Arkansas Math Outcomes Analysis 2016/17

Grade Levels: 3, 4, 5
ST Math Program: Gen-4
Analysis Type: Multi-Year
Treatment-Years: 2016/17
Baseline-Year: 2014/15 or 2015/16
Subgroup: All

Jessica Guise
MIND Research Institute

Abstract

This analysis covers all grades using ST Math in Arkansas in 2016/17. It identifies those grades with nominal or better implementation of the ST Math program, and matches them to randomly selected, similar math-performance, comparison grades. The nominal ST Math users are an aggregation of 9 grades, consisting of grades 3, 4, and 5 at 5 schools, with an average baseline z-score of 0.39 (refer to Figures 2 and 3 to see how your schools compare to those analyzed in this report). They were matched to 9 similar, randomly selected control grades at 8 schools that never used ST Math. Grade-wise growth in z-score of math proficiency was evaluated (i.e. growth in same grade, same school, from Baseline to 2016/17). Grades 3, 4, and 5 aggregated showed an ST Math effect of 0.43 z-score points.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

This is a quasi-experimental analysis at the grade-mean level. Entire grades represent the units of analysis, and outcome measures are the changes in grade-mean z-score after 1 or 2 years of using ST math. The treatment grades used the ST Math program for either 1 or 2 years, beginning in the 2015/16 or 2016/17 school year. The study hypothesis is treatment grades using ST Math will outperform similar matched control grades, using their "business as usual" conditions of instructional content and professional development. The control grades were selected to have similar demographic and math attributes to the treatment grades during the baseline year (2014/15 or 2015/16), and did not use ST Math in 2016/17. The treatment grades' selection pool was all schools using ST Math in grades 3, 4, and 5 in Arkansas. The control grades' pool was all schools not using ST Math in grades 3, 4, and 5 in Arkansas. This study method measures effectiveness of the ST Math program when nominally implemented.

1.2 Program Description

The ST Math program is a supplemental math program covering grade-level math standards. The ST Math content consists of visual representations of math standards, concepts, and procedures, presented to students as "Puzzles" of virtual manipulatives, with which they interact to pose solutions. Each time the student poses a solution, the computer visually animates the Puzzle, diagram, or symbols to show why the posed solution correctly solves, or why it does not solve, the math problem (puzzle). The Puzzles are arranged into sequential groups, called "Levels". To proceed to the next Level in sequence, the student needs to master his/her current Level. Mastering a Level requires solving 100% of the math problems, or Puzzles correctly. In this way, the program is self-paced. Students must correctly solve approximately 4-12 Puzzles, with only 1 failure and retry allowed, to proceed. Levels are sequenced together into Games and, again, the student must master each Game to get to the next Game in sequence. Games are sequenced into "Learning Objectives" (e.g. 'Fractions Concepts'). The ST Math curriculum of approximately 20-25 Learning Objectives can be rearranged in a year-long, grade-level syllabus to match district math pacing through the school year.

The Puzzles typically start with concrete representations of the math, without abstract symbols, math vocabulary, or even English words. Gradually, through subsequent Levels or Games, abstractions are introduced. For example, a Puzzle might start with "n" green blocks on the screen, and then at a subsequent Level may represent the quantity with the numeral for "n" (no green blocks anymore). In this way, three things are accomplished: i) language proficiency prerequisites to engage with the program are minimal, ii) non-mathematical distractions (e.g. back-stories for word problems) are minimized or eliminated – thereby reducing load on working memory, and iii) the actual math in the problem can be represented clearly, simply, and unambiguously.

Besides the self-paced progress made by students in their one-to-one environment, the program is designed to be referenced by teachers during their regular math instruction. It is supplemental to core or basal math instruction and instructional materials. As the great majority of grade-level math standards are covered in the ST Math digital curriculum, completion of 100% of the entire ST Math curriculum (i.e. completing every Game) is required to cover all grade-level math standards.

Teachers receive initial training, either face to face or through self-guided online instruction. The training covers account startup, as well as math learning and growth mindset goals, the pedagogical

approach to learning in a visual experiential game, monitoring and intervention of the student 1:1 game play, and connecting of ST Math content to classroom content and pacing.

