Score | 363.16 | 4.37 | 374.39 | 6.61 | 1.60 | 1, 19 | . 078 | 0.21 | . 222 | . 017 |
| \% Advanced | 27.10 | 2.25 | 32.41 | 3.42 | 1.32 | 1, 18 | . 068 | 0.54 | . 266 | . 050 |
| \% Proficient or Advanced | 56.50 | 2.01 | 61.35 | 3.05 | 1.39 | 1, 18 | . 072 | 0.55 | . 254 | . 033 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 8. ToT Analysis - Differences on 2011 CST Math in Grade 3

| Outcome | Comparison <br> Group ( $n=39$ ) |  | $\frac{\text { Treatment Group }}{(n=20)}$ |  | F | df | Partial$\eta^{2}$ | d | $p$ | BHcorrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Adjusted <br> Mean or <br> Mean \% | SE | Adjusted <br> Mean or <br> Mean \% | SE |  |  |  |  |  |  |
| Scale Score | 386.79 | 2.94 | 388.88 | 4.19 | 0.16 | 1,49 | . 003 | 0.05 | . 692 | . 033 |
| \% Advanced | 35.70 | 1.32 | 35.34 | 1.88 | 0.02 | 1,48 | . 000 | -0.04 | . 883 | . 050 |
| \% Proficient or Advanced | 61.97 | 1.56 | 64.07 | 2.17 | 0.62 | 1,48 | . 013 | 0.22 | . 437 | . 017 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 9. ToT Analysis - Differences on 2011 CST Math in Grade 4

| Outcome | Comparison <br> Group ( $\mathrm{n}=17$ ) |  | Treatment Group$(\mathrm{n}=9)$ |  | F | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Adjusted Mean or Mean \% | SE | Adjusted Mean or Mean \% | SE |  |  |  |  |  |  |
| Scale Score | 382.21 | 4.21 | 380.53 | 6.00 | 0.05 | 1,16 | . 003 | -0.03 | . 831 | . 050 |
| \% Advanced | 38.55 | 2.25 | 40.07 | 3.21 | 0.13 | 1, 15 | . 009 | 0.17 | . 719 | . 033 |
| \% Proficient or Advanced | 66.17 | 2.36 | 63.90 | 3.35 | 0.28 | 1,15 | . 018 | -0.24 | . 604 | . 017 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 10. ToT Analysis - Differences on 2011 CST Math in Grade 5

| Outcome | Comparison <br> Group ( $\mathrm{n}=16$ ) |  | $\frac{\text { Treatment Group }}{(\mathrm{n}=8)}$ |  | F | df | $\begin{gathered} \text { Partial } \\ \mathbf{\eta}^{2} \end{gathered}$ | d | $p$ | ```BH corrected critical value``` |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Adjusted Mean or Mean \% | SE | Adjusted <br> Mean or <br> Mean \% | SE |  |  |  |  |  |  |
| Scale Score | 371.00 | 4.78 | 391.65 | 7.24 | 4.81 | 1,14 | . 256 | 0.34 | .046* | .033 ${ }^{+}$ |
| \% Advanced | 25.07 | 1.97 | 33.23 | 2.99 | 4.39 | 1,13 | . 253 | 1.06 | . 056 | . 050 |
| \% Proficient or Advanced | 55.31 | 2.70 | 61.89 | 3.82 | 6.00 | 1,13 | . 316 | 0.66 | .029* | .017† |

* statistically significant at $p$-value $<0.05$, two-tailed test
$\dagger$ statistically significant at $<\mathrm{BH}$ critical value correcting for the false discovery rate under multiple testing
Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 11. ToT Analysis - Effect Sizes of 2011 CST Math Outcomes Across Grade Levels

| CST Math 2011 Outcome | Mean <br> Effect Size | $95 \%$ Cl | Z-score | BH- |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Borrected <br> critical <br> value |  |  |  |  |  |
| Mean Scale Score | 0.12 | $-0.23,0.44$ | 0.66 | .512 | .050 |
| Mean Percent Advanced | 0.29 | $-0.06,0.57$ | 1.62 | .105 | .017 |
| Mean Percent Proficient or Advanced | 0.27 | $-0.08,0.56$ | 1.51 | .131 | .033 |

# Analyses Using Comparison Schools Selected by WestEd 

## METHOD

The current study utilized a matched-comparison quasi-experimental design that matched grades that were provided with the ST Math program in LAUSD to grades in LAUSD that were not provided with ST Math. WestEd examined six outcomes of interest in its analyses: (1) grade-level 2011 CST Math mean scale scores; (2) the proportion of students in each grade who were Advanced in math; (3) the proportion of students in each grade who were either Proficient or Advanced in math; (4) grade-level 2011 CST Reading mean scale scores; (2) the proportion of students in each grade who were Advanced in reading; (3) the proportion of students in each grade who were either Proficient or Advanced in reading. To examine the effect of ST Math, separate ITT and ToT analyses were conducted, both using ANCOVA.

## IDENTIFICATION OF COMPARISON GRADES

WestEd used a matching procedure to identify comparison grades. The purpose of matching is to create two groups that are essentially equal on the observable variables known to be related to the outcome of interest. 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, 2001).

The treatment schools in the current evaluation were selected to participate by MIND in its "math initiative" and had not been provided with the ST Math program prior to the 2010-2011 school year. A total of 45 schools met these criteria and were provided with the ST Math program beginning in 2010-2011. Most schools had multiple grades participate ( $\mathrm{n}=43$ ). Two schools had only one grade participate. Twenty-nine schools had ST Math available at 2 grades, 8 schools had ST Math in 3 grades, and 6 schools had ST Math in 4 grades. ${ }^{4}$

All schools in LAUSD qualified for the pool of potential comparison grades as long as they had not previously been provided with the ST Math program. Also, because the current evaluation included only grades $2-5$, only schools that included grades $2-5$ were included in the eligible pool of LAUSD

[^3]comparison schools. Treatment grades were matched using both school- and grade-level factors known to be related to academic achievement. The grade-level factors included 2010 CST Math and 2010 CST Reading mean scale scores. The school-level factors included the percentage of Latino students, percentage of African American students, percentage of White students, percentage of students eligible for free or reduced price lunch, percentage of English language learner students, and student enrollment. Several different types of matching strategies exist (Guo \& Fraser, 2010). Propensity score matching is one well-known technique. However, propensity score matching requires a larger sample size than in the current study (Luellen, Shadish, \& Clark, 2005). ${ }^{5}$ As an alternative, WestEd used Mahalanobis distance matching to identify comparison grades (Stuart, 2009). Using the Stata macro "mahascores" and a greedy matching technique, WestEd identified a group of comparison grades within 45 schools that were matched to treatment grades. Exhibit 12 shows the sample sizes for the treatment and comparison grades used in the ITT and ToT analyses.

Exhibit 12. Treatment Group Sample Sizes

| Grade | Intent To Treat (ITT) |  |
| :---: | :---: | :---: |
| 2 | 30 | Treatment on Treated (TOT) ${ }^{\text {a }}$ |
| 2 | 38 | 10 |
| 4 | 26 | 20 |
| 5 | 14 | 8 |

$\mathrm{a}_{\text {sample sizes are equal for treatment and comparison grades }}$
To examine reliability of the matching technique, treatment and comparison grades were compared on the matching variables. The comparison and treatment groups did not significantly differ on matching characteristics. ${ }^{6}$ Exhibits in Appendix C show the results of the t-tests for each grade level.

