

West Virginia Math Outcomes Analysis 2016/17

Grade Levels: 3, 4, 5
ST Math Program: Gen-4
Analysis Type: Z-score of scale score
Treatment-Years: 2016/17
Baseline-Year: 2015/16
Subgroup: All Students

Jessica Guise
MIND Research Institute
© 2018-01-23

Abstract

This analysis evaluates grades using ST Math in West Virginia 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 46 grades, consisting of grades 3, 4, and 5 at 17 schools. Refer to Figures 2 and 3 for the math performance and demographic distributions. They were matched to 46 similar, randomly selected control grades at 44 schools that never used ST Math. Grade-wise growth in math proficiency was evaluated (i.e. growth in same grade, same school, from 2015/16 to 2016/17) on the percentage proficient, scale scores, and Z-scores of the scale scores (see Section 3.1). Grades 3, 4, and 5 aggregated showed an ST Math effect of 6.95 points at the Level 3 or 4 levels, 3.89 points at the Level 4 Level, and z-score of 0.22.

Contents

1	Introduction	5
1.1	Background	5
1.2	Program Description	5
2	Data Collection	6
2.1	Proficiency Levels Definition	6
2.2	Treatment Grades Pool and Selection	6
2.2.1	Enrollment Filter	6
2.2.2	Content Coverage Filter	7
2.3	Control Grades Pool and Selection	7
3	Data Analysis	8
3.1	Z-scores	8
3.2	Percentile Ranking	8
3.3	Final Treatment and Control	9
3.3.1	ST Math Grade-Aggregated Implementation ($\geq 85\%$ Enrollment Grades Only)	9
3.3.2	Filtering Treatment and Controls	10
3.3.3	Match of Controls to Treatment	11
3.4	Grade-Aggregated Analysis	13
3.5	Grade-Level Analysis	17
3.5.1	Grade Level Result Tables	17
3.5.2	Grade-Level Analysis of Changes in Math Level 3 or 4	18
3.5.3	Grade-Level Analysis of Changes in WVGSA Math scale scores	19
3.5.4	Grade-Level Analysis of Changes in WVGSA z-scores of scale scores	20
4	Effect Size	21
5	Findings Summary	21
6	Confounders	21
7	Reference Tables Grouped By School Year	23
8	Lists of Schools	24
8.1	Treatment Schools	24
8.2	Control Schools	25

List of Figures

1	Histogram of ST Math Percent Progress for $\geq 85\%$ Enrollment Grades 2016/17	9
2	Baseline Year Density Plots Showing Math Scores Match between TRT and CTRL - 2015/16	11
3	Baseline Year Density Plot Showing Student Need Match between TRT and CTRL	12
4	Change at each Proficiency Level for Grade-Aggregated TRT and CTRL Datasets between 2015/16 and 2016/17	13
5	Changes in WVGSA Math scale scores and z-scores (See Section 3.1) for Grade-Aggregated TRT and CTRL datasets between 2015/16 and 2016/17	14
6	Changes in Level 3 or 4 for Grade-Aggregated TRT and CTRL datasets between 2015/16 and 2016/17	15
7	Changes in Percentile Ranking for TRT and CTRL Datasets between 2015/16 and 2016/17	16
8	Changes in Percent of Students at Level 3 or 4 for TRT and CTRL Datasets between 2015/16 and 2016/17	18
9	Changes in Grade-Mean WVGSA Math scale score for TRT and CTRL Datasets between 2015/16 and 2016/17	19
10	Changes in Grade-Mean WVGSA z-score (See Section 3.1) for TRT and CTRL Datasets between 2015/16 and 2016/17	20

List of Tables

1	Proficiency Level Naming	6
2	Descriptive Statistics of ST Math Percent Progress for ≥ 85 percent Enrollment Grades	9
3	Number of ST Math Grades with ≥ 85 percent Enrollment and with ≥ 50 percent progress	9
4	Treatment Pool Filtering and Controls: Counts of Grades, Schools, and Students	10
5	Matching TRT and CTRL	12
6	Yearly Math Proficiency and Counts for TRT and CTRL Grade-Aggregated Datasets	13
7	Statistics for the Differential Changes in Math Scores Growth (TRT - CTRL) .	15
8	Grade 3 - Yearly Math Performance and Counts for TRT and CTRL Datasets	17
9	Grade 4 - Yearly Math Performance and Counts for TRT and CTRL Datasets	17
10	Grade 5 - Yearly Math Performance and Counts for TRT and CTRL Datasets	17
11	Statistics for the Differential Changes in Level 3 or 4, (TRT - CTRL)	18
12	Statistics for the Differential Changes in WVGSA Math scale scores Growth, (TRT - CTRL)	19
13	Statistics for the Differential Changes in WVGSA z-scores (See Section 3.1) Growth, (TRT - CTRL)	20
14	Cohen's d Effect Size	21
15	TRT Grades Detail Sorted by Year	23
16	CTRL Grades Detail Sorted by Year	23
17	Treatment Schools (TRT Dataset)	24
18	Matched Control Schools (CTRL Dataset)	25