To achieve nominal progress through the program, there is a time-on-task requirement. While student progress rates through the program vary, MIND Research Institute has found that consistent application of 90 minutes per week throughout the school year is sufficient to get most students through at least half of the ST Math Learning Objectives. Students are recommended to use the program in school for at least two 45-minute sessions per week, or 90 minutes per week, over about 35 weeks. Analyses of ST Math usage have shown that consistently following this schedule throughout the school year is usually sufficient to achieve 50% or more Progress through ST Math content. Progress is a percentage of ST Math content coverage, and is defined as Levels completed by the student, divided by the total number of Levels in the curriculum. In addition, MIND's historical analyses have shown that it is necessary to complete at least 50% of the program in order to expect significantly higher performance compared to non-users.

2 Data Collection

Since this analysis uses grades as the unit of analysis, and states publish grade-mean state standardized test scores, the data for student math outcomes is collected from each state education agency's research files (retrieved from state websites). The school-level demographic data is also collected from the MDR (Market Data Retrieval, Shelton CT) database. The treatment students use ST Math student accounts served by MIND. Student ST Math usage data is aggregated to grade-level means by MIND.

2.1 Treatment Grades Pool and Selection

The Treatment grades pool originated with all schools and grades using ST Math in Arkansas. From these schools, every grade that had used the ST Math program in 2016/17 was identified. They comprise the Treatment grades pool for this evaluation of 1-year usage.

Because the analysis uses grade-mean data, such as grade-mean scale scores or grade-mean proficiency level percentages, it is necessary that the program also be a grade-wide treatment, with the great majority of students in each grade receiving treatment. Otherwise, the grade-means reported by the state of 100% of *tested* students would not be valid measures of a smaller fraction of *treatment* students. MIND's site implementation requirement is that an entire grade, including all teachers and all classes within that grade, use the ST Math program. We validate how closely this is the case for each individual treatment grade by comparing the number of ST Math student accounts at a grade level to the reported enrollment at that grade level. We discard from the Treatment pool any grade with a ratio of ST Math student accounts to reported grade enrollment lower than 85%.

Furthermore, the outcomes measure is a summative year-end test, i.e. that state's standardized math assessment. The math assessment thus covers all the math standards for that entire grade level. Meanwhile, the ST Math program curriculum (arranged into Learning Objectives) is also aligned to that state's math standards. To infer that the ST Math content is having a valid effect on student outcomes on the summative assessment, we discard any grade with grade-mean of ST Math Progress for its students lower than 50% by year-end.

Progress is a percentage, and is defined as Levels completed by the student, divided by the total number of Levels in the grade-level curriculum. Note that student achievement of at least 50%

progress in ST Math is accomplished primarily by teacher assignment of computer session time to students. With sufficient time on task, students make progress. The program helps them self-pace through providing real-time informative feedback for each puzzle.

2.2 Control Grades Pool and Selection

The control grades are randomly selected from a control pool of schools in Arkansas. Though they are randomly selected, they are also matched to be similar to the Treatment grades' math attributes and demographics during the baseline year. The matched attributes include:

- z-score of math proficiency
- percentage of students receiving free or reduced lunch (using the demographic data from MDR).

To mitigate the risk of randomly picking a set of Control grades that generates an outlier for effect, a Monte Carlo approach is used to perform many random picks. The control pool's size is large enough that there are many possible "picks" of closely matched control grades.

One hundred randomly matched picks are made and sets of matched control grades are generated. For each set, the quality of the match as well as the math growth of the potential control set is evaluated. Some picked sets have high average math growth, some have low average math growth. From the set of all picks, a median pick is chosen. This avoids either an unlikely overestimate, or underestimate, of the Control grades' growth. When multiple median picks exist, the control set with the minimal math score differences in the baseline year is chosen.

3 Data Analysis

The set of all schools and grades using ST Math in Arkansas is evaluated for Enrollment percentage and Progress percentage parameters. A filtered Treatment set (TRT) of all ST Math grades with $\geq$ 85% Enrollment and $\geq$ 50% Progress is identified. State math assessment data is tabulated. A matching set of Control grades based on baseline year state math assessment is selected.

Changes in math performance, i.e. the difference in math performance of a grade from a baseline year to the final year, are evaluated and tabulated. Statistical tests of the significance of the difference in math performance changes between Treatment grades and Control grades are performed. Finally, a grade-by-grade disaggregation is performed.