## DATA AND ANALYSES

ANCOVA models were used to examine the effects of ST Math on 2011 CST Math mean scale scores, proportion of students at the Advanced level on the 2011 CST Math assessment, and the proportion of students at the Proficient or Advanced levels on the 2011 CST Math assessment. English Language Arts outcomes were included to address history effects. History effects occur when something other than the intervention occurs at the same time that affects the outcome. For example, if a school-wide program focusing on improving test scores occurs at the same time as the implementation of ST Math, it is impossible to determine whether the changes in outcomes are due to the school-wide reform or to the ST Math program. One would not expect to see improvements

[^4]in ELA outcomes due to a mathematics-related intervention. So, if improvements in mathematics scores are found, but are not found for ELA scores, it lends validity to the effect of the ST Math program. As such, WestEd also examined the effects of ST Math on 2011 CST Reading mean scale scores, proportion of students at the Advanced level on the 2011 CST Reading assessment, and the proportion of students at the Proficient or Advanced levels on the 2011 CST Reading assessment.

The ANCOVA models included as covariates the variables used in the matching procedure. Schoollevel percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students were included as school-level covariates in all the analyses as was school-level student enrollment size. Grade-level 2010 CST-math mean scale scores and grade-level 2010 CST Reading mean scale scores were included as covariate in all ANCOVAs. For the proportion of students at the Proficient or Advanced levels and the proportion of students at the Advanced levels outcomes, the corresponding baseline estimate in the corresponding content area (i.e., reading or mathematics) was included in the ANCOVA model. The calculation of the standardized mean difference (d) was calculated using the grade-level standard deviation for the 2010-2011 mathematics score, corrected for clustering (WWC, 2011).

WestEd conducted two series of analyses: ITT, which included all grades that were provided with ST Math, and ToT, which included only grades where at least 85 percent of students were enrolled in ST Math and at least 50 percent of the material was covered. Because the outcome analyses are at the grade-level, when a treatment school was excluded from a particular grade level analysis (i.e., ST Math was not offered for that grade within the school), the corresponding matched-comparison school was also excluded from the analysis.

In order to calculate the effect size for CST mean scale scores, WestEd used the grade-level standard deviations for the CST Math exam in 2011 in LAUSD and divided the standard deviations by the square root of 0.20 . The denominator in this equation includes the intraclass correlation (ICC) of .20, which accounts for clustering of students within grades and is an accepted norm (What Works Clearinghouse Handbook 2.1, 2011). This provides a student-level estimate. The use of the studentlevel standard deviations provides a more accurate estimate of the effect size than grade- or schoollevel standards deviations (Lipsey et al., 2012), and the effect sizes calculated using the student-level standard deviations were comparable to effect sizes from other electronic mathematics programs (Slavin \& Lake, 2008; Cheung \& Slavin 2011). School-level standard deviations were used in the calculation of the effect sizes for the other two outcomes: the proportion of students scoring at the Advanced level in math and the proportion of students scoring Proficient or Advanced in math. This was because student- or grade-level standard deviations were not available for these outcomes.

As previously described, to pool effect sizes across grades a meta-analysis was conducted using Lipsey and Wilson's (2001) guidelines for calculating the average effect size using standardized mean
differences weighted by the inverse variance. It was calculated using Lipsey and Wilson's (2001) SPSS macro MEANES. 7

Significance testing (whether the effect significantly differs from zero) was conducted using z-scores and associated p-values. WestEd examined multiple outcomes in each grade and across grades. As the number of outcome comparisons increases, the likelihood of Type-I error increases (i.e., identifying a statistically significant difference when one does not actually exist). To address this issue, WestEd used the Benjamini-Hochberg $(\mathrm{BH})$ correction for comparison of each group of three outcomes within a grade-level (Benjamini \& Hochberg, 1995; Schochet, 2008). ${ }^{8}$ The results of the analyses are presented both with and without the BH corrections.

## RESULTS FOR THE ITT ANALYSES

## MATH OUTCOMES

The results for the ITT analyses of 2011 CST Math outcomes, by grade level, appear in Exhibits 13 through 16. The analyses revealed that, after controlling for grade-level baseline achievement scores and school-level demographic variables, fourth grades that were provided the ST Math program had significantly higher 2011 CST Math mean scale scores compared to fourth grades that were not provided the ST Math program. This finding was no longer statistically significant when correcting for multiple comparisons. No other math outcomes in the fourth grade and no math outcomes in grades 2,3 , or 5 were statistically significant.

The meta-analysis of the combined grade-level results for each outcome (Exhibit 17) revealed statistically significant differences for the proportion of students who scored at the Advanced level and for the proportion of students who scored at either the Proficient or Advanced level. These differences were statistically significant both with and without the correction for multiple comparisons. No statistically significant difference was found between groups on 2011 CST Math mean scale scores.

[^5]Exhibit 13. ITT Analysis - Differences on 2011 CST Math in Grade 2

| Outcome | Adjusted Mean or Mean \% |  | SE | F | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\begin{aligned} & \text { Comparison } \\ & \text { Group } \\ & (n=30) \\ & \hline \end{aligned}$ | $\begin{gathered} \text { Treatment } \\ \text { Group } \\ (n=30) \end{gathered}$ |  |  |  |  |  |  |  |
| Scale Score | 360.23 | 367.05 | 3.98 | 1.39 | 1, 50 | . 027 | 0.11 | . 245 | . 050 |
| \% Advanced | 25.25 | 29.32 | 1.78 | 2.44 | 1,49 | . 047 | 0.42 | . 125 | . 033 |
| \% Proficient or Advanced | 53.54 | 58.66 | 1.95 | 3.24 | 1,49 | . 062 | 0.49 | . 078 | . 017 |

Notes: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 14. ITT Analysis - Differences on 2011 CST Math in Grade 3

| Outcome | Adjusted Mean or Mean \% |  | SE | F | df | Partial $\eta^{2}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group (n=38) | $\begin{aligned} & \hline \text { Treatment } \\ & \text { Group } \\ & \text { (n=38) } \end{aligned}$ |  |  |  |  |  |  |  |
| Scale Score | 384.91 | 387.33 | 2.79 | 0.36 | 1,66 | . 005 | 0.05 | . 551 | . 033 |
| \% Advanced | 34.50 | 34.58 | 1.30 | 0.00 | 1, 65 | . 000 | 0.01 | . 967 | . 050 |
| \% Proficient or Advanced | 60.34 | 63.21 | 1.63 | 0.39 | 1, 65 | . 006 | 0.29 | . 536 | . 017 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 15. ITT Analysis - Differences on 2011 CST Math in Grade 4

| Outcome | Adjusted Mean or Mean \% |  | SE | F | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH- corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\begin{gathered} \text { Comparison } \\ \text { Group } \\ (\mathrm{n}=26) \\ \hline \end{gathered}$ | Treatment Group ( $\mathrm{n}=26$ ) |  |  |  |  |  |  |  |
| Scale Score | 374.68 | 385.08 | 3.26 | 4.86 | 1, 42 | . 104 | 0.19 | .033* | . 017 |
| \% Advanced | 37.03 | 42.09 | 1.75 | 3.99 | 1, 41 | . 089 | 0.58 | . 053 | . 033 |
| \% Proficient or Advanced | 61.85 | 66.31 | 1.83 | 2.81 | 1,41 | . 064 | 0.49 | . 101 | . 050 |