3.1 Z-scores

In order to analyze across all states with different math assessments, a new z-score of that test's math proficiency is calculated. For each year being analyzed, by grade, a z-score takes the difference of the grade mean percent proficient and the mean of all percent proficient statewide for that year, and then divides it by the standard deviation of all percent proficient statewide for that year. Here is a fictional example to illustrate the calculation of a z-score for the 2015/16 exam:

School A, Grade 3, Percent Proficient: 70
Average across all schools statewide, Grade 3: 50
Standard deviation across all schools statewide, Grade 3: 20
Z-score=((School A, Grade 3, Percent Proficient)-(Average across all schools, Grade 3))/(Standard deviation across all schools, Grade 3)

$$\text{Z-score}=\frac{70-50}{20}=1$$

The z-score is calculated for every grade across all years being analyzed, using the full state data set of schools for the averages and standard deviations. The use of z-scores is a valid statistical method to normalize any dataset and to enable analysis across otherwise uncomparable exams. In this report, we only analyze z-scores.

3.2 Percentile Ranking

These newly calculated z-scores can then be converted into a percentile ranking. Each percentile ranking shows the grade's performance relative to the others in that year and grade. For example, for a specific grade 3, a percentile ranking of 50 shows that this grade 3 performed at the average of all third grades in the state for that testing year.

3.3 Final Treatment and Control

3.3.1 ST Math Grade-Aggregated Implementation ($\geq$ 85% Enrollment Grades Only)

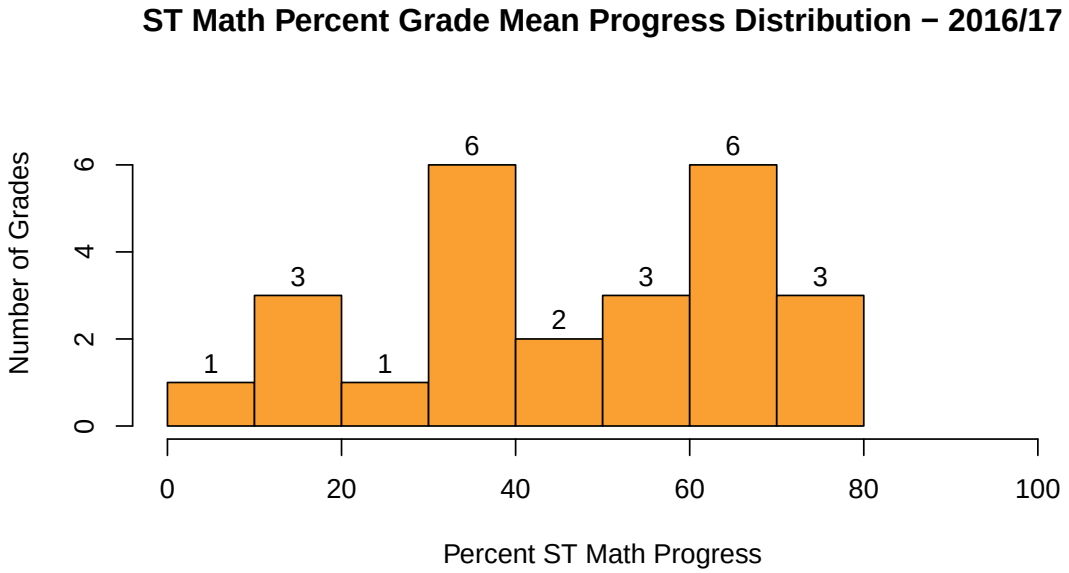

Figure 1: Histogram of ST Math Percent Progress for $\geq$ 85% Enrollment Grades 2016/17

For all ST Math grades with Enrollment $\geq$ 85%, Figure 1 shows the frequency distribution of grade-average Progress percentage through the program. Note that we will only be using grades with $\geq$ 50% Progress as the Treatment Group.

Table 1 provides descriptive statistics of the Progress distribution. Table 2 shows the number of remaining treatment grades after applying enrollment and progress filters.

	Min.	Max.	Average	S.D.
ST Math % Progress	9.7	79.0	47.0	22.1

Table 1: Descriptive Statistics of ST Math Percent Progress for >= 85 percent Enrollment Grades

Grades with >= 85% Enrollment:	19
Grades with in addition >= 50% Progress:	9

Table 2: Number of ST Math Grades with >= 85 percent Enrollment and with >= 50 percent progress

3.3.2 Filtering Treatment and Controls

Table 3 shows the total number of grades in the Treatment pool, the number of grades that exceeded the 85% Enrollment figure, and also the 50% Progress filter. Other rows in the table indicate counts of numbers of students (2016/17 from state testing count) and counts of number of schools represented. The number of matched Control (CTRL) grades, students, and schools is also shown.

	Grade 3	Grade 4	Grade 5	Total
ST Math Using Grades	9	7	9	25
ST Math Using Schools	9	7	9	11
ST Math Students	697	592	675	1964
ST Math Grades (Enroll >= 85%)	8	6	5	19
TRT Grades (Enroll >= 85% & Prog >= 50%)	3	2	4	9
TRT Schools (Enroll >= 85% & Prog >= 50%)	3	2	4	5
TRT Students (Enroll >= 85% & Prog >= 50%)	242	188	402	832
CTRL Grades	3	2	4	9
CTRL Schools	3	2	4	8
CTRL Students	179	195	353	727

Table 3: Treatment Pool Filtering and Controls: Counts of Grades, Schools, and Students

3.3.3 Match of Controls to Treatment

Figure 2 shows the density plot of the baseline z-scores for treatment grades overlayed on control grades, showing the closeness of the match obtained between Treatment and Control sets of grades in the baseline year. It is important to keep in mind that we only have a small number of treatment and control grades (9) and that the Control set was arrived at through a Monte Carlo process (see Section 2.3) rather than a closest math performance match.

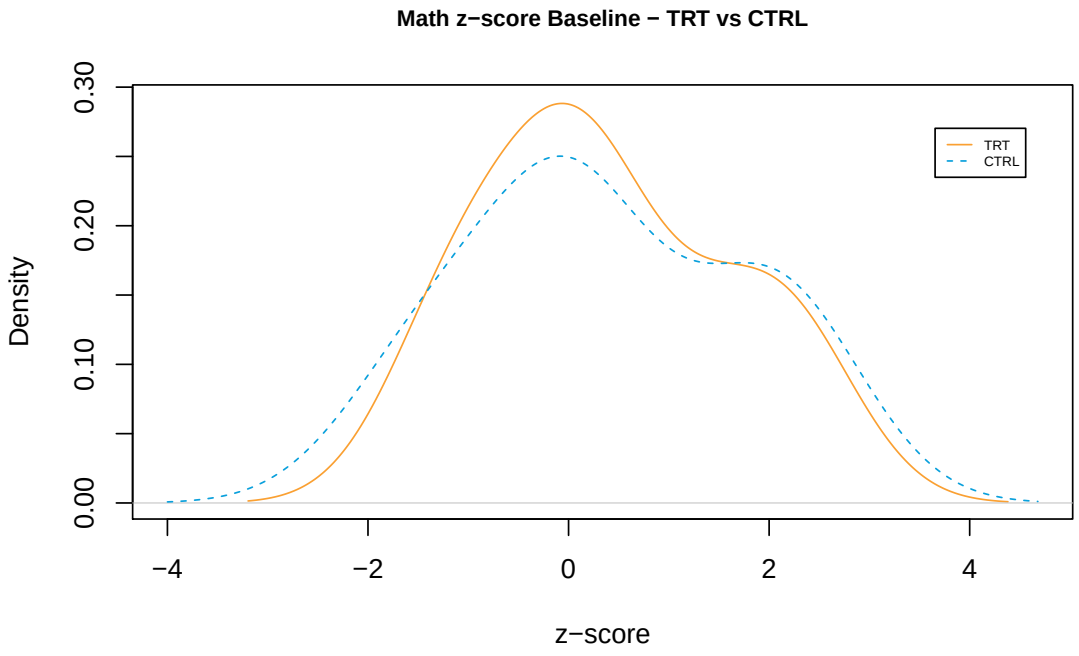

Figure 2: Baseline Year Density Plots Showing Math z-scores Match between TRT and CTRL - Baseline

Similarly, Figure 3 shows the density plot of the percentage of students needing free or reduced lunch for treatment grades overlayed on control grades, showing the closeness of the match obtained between Treatment and Control sets of grades.