* statistically significant at $p$-value $<0.05$, two-tailed test

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 16. ITT Analysis - Differences on 2011 CST Math in Grade 5

| Outcome | Adjusted Mean or Mean \% |  | SE | F | df | Partial$\eta^{2}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group ( $\mathrm{n}=14$ ) | Treatment Group ( $\mathrm{n}=14$ ) |  |  |  |  |  |  |  |
| Scale Score | 363.31 | 380.75 | 6.43 | 3.49 | 1,18 | . 162 | 0.27 | . 078 | . 017 |
| \% Advanced | 22.69 | 29.38 | 2.51 | 3.32 | 1, 17 | . 163 | 0.74 | . 086 | . 033 |
| \% Proficient or Advanced | 50.37 | 58.92 | 3.61 | 2.62 | 1,17 | . 133 | 0.66 | . 124 | . 050 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 17. IT'T Analysis - Effect Sizes of 2011 CST Math Outcomes Across Grade Levels

| CST Math 2011 Outcome | Mean <br> Effect Size | $95 \% \mathrm{CI}$ | Z-score | $p$ | BH-corrected <br> critical value |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Mean Scale Score | 0.13 | $-0.14,0.38$ | 0.94 | .345 | .050 |
| Mean Percent Advanced | 0.33 | $0.07,0.55$ | 2.47 | $.014^{*}$ | $.033^{\dagger}$ |
| Mean Percent Proficient or Advanced | 0.41 | $0.17,0.61$ | 3.19 | $.001^{*}$ | $.017+$ |

* statistically significant at $p$-value $<0.05$, two-tailed test
$\dagger$ statistically significant at $<\mathrm{BH}$ critical value correcting for the false discovery rate under multiple testing


## READING OUTCOMES

The results for the ITT analyses of 2011 CST Reading outcomes, by grade level, appear in Exhibits 18 through 21. The analyses revealed that, after controlling for grade-level baseline achievement scores and school-level demographic variables, fifth grades that were provided the ST Math program had a significantly greater percentage of students at the Proficient or Advanced levels of reading compared to fifth grades that were not provided the ST Math program. This finding remained statistically significant when correcting for multiple comparisons. No other reading outcomes in the fifth grade and no reading outcomes in grades 2,3 , or 4 were statistically significant. When examining the effect of ST Math on reading across grades using the meta-analytic technique, none of the mean effect sizes were statistically significant (Exhibit 22).

Exhibit 18. ITT Analysis - Differences on 2011 CST Reading in Grade 2

| Outcome | Adjusted Mean or Mean \% |  | SE | $F$ | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group $(\mathrm{n}=30)$ | $\begin{aligned} & \text { Treatment } \\ & \text { Group } \\ & (\mathrm{n}=30) \end{aligned}$ |  |  |  |  |  |  |  |
| Mean Scale Score | 341.75 | 343.02 | 2.46 | 0.13 | 1,50 | . 002 | 0.03 | . 725 | . 033 |
| \% Advanced | 17.03 | 18.47 | 1.30 | 0.57 | 1,49 | . 012 | 0.21 | . 453 | . 017 |
| \% Proficient or <br> Advanced | 44.78 | 45.05 | 1.81 | 0.10 | 1,49 | . 000 | 0.03 | . 920 | . 050 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Reading. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Reading.

Exhibit 19. ITT Analysis - Differences on 2011 CST Reading in Grade 3

| Outcome | Adjusted Mean or Mean \% |  | SE | $F$ | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group ( $\mathrm{n}=38$ ) | Treatment Group ( $\mathrm{n}=38$ ) |  |  |  |  |  |  |  |
| Mean Scale Score | 323.07 | 322.87 | 1.66 | 0.01 | 1,66 | . 000 | -0.01 | . 932 | . 050 |
| \% Advanced | 8.68 | 7.63 | 0.65 | 1.23 | 1,65 | . 019 | -0.27 | . 272 | . 033 |
| \% Proficient or Advanced | 31.68 | 30.49 | 1.44 | 1.29 | 1,65 | . 019 | -0.14 | . 261 | . 017 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Reading. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Reading.

Exhibit 20. ITT Analysis - Differences on 2011 CST Reading in Grade 4

| Outcome | Adjusted Mean or Mean \% |  | SE | $F$ | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group ( $\mathrm{n}=26$ ) | Treatment Group ( $\mathrm{n}=26$ ) |  |  |  |  |  |  |  |
| Mean Scale Score | 349.61 | 350.95 | 1.89 | 0.24 | 1,42 | . 006 | 0.04 | . 627 | . 033 |
| \% Advanced | 21.32 | 24.41 | 1.12 | 3.49 | 1,41 | . 078 | 0.55 | . 069 | . 017 |
| \% Proficient or Advanced | 51.18 | 50.25 | 1.93 | 0.01 | 1,41 | . 000 | -0.09 | . 906 | . 050 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Reading. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Reading.

Exhibit 21. ITT Analysis - Differences on 2011 CST Reading in Grade 5

| Outcome | Adjusted Mean or Mean \% |  | SE | F | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group ( $\mathrm{n}=14$ ) | $\begin{aligned} & \hline \text { Treatment } \\ & \text { Group } \\ & (\mathrm{n}=14) \end{aligned}$ |  |  |  |  |  |  |  |
| Mean Scale Score | 337.88 | 342.65 | 2.61 | 1.59 | 1,18 | . 081 | 0.16 | . 224 | . 033 |
| \% Advanced | 16.08 | 18.06 | 1.76 | 0.60 | 1,17 | . 034 | 0.31 | . 451 | . 050 |
| \% Proficient or Advanced | 39.89 | 47.40 | 1.76 | 7.60 | 1,17 | . 309 | 1.18 | .013* | .017† |

* statistically significant at $p$-value $<0.05$, two-tailed test
$\dagger$ statistically significant at $<\mathrm{BH}$ critical value correcting for the false discovery rate under multiple testing
Notes: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Reading. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Reading.

Exhibit 22. ITT Analysis - Effect Sizes of 2011 CST Reading Outcomes Across Grade Levels

| CST Reading 2011 Outcome | Mean <br> Effect Size | $95 \%$ Cl | Z-score | p-value | BH-corrected <br> critical value |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Mean Scale Score | 0.04 | $-0.23,0.30$ | 0.25 | .800 | .050 |
| Mean Percent Advanced | 0.12 | $-0.15,0.37$ | 0.85 | .395 | .017 |
| Mean Percent Proficient or Advanced | 0.07 | $-0.20,0.33$ | 0.53 | .597 | .033 |

## RESULTS FOR THE TOT ANALYSES

## MATH OUTCOMES

The ToT analyses examined math outcomes for grades that were provided with the ST Math software program, and where at least 85 percent of the students logged into ST Math and completed at least 50 percent of the material. Grades that were provided with the ST Math program but that did not meet these conditions were excluded from the ToT analyses along with their matched grades. The results for the ANCOVAs of 2011 CST Math outcomes, by grade level, using the ToT sample appear in Exhibits 23 through 26. After controlling for baseline grade-level achievement scores and school-level demographic factors, fourth grades that were provided the ST Math program had a significantly greater percentage of students at the Advanced level compared to fourth grades that were not provided the ST Math program. This difference was statistically significant with and without the correction for multiple comparisons. No other statistically significant differences were found in grade 4 and no statistically significant differences were found for grades 2,3 , or 5 .