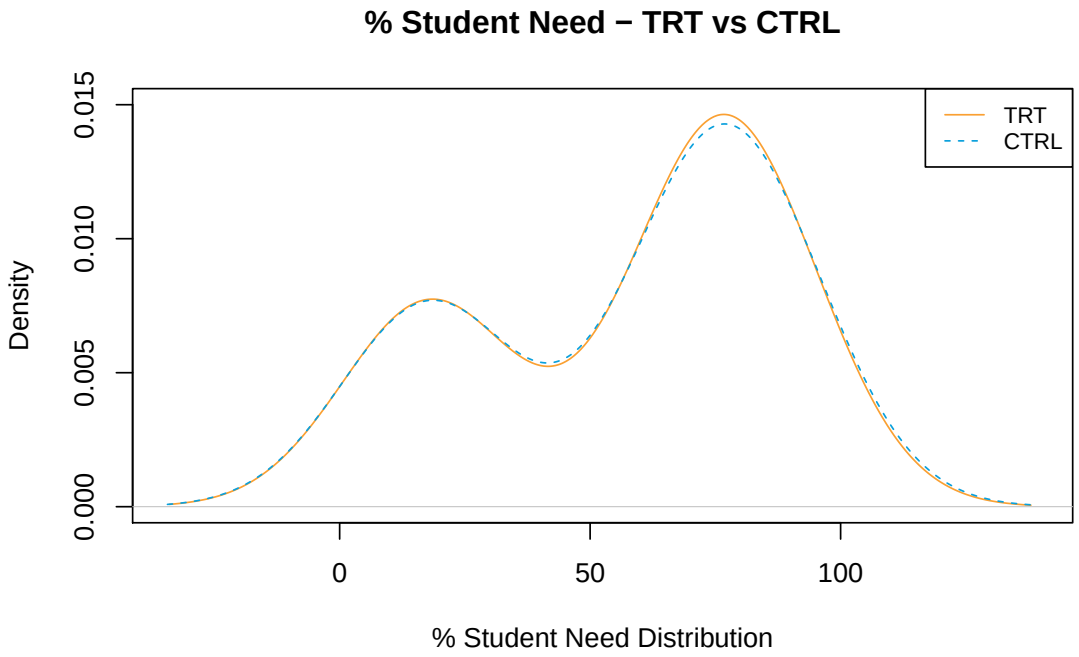

Figure 3: Baseline Year Density Plot Showing Student Need Match between TRT and CTRL

Table 4 shows the difference of the means of Treatment versus Control in the baseline year, with accompanying p-values, for mean z-score and for percent of students receiving free or reduced lunch. The large p-values show the differences between the Treatment and Control grades are not statistically significant.

	Mean(TRT)	SD(TRT)	Mean(CTRL)	SD(CTRL)	Estimate	P-Value
Z-score - Baseline	0.39	1.25	0.39	1.39	-0.00	1.00
Percent Free or Reduced Lunch	57.33	29.88	57.44	30.10	-0.11	0.99

Table 4: Matching TRT and CTRL

3.4 Grade-Aggregated Analysis

Table 5 shows for both Treatment (TRT) and Control (CTRL) aggregation across grades of z-score distributions. The far right column also shows the average ST Math Progress for the TRT set.

	# Grades	# Schools	# Students	Z-score	Percentile	ST Math Per Comp.
TRT.Baseline	9	5	828	0.39	57.78	–
TRT.16.17	9	5	791	0.77	72.89	67.4
TRT.Delta	–	–	–	0.38	15.11	–
CTRL.Baseline	9	8	699	0.39	57.44	–
CTRL.16.17	9	8	727	0.35	57.11	–
CTRL.Delta	–	–	–	-0.04	-0.33	–

Table 5: All Grades Together Growth

Figure 4 shows the changes in z-scores of math proficiency for the grade-aggregated Treatment and Control sets.

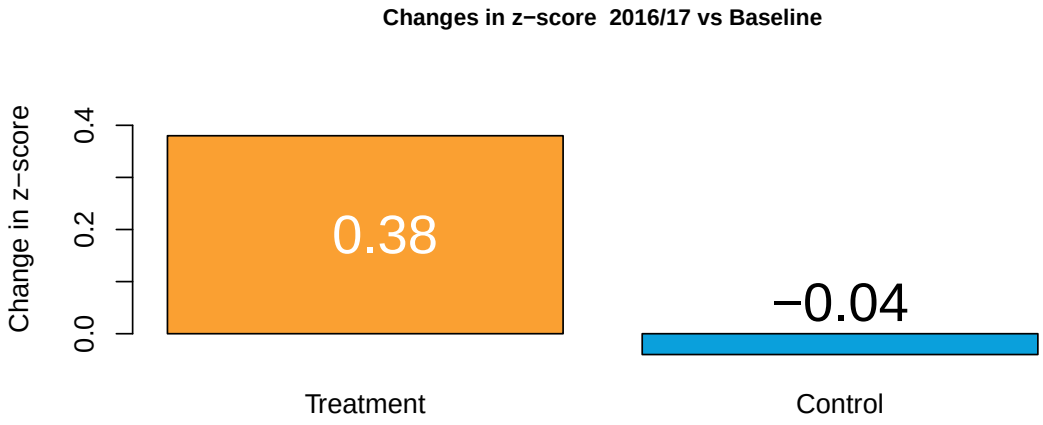

Figure 4: Changes in z-score for grade-Aggregated TRT and CTRL datasets between Baseline and 2016/17

Table 6 shows the statistics for the *differences* in changes between TRT and CTRL (Treatment - Control) for these same z-score changes as in the above figures. [1]

	Estimate	P-Value	Int.Low	Int.High
Z-score	0.43	0.33	-0.47	1.32

Table 6: Statistics for the Differential Changes in Math Scores Growth (TRT - CTRL)

[1] * statistically significant $p<0.05$

Finally, Figure 5 shows the changes in mean percentile ranking between TRT and CTRL.

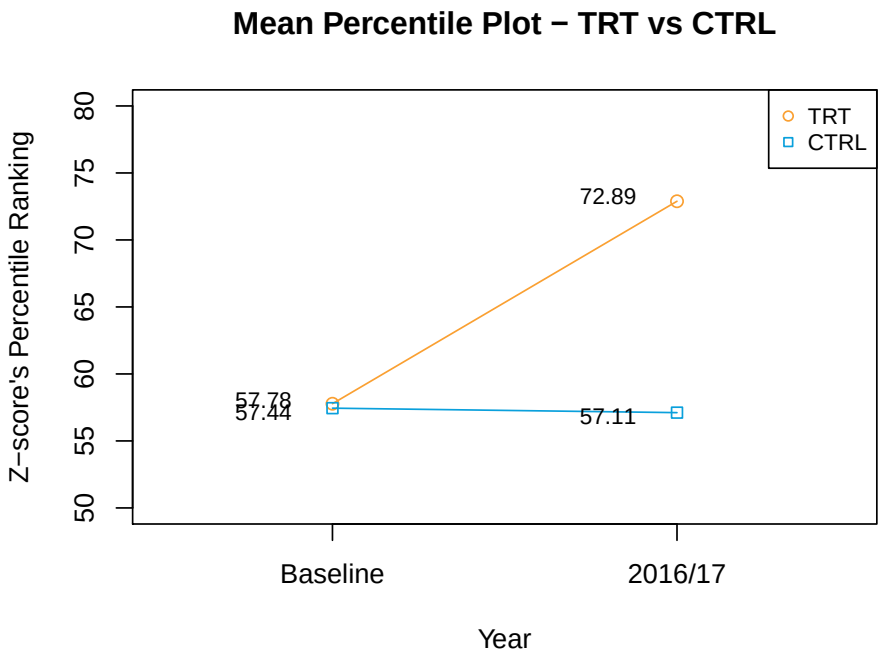

Figure 5: Changes in Percentile Ranking for TRT and CTRL Datasets between Baseline and 2016/17

3.5 Grade-Level Analysis

3.5.1 Grade Level Result Tables

The following tables (Table 7, 8, and 9) present a disaggregation of results by grade level. The far right column in each table also shows the average ST Math Progress for the TRT set.

	# Grades	# Schools	# Students	Z-score	Percentile	ST Math Per Prog.
TRT.Baseline	3	3	224	0.30	58.67	–
TRT.16.17	3	3	224	0.62	68.33	61.83
TRT.Delta	–	–	–	0.32	9.67	–
CTRL.Baseline	3	3	179	0.68	70.00	–
CTRL.16.17	3	3	179	0.29	54.67	–
CTRL.Delta	–	–	–	-0.39	-15.33	–

Table 7: Grade 3 - Yearly Math Performance and Counts for TRT and CTRL Datasets

	# Grades	# Schools	# Students	Z-score	Percentile	ST Math Per Prog.
TRT.Baseline	2	2	196	0.42	55.00	–
TRT.16.17	2	2	196	0.95	79.50	72.5
TRT.Delta	–	–	–	0.53	24.50	–
CTRL.Baseline	2	2	195	0.48	54.50	–
CTRL.16.17	2	2	195	0.86	73.00	–
CTRL.Delta	–	–	–	0.39	18.50	–

Table 8: Grade 4 - Yearly Math Performance and Counts for TRT and CTRL Datasets

	# Grades	# Schools	# Students	Z-score	Percentile	ST Math Per Prog.
TRT.Baseline	4	4	371	0.44	58.50	–
TRT.16.17	4	4	371	0.80	73.00	69.02
TRT.Delta	–	–	–	0.36	14.50	–
CTRL.Baseline	4	4	353	0.12	49.50	–
CTRL.16.17	4	4	353	0.13	51.00	–
CTRL.Delta	–	–	–	0.00	1.50	–