The meta-analysis of the combined grade-level results for each outcome (Exhibit 26) revealed no statistically significant differences between groups on 2011 CST Math mean scale scores or the proportion of students who scored at the Advanced level. A significant effect size was found for the
proportion of students who scored at either the Proficient or Advanced level. However, the effect was not statistically significant after applying the correction for multiple comparisons.

Exhibit 23. ToT Analysis - Differences on 2011 CST Math in Grade 2

| Outcome | Adjusted Mean or Mean \% |  |  | $F$ | df | Partial $\eta^{2}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group ( $\mathrm{n}=10$ ) | $\begin{gathered} \text { Treatment } \\ \text { Group } \\ (\mathrm{n}=10) \end{gathered}$ | SE |  |  |  |  |  |  |
| Mean Scale Score | 363.19 | 381.13 | 9.21 | 1.50 | 1,10 | . 131 | 0.27 | . 248 | . 033 |
| \% Advanced | 27.70 | 35.90 | 4.51 | 1.29 | 1, 9 | . 125 | 0.61 | . 286 | . 050 |
| \% Proficient or Advanced | 54.26 | 64.14 | 4.33 | 2.03 | 1,9 | . 184 | 0.76 | . 188 | . 017 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 24. ToT Analysis - Differences on 2011 CST Math in Grade 3

| Outcome | Adjusted Mean or Mean \% |  | SE | F | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group ( $\mathrm{n}=20$ ) | $\begin{gathered} \text { Treatment } \\ \text { Group } \\ (\mathrm{n}=20) \end{gathered}$ |  |  |  |  |  |  |  |
| Mean Scale Score | 388.07 | 391.29 | 2.28 | 0.93 | 1,30 | . 030 | 0.09 | . 342 | . 017 |
| \% Advanced | 37.04 | 36.41 | 1.28 | 0.12 | 1,29 | . 004 | -0.11 | . 736 | . 050 |
| \% Proficient or Advanced | 63.99 | 65.16 | 1.43 | 0.32 | 1,29 | . 011 | 0.19 | . 577 | . 033 |

[^6]Exhibit 25. ToT Analysis - Differences on 2011 CST Math in Grade 4

| Outcome | Adjusted Mean or Mean \% |  | SE | $F$ | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group $(\mathrm{n}=9)$ | Treatment Group ( $\mathrm{n}=9$ ) |  |  |  |  |  |  |  |
| Mean Scale Score | 378.06 | 392.40 | 4.54 | 4.41 | 1, 8 | . 356 | 0.22 | . 069 | . 033 |
| \% Advanced | 39.45 | 43.88 | 0.89 | 10.61 | 1,7 | . 603 | 1.76 | .014* | .017† |
| \% Proficient or Advanced | 62.81 | 68.74 | 3.13 | 1.59 | 1,7 | . 185 | 0.67 | . 248 | . 050 |

* statistically significant at $p$-value $<0.05$, two-tailed test
$\dagger$ statistically significant at $<\mathrm{BH}$ critical value correcting for the false discovery rate under multiple testing
Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 26. ToT Analysis - Differences on 2011 CST Math in Grade 5

| Outcome | Adjusted Mean or Mean \% |  | SE | $F$ | df | $\begin{aligned} & \text { Parti } \\ & \text { al } \eta^{2} \end{aligned}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group $(\mathrm{n}=8)$ | Treatment Group ( $\mathrm{n}=8$ ) |  |  |  |  |  |  |  |
| Mean Scale Score | 372.13 | 388.82 | 9.08 | 1.50 | 1,6 | . 200 | 0.25 | . 267 | . 017 |
| \% Advanced | 25.53 | 32.10 | 3.62 | 1.43 | 1,5 | . 222 | 0.69 | . 286 | . 033 |
| \% Proficient or Advanced | 55.64 | 63.61 | 5.34 | 0.94 | 1,5 | . 158 | 0.56 | . 377 | . 050 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

Exhibit 26. ToT Analysis - Effect Sizes of 2011 CST Math Outcomes Across Grade Levels

| 2011 CST Math Outcome | Mean <br> Effect Size | $95 \% \mathrm{Cl}$ | Z-score | $p$ | BH-corrected <br> critical value |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Mean Scale Score | 0.18 | $-0.22,0.53$ | 0.87 | .384 | .050 |
| Mean Percent Advanced | 0.40 | $-0.01,0.69$ | 1.90 | .058 | .033 |
| Mean Percent Proficient or Advanced | 0.43 | $0.05,0.70$ | 2.19 | $.029 *$ | .017 |

* statistically significant at $p$-value $<0.05$, two-tailed test


## READING OUTCOMES

The results for the ANCOVAs of 2011 CST Reading outcomes, by grade level, using the ToT sample appear in Exhibits 27 through 30. The analyses revealed that, after controlling for grade-level baseline achievement scores and school-level demographic variables, third grades that were provided the ST Math program had a significantly smaller percentage of students at the Proficient or

Advanced levels of reading compared to third grades that were not provided the ST Math program. In addition, after controlling for grade-level baseline achievement scores and school-level demographic variables, fifth grades that were provided the ST Math program had a significantly greater percentage of students at the Proficient or Advanced levels of reading compared to fifth grades that were not provided the ST Math program. Both these finding remained statistically significant when correcting for multiple comparisons. No other reading outcomes in the third or fifth grades were statistically significant and none of the reading outcomes in grades 2 or 4 were statistically significant. When examining the effect of ST Math on reading across grades using the meta-analytic technique, none of the mean effect sizes were statistically significant (Exhibit 31).

Exhibit 27. ToT Analysis - Differences on 2011 CST Reading in Grade 2

| Outcome | Adjusted Mean or Mean \% |  | SE | F | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group ( $\mathrm{n}=10$ ) | Treatment Group ( $\mathrm{n}=10$ ) |  |  |  |  |  |  |  |
| Mean Scale Score | 340.68 | 348.15 | 5.07 | 0.86 | 1,10 | . 079 | 0.20 | . 376 | . 017 |
| \% Advanced | 17.79 | 20.71 | 3.55 | 0.27 | 1,9 | . 029 | 0.27 | . 619 | . 033 |
| \% Proficient or Advanced | 45.91 | 47.29 | 3.73 | 0.05 | 1,9 | . 006 | 0.12 | . 826 | . 050 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Reading. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Reading.

Exhibit 28. ToT Analysis - Differences on 2011 CST Reading in Grade 3

| Outcome | Adjusted Mean or Mean \% |  | SE | $F$ | df | $\begin{aligned} & \text { Parti } \\ & \text { al } \eta^{2} \end{aligned}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group ( $\mathrm{n}=20$ ) | $\begin{aligned} & \hline \text { Treatment } \\ & \text { Group } \\ & (\mathrm{n}=20) \end{aligned}$ |  |  |  |  |  |  |  |
| Mean Scale Score | 325.32 | 322.08 | 1.70 | 1.71 | 1,30 | . 054 | -0.13 | . 202 | . 050 |
| \% Advanced | 8.91 | 6.79 | 0.71 | 4.14 | 1,29 | . 125 | -0.69 | . 051 | . 033 |
| \% Proficient or Advanced | 34.41 | 28.54 | 1.60 | 6.29 | 1,29 | . 178 | -0.84 | .018* | .017† |

* statistically significant at $p$-value $<0.05$, two-tailed test
$\dagger$ statistically significant at $<\mathrm{BH}$ critical value correcting for the false discovery rate under multiple testing
Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Reading. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Reading.