Table 9: Grade 5 - Yearly Math Performance and Counts for TRT and CTRL Datasets

3.5.2 Grade-Level Analysis of Changes in Math Z-score

Figure 6 shows the difference in the change of z-scores, for the TRT and CTRL datasets, disaggregated by grade:

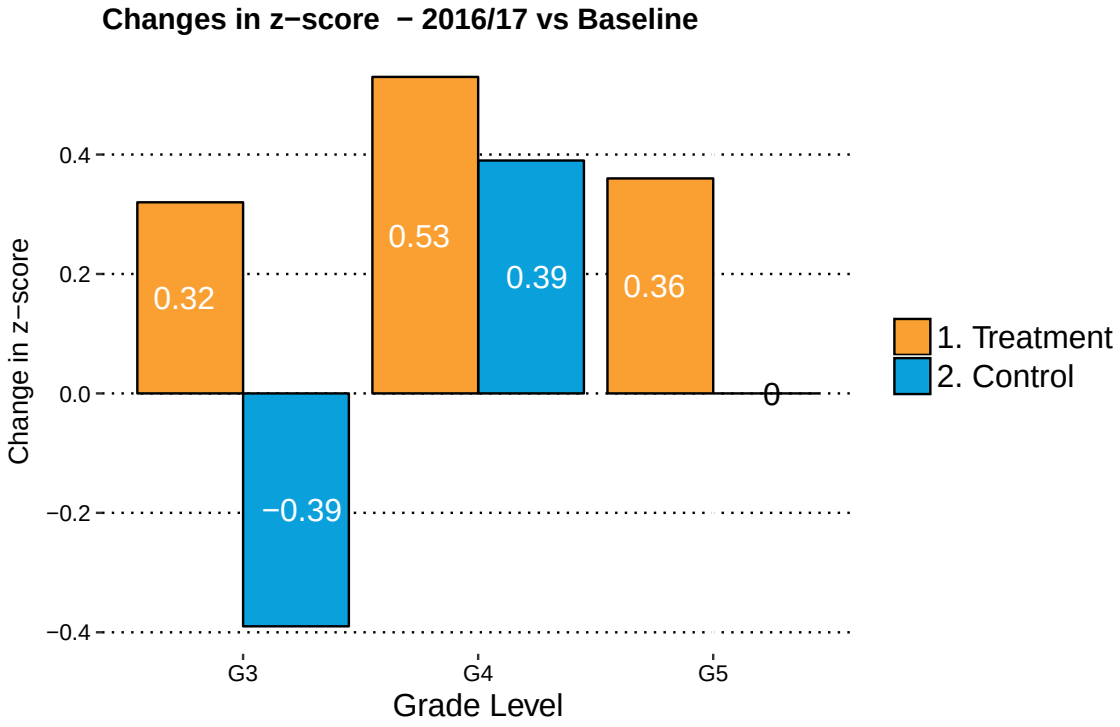

Figure 6: Changes in z-scores for TRT and CTRL Datasets between Baseline and 2016/17

Table 10 shows the statistics for the *differences* in changes between TRT and CTRL (Treatment - Control) z-scores as shown in Figure 6.

	Estimate	P-Value	Int.Low	Int.High
Grade 5	0.36	0.67	-1.58	2.30

Table 10: Statistics for the Differential Changes in z-scores (TRT - CTRL)

4 Effect Size

The following table shows the effect sizes for z-scores.

	Z-score Effect Size
Grade 3	0.79
Grade 4	0.06
Grade 5	0.22
All Grades	0.31

Table 11: Cohen's d Effect Size

5 Findings Summary

Grades 3, 4, and 5 using ST Math for the year 2016/17 averaged 47% ST Math Progress. 12/25 grades (48%) averaged covering more than 50% of ST Math content. No statistically significant differences were found in this analysis due to the small number of treatment grades.