Exhibit 29. ToT Analysis - Differences on 2011 CST Reading in Grade 4

| Outcome | Adjusted Mean or Mean \% |  | SE | $F$ | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group ( $\mathrm{n}=9$ ) | $\begin{gathered} \text { Treatment } \\ \text { Group } \\ (\mathrm{n}=9) \end{gathered}$ |  |  |  |  |  |  |  |
| Mean Scale Score | 357.34 | 357.05 | 3.05 | 0.00 | 1,8 | . 000 | -0.01 | . 952 | . 050 |
| \% Advanced | 25.98 | 27.69 | 1.51 | 0.57 | 1,7 | . 075 | 0.40 | . 475 | . 033 |
| \% Proficient or Advanced | 56.80 | 52.42 | 2.06 | 1.96 | 1,7 | . 219 | -0.75 | . 204 | . 017 |

Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Reading. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Reading.

Exhibit 30. ToT Analysis - Differences on 2011 CST Reading in Grade 5

| Outcome | Adjusted Mean or Mean \% |  | SE | $F$ | df | $\begin{gathered} \text { Partial } \\ \eta^{2} \end{gathered}$ | d | $p$ | BH-corrected critical value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Comparison Group $(\mathrm{n}=8)$ | Treatment Group $(\mathrm{n}=8)$ |  |  |  |  |  |  |  |
| Mean Scale Score | 345.10 | 347.19 | 4.32 | 0.10 | 1,6 | . 017 | 0.07 | . 759 | . 033 |
| \% Advanced | 19.26 | 19.00 | 3.31 | 0.00 | 1,5 | . 001 | -0.03 | . 960 | . 050 |
| \% Proficient or Advanced | 42.86 | 51.89 | 1.22 | 20.91 | 1,5 | . 807 | 2.80 | .006* | .017† |

* statistically significant at $p$-value $<0.05$, two-tailed test
$\dagger$ statistically significant at $<\mathrm{BH}$ critical value correcting for the false discovery rate under multiple testing Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Reading. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Reading.

Exhibit 31. ToT Analysis - Effect Sizes of 2011 CST Reading Outcomes Across Grade Levels

| 2011 CST Reading Outcome | $\begin{array}{c}\text { Mean } \\ \text { Effect Size }\end{array}$ | $95 \%$ CI | Z-score | $\begin{array}{c}\text { B-H }\end{array}$ |
| :--- | :---: | :---: | :---: | :---: | :---: |
| corrected |  |  |  |  |
| critical |  |  |  |  |
| value |  |  |  |  |$]$

## Discussion and Next Steps

WestEd examined the effects of the ST Math software program in LAUSD through two sets of analyses, each using a different comparison group to estimate program effects. One set of analyses used a comparison group selected by MIND while the second set used a matched-comparison group independently selected by WestEd. When using the full set of grades (i.e., a ITT analyses) both sets of analyses revealed statistically significant differences on math outcomes between the grades that were provided with ST Math and grades that were not provided with the program. However, these significant differences were found for different grades depending on which comparison group was used. Using the MIND-selected group, differences were found in grades 2 across all math outcomes and in one math outcome for grade 3. Using the WestEd-selected group, a difference was found on a single math outcome in grade 4 . All the aforementioned differences remained statistically significant after controlling for multiple comparisons.

Both sets of analyses found that the pooled effect of the program across grades was statistically significant on two outcomes: the proportion of students who scored at the Advanced level, and the proportion who scored at the Proficient or Advanced levels. These differences remained statistically significant after controlling for multiple comparisons. When analyzing the effects of ST Math using only grades with high levels of implementation (i.e., a ToT analysis), the sample size was reduced and fewer significant differences were found. No significant differences were found when pooling effects across grades in the ToT analysis.

ELA outcomes were examined to assess history effects, that is, whether positive finding on math outcomes may be due to an event or condition (e.g., another supplemental program) at the treatment schools other than ST Math. Although statistically significant differences were found with several reading outcomes, they were inconsistent both in terms of the grade levels where they occurred and the direction of the effect. For the ITT analyses, one ELA outcome showed a significant positive effect for ST Math in $4^{\text {th }}$ grade. For the ToT analysis, one ELA outcome showed a significant positive effect for ST Math in $5^{\text {th }}$ grade and another showed a significant negative effect for ST Math in 3rd grade. None of the significant findings for ELA outcomes were in the same grades as the significant math outcomes for the corresponding analyses (i.e., ITT or ToT). In addition, there were no significant findings for ELA outcomes when pooling effects across grades. Therefore, it is inconclusive whether the findings for ELA outcomes indicate that the positive findings for ST Math are due to some other factor associated with the treatment schools.

It is likely that a greater number of statistically significant differences would have been found at individual grade levels, particularly for math outcomes, if a larger sample of grades had been included in the analyses. For example, when using the WestEd-selected comparison group, the effect size (d) for a number of math outcomes in grades 2 , 4 , and 5 was over 0.40 with the largest being 0.76. Assuming similar effect sizes as those found in the current study, pooling effects across a larger number of schools (hence, grades) would likely reveal statistically significant differences in math outcomes for more grades.

Future research on the program could be strengthened in several ways. One way is to obtain gradelevel outcome data for more than a single school year. 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 or not program effects across grades increase with each year of exposure. In addition, obtaining student-level math outcomes would further strengthen research on the program. It would allow for a more precise estimate of standard errors and would allow researchers to assess any impacts of the program on individual students over time, either due to multiple years of exposure or to long term effects after exposure ends. Finally, despite the careful matching of treatment and comparison schools on observable characteristics, it is possible that 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 the groups differed on other characteristics besides exposure to ST Math impedes any causal conclusion (Shadish, Cook, \& Campbell, 2001). However, a future randomized-control trial of ST Math, that is carefully executed, would allow for such a conclusion.

## References

Akhavan, S. (2012a). Los Angeles Unified Math outcomes analysis 2011: Grades using the ST Math program only in 2010/11. Unpublished manuscript dated November 14, 2012. Santa Ana, CA: MIND Research Institute.

Akhavan, S. (2012b). Los Angeles Unified Math outcomes analysis 2011: Grades using the ST Math program only in 2010/11. Unpublished manuscript dated November 17, 2012. Santa Ana, CA: MIND Research Institute.

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.

Cheung, A.C.K., \& Slavin, R.E. (2011). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education.

Cohen, J., Cohen, P., West, S., \& Aiken, L. (2003). Applied multiple regression/correlation analysis for the behavioral sciences ( $3^{\text {rd }}$ ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Gupta, S.K. (2011). Intention-to-treat concept: A review. Perspectives in Clinical Research, 2, 109-112.
Lipsey, M.W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M.W., Roberts, M., Anthony, K.S., \& Busick, M.D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncser

Lipsey, M.W., \& Wilson, D.B. (2001). Practical meta-analysis. Thousand Oaks, CA: Sage Publications.

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

Rubin, D.B., \& Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. Journal of American Statistical Association, 95, 573-585.

Schochet, P.Z. (2008). Technical methods report: Guidelines for multiple testing in impact evaluations. Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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

Slavin, R.E., \& Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. Review of Educational Research, 78, 427-515. Doi: 10.3102/0034654308317473

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

Tabachnik, B., \& Fidell, L. (2006). Using multivariate statistics ( $5^{\text {th }} \mathrm{ed}$ ). Boston, MA: Allyn and Bacon.

What Works Clearinghouse. (Sept. 2011). Procedures and Standards Handbook v.2.1. Retrieved 4/2/13. http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19

# Appendix A. Baseline Comparisons with Comparison Groups Selected by the MIND Research Institute 

## ITT SAMPLE

Exhibit A1. Treatment and Comparison Groups Grade 2

| Variable | $\frac{\text { Comparison Group }}{(n=55)}$ |  | $\frac{\text { Treatment Group }}{(\mathrm{n}=32)}$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard <br> Deviation | Mean | Standard <br> Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST Reading Scale Scores | 341.49 | 17.56 | 333.43 | 15.26 | -2.17 | .033* | -0.48 |
| 2010 CST Math Scale Scores | 350.59 | 25.17 | 344.67 | 21.38 | -1.12 | . 267 | -0.25 |
| School-Level Variables |  |  |  |  |  |  |  |
| Percentage Latino Students | 77.75 | 24.72 | 87.69 | 12.75 | 2.11 | .037* | 0.47 |
| Percentage African American Students | 12.82 | 21.59 | 9.41 | 13.00 | 0.81 | . 420 | 0.18 |
| Percentage White Students | 4.75 | 12.59 | 1.09 | 2.68 | -1.62 | . 110 | -0.36 |
| Percentage of Students Eligible for FRPL | 86.24 | 19.91 | 97.16 | 6.09 | 3.02 | .003* | 0.67 |
| Percent of English Language Learners | 36.82 | 15.51 | 46.59 | 12.92 | 3.01 | .003* | 0.67 |
| Student Enrollment Size | 419.51 | 191.70 | 493.31 | 190.77 | 1.74 | . 086 | 0.39 |

* statistically significant at $p$-value $<0.05$, two-tailed test

Exhibit A2. Treatment and Comparison Groups Grade 3

| Variable | $\frac{\text { Comparison Group }}{(n=65)}$ |  | $\begin{aligned} & \text { Treatment Group } \\ & (n=40) \end{aligned}$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard <br> Deviation | Mean | Standard <br> Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST Reading Scale Scores | 320.09 | 15.67 | 31532 | 16.23 | -1.49 | . 138 | -0.30 |
| 2010 CST Math Scale Scores | 365.11 | 23.99 | 361.47 | 24.14 | -0.75 | . 453 | 0.09 |

School-Level Variables

| Percentage Latino Students | 77.60 | 23.96 | 86.48 | 11.79 | 2.19 | $.031^{*}$ | 0.44 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Percentage African American <br> Students | 13.12 | 20.16 | 10.00 | 12.24 | -0.88 | .379 | -0.18 |
| Percentage White Students | 5.65 | 13.06 | 1.28 | 2.67 | -2.09 | $.039^{*}$ | -0.42 |
| Percentage of Students | 88.82 | 16.52 | 96.50 | 6.49 | 2.81 | $.006^{*}$ | 0.56 |


| Variable | $\frac{\text { Comparison Group }}{(\mathrm{n}=65)}$ |  | $\frac{\text { Treatment Group }}{(\mathrm{n}=40)}$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Eligible for FRPL |  |  |  |  |  |  |  |
| Percent of English Language Learners | 37.06 | 15.07 | 46.18 | 12.56 | 3.20 | .002* | 0.64 |
| Student Enrollment Size | 471.75 | 228.39 | 483.78 | 183.87 | 0.28 | . 779 | 0.06 |

* statistically significant at $p$-value $<0.05$, two-tailed test

Exhibit A3. Treatment and Comparison Groups Grade 4

| Variable | $\frac{\text { Comparison Group }}{(\mathrm{n}=49)}$ |  | $\frac{\text { Treatment Group }}{(n=26)}$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard Deviation | Mean | Standard <br> Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST Reading Scale Scores | 349.71 | 15.46 | 343.38 | 15.02 | -1.71 | . 093 | -0.41 |
| 2010 CST Math Scale Scores | 368.71 | 19.31 | 366.85 | 19.51 | -0.39 | . 695 | -0.09 |
| School-Level Variables |  |  |  |  |  |  |  |
| Percentage Latino Students | 77.80 | 23.89 | 84.81 | 13.33 | 1.38 | . 171 | 0.33 |
| Percentage African American Students | 11.41 | 20.07 | 11.42 | 13.84 | 0.00 | . 997 | 0.00 |
| Percentage White Students | 6.27 | 12.87 | 1.58 | 3.19 | -1.82 | . 072 | -0.44 |
| Percentage of Students Eligible for FRPL | 86.82 | 14.18 | 95.42 | 7.46 | 2.88 | .005* | 0.70 |
| Percent of English Language Learners | 33.86 | 15.58 | 43.19 | 11.47 | 2.69 | .009* | 0.65 |
| Student Enrollment Size | 422.49 | 174.65 | 477.81 | 176.45 | 1.30 | . 197 | 0.32 |

* statistically significant at $p$-value $<0.05$, two-tailed test

Exhibit A4. Treatment and Comparison Groups Grade 5

| Variable | $\frac{\text { Comparison Group }}{(\mathrm{n}=25)}$ |  | $\frac{\text { Treatment Group }}{(\mathrm{n}=14)}$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard <br> Deviation | Mean | Standard <br> Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST Reading Scale Scores | 338.34 | 11.10 | 332.08 | 11.99 | -1.64 | . 109 | -0.55 |
| 2010 CST Math Scale Scores | 352.44 | 22.39 | 348.98 | 20.95 | -0.47 | . 639 | -0.16 |
| School-Level Variables |  |  |  |  |  |  |  |
| Percentage Latino Students | 82.12 | 19.12 | 89.50 | 9.39 | 1.35 | . 185 | 0.45 |
| Percentage African American Students | 9.20 | 14.99 | 6.29 | 7.49 | -0.68 | . 502 | -0.23 |
| Percentage White Students | 2.60 | 3.43 | 2.21 | 3.95 | -0.32 | . 751 | -0.11 |
| Percentage of Students Eligible for FRPL | 90.80 | 12.39 | 96.93 | 8.12 | 1.66 | . 106 | 0.55 |
| Percent of English Language Learners | 34.76 | 13.26 | 46.00 | 10.30 | 2.74 | .009* | 0.91 |
| Student Enrollment Size | 477.64 | 225.07 | 521.64 | 195.57 | 0.61 | . 544 | 0.20 |

* statistically significant at $p$-value $<0.05$, two-tailed test


## TOT SAMPLE

Exhibit A5. Treatment and Comparison Groups Grade 2

| Variable | $\frac{\text { Comparison Group }}{(\mathrm{n}=19)}$ |  | Treatment Group$(n=10)$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard Deviation | Mean | Standard Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST Reading Scale Scores | 340.95 | 12.38 | 337.85 | 15.93 | -0.58 | . 566 | -0.23 |
| 2010 CST Math Scale Scores | 350.83 | 16.51 | 352.41 | 18.29 | 0.24 | . 815 | 0.09 |
| School-Level Variables |  |  |  |  |  |  |  |
| Percentage Latino Students | 87.84 | 10.61 | 85.00 | 19.93 | -0.56 | . 582 | -0.22 |
| Percentage African American Students | 3.95 | 5.03 | 11.90 | 18.30 | 1.80 | . 084 | 0.70 |
| Percentage White Students | 3.16 | 5.49 | 0.70 | 1.06 | -1.39 | . 176 | -0.54 |
| Percentage of Students Eligible for FRPL | 91.11 | 10.51 | 96.30 | 4.92 | 1.47 | . 153 | 0.57 |
| Percent of English Language Learners | 40.58 | 11.39 | 46.00 | 16.51 | 1.04 | . 306 | 0.41 |
| Student Enrollment Size | 511.42 | 262.67 | 454.60 | 210.55 | -0.59 | . 560 | -0.23 |