6 Confounders

Despite best efforts in minimizing confounders to the results of this analysis, there still remain a few input variables that could be significant in affecting differences of state test score outcomes between the Treatment and Control sets. One issue is the lack of randomization of grades chosen to receive the ST Math treatment. Instead of randomized selection, Treatment grades are self-selected. Self-selection can be an indication of districts or schools with a focus on math, an appetite for change, and with a spotlight on math training. Furthermore, not all grades using the ST Math program are chosen for analysis. Each grade must pass two specific filters to be considered for the Treatment set: the first being an enrollment filter of at least 85% of students in each grade using the program, and the second being a progress filter of at least 50% of the program completed on average by students in that grade. These filters might indicate relatively high-functioning schools with a team of relatively effective teachers in that grade, thus resulting in better instruction overall. A mitigation of this possible confounder is our selection of treatment groups on the grade level, rather than the teacher level, so there is no cherry picking of teachers: the full range of teachers in each grade is included. Moreover, the specific teachers may often be the same in the baseline year as in the current year, so the Treatment growth is not due to teacher differences. Finally, a possible confounder lies in the "business as usual" conditions at the matched control grades chosen for each analysis. It's unknown whether these control grades used other programs that could affect the comparison of the two sets of grades. The Monte Carlo Method is used to mitigate the possibility of control picks being favorable or unfavorable (see Section 2.3).

7 Reference Tables Grouped By School Year

The following tables show grade-level details, grouped by school year and for treatment (Table 12) and controls (Table 13) separately.

	# Grades	# Schools	# Students	Z-score	Percentile	ST Math Per Comp.
Grade 3 (Baseline)	3	3	224	0.30	58.67	–
Grade 4 (Baseline)	2	2	196	0.42	55.0	–
Grade 5 (Baseline)	4	4	371	0.44	58.5	–
All Grades (Baseline)	9	5	828	0.39	57.78	–
Grade 3 (16.17)	3	3	224	0.62	68.33	61.83
Grade 4 (16.17)	2	2	196	0.95	79.5	72.5
Grade 5 (16.17)	4	4	371	0.80	73.0	69.02
All Grades (16.17)	9	5	791	0.77	72.89	67.4

Table 12: TRT Grades Detail Sorted by Year

	# Grades	# Schools	# Students	Z-score	Percentile	ST Math Per Comp.
Grade 3 (Baseline)	3	3	179	0.68	70.00	–
Grade 4 (Baseline)	2	2	195	0.48	54.5	–
Grade 5 (Baseline)	4	4	353	0.12	49.5	–
All Grades (Baseline)	9	8	699	0.39	57.44	–
Grade 3 (16.17)	3	3	179	0.29	54.67	–
Grade 4 (16.17)	2	2	195	0.86	73.0	–
Grade 5 (16.17)	4	4	353	0.13	51.0	–
All Grades (16.17)	9	8	727	0.35	57.11	–

Table 13: CTRL Grades Detail Sorted by Year

8 Lists of Schools

8.1 Treatment Schools

The following table lists the treatment schools and grades (after 85% enrollment and 50% progress filtering) used in the analysis.

State	PID	IID	District Name	School Name	GRADE
	4916906	BER5M4	SPRINGDALE SCHOOL DISTRICT	BERNICE YOUNG ELEMENTARY	3, 4, 5
	4015655	GEO5M4	SPRINGDALE SCHOOL DISTRICT	GEORGE ELEMENTARY SCHOOL	5
	32417	LAW5KL	PULASKI COUNTY SPECIAL SCHOOL DISTRICT	LAWSON ELEMENTARY SCHOOL	3
	11713943	SON5M4	SPRINGDALE SCHOOL DISTRICT	SONORA ELEMENTARY SCHOOL	3, 4, 5
	35653	WES5M4	SPRINGDALE SCHOOL DISTRICT	WESTWOOD ELEMENTARY SCHOOL	5

Table 14: Treatment Schools (TRT Dataset)

8.2 Control Schools

The following table lists the control schools and grades (matched control grades to treatment grades) used in the analysis.

State	PID	District Name	School Name	GRADE
	28739	FLIPPIN SCHOOL DISTRICT	FLIPPIN ELEMENTARY SCHOOL	3
	31580	LITTLE ROCK SCHOOL DISTRICT	FOREST PARK ELEMENTARY SCHOOL	4, 5
	2045563	ROGERS SCHOOL DISTRICT	GRACE HILL ELEMENTARY SCHOOL	5
	35902	RIVERVIEW SCHOOL DISTRICT	JUDSONIA ELEMENTARY SCHOOL	3
	35938	RIVERVIEW SCHOOL DISTRICT	KENSETT ELEMENTARY SCHOOL	5
	26092	NASHVILLE SCHOOL DISTRICT	NASHVILLE ELEMENTARY SCHOOL	5
	20244	STUTTGART SCHOOL DISTRICT	PARK AVENUE ELEMENTARY SCHOOL	4
	35392	FAYETTEVILLE SCHOOL DISTRICT	ROOT ELEMENTARY SCHOOL	3

Table 15: Matched Control Schools (CTRL Dataset)