Exhibit A6. Treatment and Comparison Groups Grade 3

| Variable | $\frac{\text { Comparison Group }}{(n=39)}$ |  | $\frac{\text { Treatment Group }}{(\mathrm{n}=20)}$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard Deviation | Mean | Standard Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST Reading Scale Scores | 317.38 | 10.99 | 319.06 | 12.65 | 0.53 | . 600 | 0.15 |
| 2010 CST Math Scale Scores | 367.00 | 20.06 | 366.59 | 22.13 | -0.07 | . 943 | -0.02 |
| School-Level Variables |  |  |  |  |  |  |  |
| Percentage Latino Students | 86.90 | 17.84 | 87.40 | 13.57 | 0.11 | . 912 | 0.03 |
| Percentage African American Students | 6.56 | 14.30 | 9.00 | 14.15 | 0.62 | . 537 | 0.17 |
| Percentage White Students | 1.90 | 4.76 | 0.90 | 1.65 | -0.15 | . 368 | -0.04 |
| Percentage of Students Eligible for FRPL | 93.92 | 9.71 | 96.80 | 5.89 | 1.21 | . 230 | 0.33 |
| Percent of English Language Learners | 42.77 | 12.08 | 45.35 | 13.50 | 0.75 | . 458 | 0.21 |
| Student Enrollment Size | 530.49 | 177.02 | 471.25 | 187.71 | 1.19 | . 238 | 0.33 |

Exhibit A7. Treatment and Comparison Groups Grade 4

| Variable | $\frac{\text { Comparison Group }}{(\mathrm{n}=17)}$ |  | $\frac{\text { Treatment Group }}{(\mathrm{n}=9)}$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard Deviation | Mean | Standard Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST Reading Scale Scores | 348.48 | 16.90 | 346.21 | 14.62 | -0.34 | . 736 | -0.14 |
| 2010 CST Math Scale Scores | 363.84 | 20.09 | 366.34 | 20.97 | 0.30 | . 768 | 0.12 |
| School-Level Variables |  |  |  |  |  |  |  |
| Percentage Latino Students | 81.06 | 27.07 | 85.67 | 17.35 | 0.46 | . 649 | 0.19 |
| Percentage African American Students | 10.41 | 21.65 | 7.56 | 17.50 | -0.34 | . 737 | -0.14 |
| Percentage White Students | 5.12 | 14.40 | 3.00 | 4.24 | -0.43 | . 673 | -0.18 |
| Percentage of Students Eligible for FRPL | 88.41 | 22.67 | 94.89 | 7.64 | 0.83 | . 417 | 0.34 |
| Percent of English Language Learners | 36.88 | 17.87 | 42.44 | 13.55 | 0.82 | . 423 | 0.34 |
| Student Enrollment Size | 507.06 | 307.26 | 419.56 | 139.00 | -0.81 | . 428 | -0.34 |

Exhibit A8. Treatment and Comparison Groups Grade 5

| Variable | $\frac{\text { Comparison Group }}{(\mathrm{n}=16)}$ |  | $\frac{\text { Treatment Group }}{(n=8)}$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard Deviation | Mean | Standard <br> Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST Reading Scale Scores | 338.09 | 10.69 | 336.53 | 13.99 | -0.31 | . 763 | -0.13 |
| 2010 CST Math Scale Scores | 350.06 | 21.98 | 349.49 | 21.52 | -0.06 | . 953 | -0.03 |
| School-Level Variables |  |  |  |  |  |  |  |
| Percentage Latino Students | 69.25 | 32.20 | 88.25 | 9.29 | 1.62 | . 120 | 0.70 |
| Percentage African American Students | 19.13 | 25.77 | 4.88 | 5.28 | -1.53 | . 140 | -0.66 |
| Percentage White Students | 6.63 | 11.60 | 3.88 | 4.64 | -0.64 | . 529 | -0.28 |
| Percentage of Students Eligible for FRPL | 83.44 | 22.34 | 94.63 | 10.41 | 1.34 | . 196 | 0.58 |
| Percent of English Language Learners | 30.06 | 17.33 | 43.63 | 11.25 | 2.00 | . 058 | 0.87 |
| Student Enrollment Size | 542.94 | 344.92 | 424.25 | 136.64 | -0.93 | . 363 | -0.40 |

## Appendix B. Example of the Meta-analytic Technique Used to Pool Effect Sizes Across Grades

The following is an example of the meta-analytic technique described in the body of the report as it was applied to the analysis of combined effect sizes across grades for the ITT analysis using the comparison group selected by the MIND Research Institute.

The meta-analysis combined the effect sizes across grades for each outcome, using the weighted effect sizes. Following equations 1 and 2 in the body of the report, the following equation was used to calculate each inverse variance for each grade-level effect (grade 2 used as example):

$$
\begin{equation*}
\omega=\frac{1}{S E^{2}}=\frac{1}{.22^{2}}=20 \tag{4}
\end{equation*}
$$

where $\omega$ is the inverse variance weight and $S E$ is equal to:

$$
\begin{equation*}
S E=\sqrt{\frac{n_{G 1}+n_{G 2}}{n_{G 1} n_{G 2}}+\frac{E S^{2}}{2\left(n_{G 1}+n_{G 2}\right)}}=\sqrt{\frac{55+32}{55 * 32}+\frac{.25^{2}}{2(55+32)}}=\sqrt{.05}=.22 \tag{5}
\end{equation*}
$$

An inverse variance weight was calculated for each grade, for each outcome using formulas 1 and 2. Once each inverse variance weight was calculated, the effect size estimates were multiplied by their respective inverse variance weight and then divided by the sum of the inverse variance weights (equation 3) for each outcome.

# Appendix C. Baseline Comparisons Using Comparison Groups Selected by WestEd 

Exhibit C1. Treatment and Comparison Groups - Grade 2

| Variable | $\frac{\text { Comparison Group }}{(\mathrm{n}=30)}$ |  | $\frac{\text { Treatment Group }}{(n=30)}$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard <br> Deviation | Mean | Standard <br> Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST Reading Scale Scores | 339.27 | 14.10 | 333.85 | 15.62 | -1.41 | . 164 | -0.36 |
| 2010 CST Math Scale Scores | 350.97 | 21.50 | 346.10 | 20.65 | -0.89 | . 375 | -0.23 |
| School-Level Variables |  |  |  |  |  |  |  |
| Percentage Latino Students | 87.23 | 11.46 | 88.17 | 12.69 | 0.30 | . 766 | 0.08 |
| Percentage African American Students | 7.67 | 8.98 | 8.93 | 12.82 | 0.44 | . 659 | 0.11 |
| Percentage White Students | 2.60 | 5.41 | 1.10 | 2.76 | -1.35 | . 181 | -0.35 |
| Percentage of Students Eligible for FRPL | 94.93 | 8.22 | 97.30 | 6.13 | 1.26 | . 211 | 0.33 |
| Percent of English Language Learners | 43.77 | 10.34 | 46.97 | 13.06 | 1.05 | . 297 | 0.27 |
| Student Enrollment Size | 502.07 | 181.56 | 500.10 | 195.26 | -0.04 | . 968 | -0.01 |

Note: Degrees of freedom for all tests was 62.
Exhibit C2. Treatment and Comparison Groups - Grade 3

| Variable | Comparison Group ( $\mathrm{n}=38$ ) |  | Treatment Group ( $\mathrm{n}=38$ ) |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard Deviation | Mean | Standard <br> Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST Reading Scale Scores | 319.92 | 14.97 | 316.08 | 15.99 | -1.08 | . 282 | -0.25 |
| 2010 CST Math Scale Scores | 369.27 | 23.03 | 362.55 | 23.08 | -1.27 | . 208 | -0.29 |
| School-Level Variables |  |  |  |  |  |  |  |
| Percentage Latino Students | 86.11 | 1136 | 86.79 | 11.72 | 0.26 | . 797 | 0.06 |
| Percentage African American Students | 7.82 | 9.35 | 9.66 | 12.07 | 0.74 | . 459 | 0.11 |
| Percentage White Students | 3.08 | 5.39 | 1.29 | 2.73 | -1.83 | . 072 | -0.42 |
| Percentage of Students Eligible for FRL | 94.50 | 7.69 | 96.58 | 6.55 | 1.27 | . 208 | 0.29 |
| Percent of English Language Learners | 43.68 | 10.09 | 46.45 | 12.66 | 1.05 | . 286 | 0.24 |
| Student Enrollment | 493.68 | 16904 | 488.63 | 187.46 | -0.12 | . 902 | -0.03 |

Note: Degrees of freedom for all tests was 78.

Exhibit C3. Treatment and Comparison Groups - Grade 4

| Variable | $\frac{\text { Comparison Group }}{(n=26)}$ |  | Treatment Group$\text { ( } \mathrm{n}=26 \text { ) }$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard <br> Deviation | Mean | Standard <br> Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST Reading Scale Scores | 344.07 | 17.55 | 343.38 | 15.02 | -0.15 | . 880 | -0.04 |
| 2010 CST Math Scale Scores | 367.77 | 25.10 | 366.85 | 19.51 | -0.15 | . 884 | -0.04 |
| School-Level Variables |  |  |  |  |  |  |  |
| Percentage Latino Students | 84.58 | 14.41 | 84.81 | 13.33 | 0.06 | . 952 | 0.02 |
| Percentage African American Students | 8.19 | 9.90 | 11.42 | 13.84 | 0.97 | . 337 | 0.27 |
| Percentage White Students | 3.88 | 7.56 | 1.58 | 3.19 | -1.43 | . 158 | -0.40 |
| Percentage of Students Eligible for FRL | 93.58 | 11.17 | 95.42 | 7.46 | 0.70 | . 487 | 0.19 |
| Percent of English Language Learners | 42.19 | 11.27 | 43.19 | 11.47 | 0.32 | . 753 | 0.09 |
| Student Enrollment | 496.54 | 162.69 | 477.81 | 176.45 | -0.40 | . 692 | -0.11 |

Note: Degrees of freedom for all tests was 50.
Exhibit C4. Treatment and Comparison Groups - Grade 5

| Variable | $\begin{gathered} \text { Comparison Group } \\ (n=14) \end{gathered}$ |  | $\frac{\text { Treatment Group }}{(n=14)}$ |  | $t$ | $p$ | d |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | Standard <br> Deviation | Mean | Standard Deviation |  |  |  |
| Grade-Level Variables |  |  |  |  |  |  |  |
| 2010 CST-Reading Scale Scores | 337.26 | 11.56 | 332.08 | 11.99 | -1.17 | . 255 | -0.44 |
| 2010 CST-Math Scale Scores | 358.64 | 21.25 | 348.98 | 20.95 | -1.21 | . 237 | -0.46 |
| School-Level Variables |  |  |  |  |  |  |  |
| Percentage Latino Students | 85.36 | 15.68 | 89.50 | 9.39 | 0.85 | . 404 | 0.32 |
| Percentage African American Students | 5.57 | 7.26 | 6.29 | 7.49 | 0.26 | . 800 | 0.10 |
| Percentage White Students | 5.36 | 8.41 | 2.21 | 3.95 | -1.27 | . 217 | -0.48 |
| Percentage of Students Eligible for FRL | 92.64 | 13.07 | 96.93 | 8.12 | 1.04 | . 307 | 0.39 |
| Percent of English Language Learners | 44.29 | 11.04 | 46.00 | 10.30 | 0.43 | . 674 | 0.16 |
| Student Enrollment | 523.50 | 178.98 | 521.64 | 195.57 | -0.03 | . 979 | -0.01 |

Note: Degrees of freedom for all tests was 26.


[^0]:    ${ }^{1}$ The procedures for selecting the comparison grades are described in detail in Akhavan (2012a, b).

[^1]:    ${ }^{2}$ There was some nesting of grades within schools because some schools implemented ST Math at multiple grades, whereas other schools implemented ST Math in only a single grade. With nested data such as these, typically hierarchical linear modeling (HLM; Cohen, Cohen, Aiken \& West, 2001) is employed to account for the clustering of grades within schools and to prevent inflation of standard error estimates. However, there were too few schools where ST Math was implemented in multiple grades to use HLM, or to otherwise adjust for clustering of grades within schools in the meta-analysis.

[^2]:    ${ }^{3}$ The Benjamini-Hochberg correction was chosen because the power losses are smaller compared to other multiple comparison correction procedures (Schochet, 2008).

[^3]:    ${ }^{4}$ For the treatment-on-treated analysis, 10 schools implemented ST Math at 1 grade, 17 schools implemented ST Math at 2 grades, and 1 school implemented ST Math at 3 grades. No schools implemented ST Math at all 4 grades.

[^4]:    ${ }^{5}$ Propensity score matching was originally utilized in the analysis, but the resulting matches were unacceptable (Guo \& Fraser, 2010).
    ${ }^{6}$ There were a number of statistically significant differences between groups on these same factors for the MIND-selected comparison group. This indicates that the WestEd-selected comparison group was, on the whole, a better match than the one selected by MIND.

[^5]:    ${ }^{7}$ This technique is discussed in detail in the section of this report describing the analyses using the comparison group selected by MIND. An example of how the technique was employed using data from this study can be found in Appendix B.
    ${ }^{8}$ The Benjamini-Hochberg correction was chosen because the power losses are smaller compared to other multiple comparison correction procedures (Schochet, 2008).

[^6]:    Note: All outcomes adjusted for grade-level 2010 CST Math mean scale scores and 2010 CST Reading mean scale scores. All outcomes adjusted for school-level percentages of Latino students, African American students, White students, students eligible for free or reduced price lunch, and English language learner students, and school-level student enrollment size. Mean percent Advanced also adjusted for grade-level proportions of students who scored at the Advanced level on the 2010 CST Math. Mean percent Proficient or Advanced also adjusted for grade-level proportions of students who scored at the Proficient or Advanced level on the 2010 CST Math.