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 1-year changes in grade-mean WVGSAs Level 3 or 4 percentages. The treatment grades used the ST Math program for 1 year, beginning in the 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 (See Figures 2 and 3) to the treatment grades during the baseline year (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 West Virginia. The control grades’ pool was all schools not using ST Math in grades 3, 4, and 5 in West Virginia. 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 West Virginia 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 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 Proficiency Levels Definition

The following (Table 1) is West Virginia’s proficiency level descriptions:

Proficiency Level	State Proficiency Level Name
L1	Level 1
L2	Level 2
L3	Level 3
L4	Level 4

Table 1: Proficiency Level Naming

2.2 Treatment Grades Pool and Selection

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

2.2.1 Enrollment Filter

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 West Virginia’s 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%.

2.2.2 Content Coverage Filter

Furthermore, the outcomes measure is a summative year-end test, i.e. West Virginia’s standardized math assessment (WVGSA). 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 West Virginia 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.

2.3 Control Grades Pool and Selection

The control grades are randomly selected from a control pool of schools in West Virginia. Though they are randomly selected, they are also matched to be similar to the Treatment grades’ math attributes and demographics during the baseline 2015/16 year. The matched attributes include:

- scale score
- student percentages at each math proficiency level
- 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 West Virginia 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

When states change their state assessment throughout the years, they also change the range of possible scale scores achieved on the exam. This makes it difficult to compare changes in grade mean scale scores across years with a different exam. To deal with this issue, a new z-score is calculated. For each year being analyzed, by grade, a z-score takes the difference of the grade mean scale score and the mean of all scale scores statewide for that year, and then divides it by the standard deviation of all scale scores statewide for that year. Here is a fictional example to illustrate the calculation of a z-score for the 2015/16 exam:

$$\begin{aligned} &\text{School A, Grade 3, Mean scale score: } 300 \\ &\text{Average across all schools statewide, Grade 3: } 350 \\ &\text{Standard deviation across all schools statewide, Grade 3: } 30 \\ \text{Z-score} &= ((\text{School A, Grade 3, Mean scale score}) - (\text{Average across all schools, Grade 3})) / (\text{Standard} \\ &\quad \text{deviation across all schools, Grade 3}) \\ \text{Z-score} &= \frac{300 - 350}{30} = -1.67 \end{aligned}$$

The z-score is calculated for every grade across all years being analyzed, using the full state data set of West Virginia 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 will include both mean scale scores and their accompanying 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)

ST Math Percent Grade Mean Progress Distribution – 2016/17

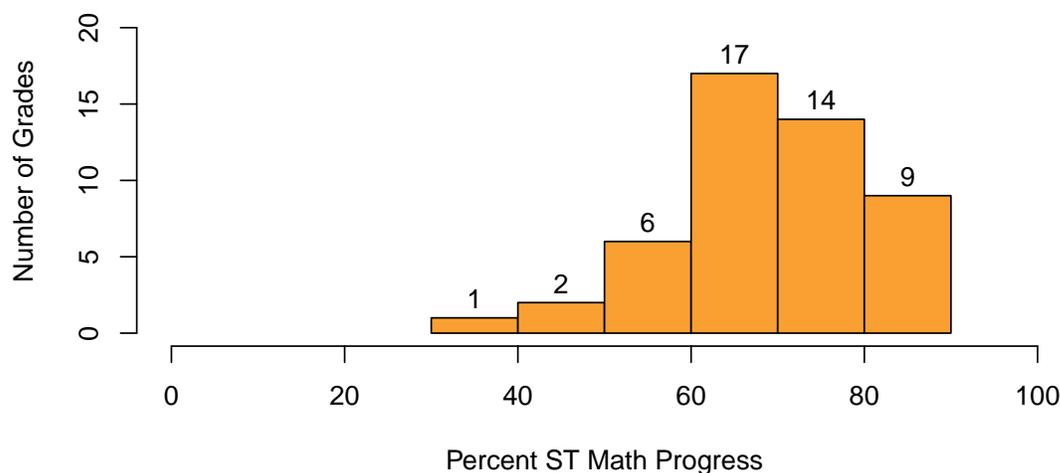


Figure 1: Histogram of ST Math Percent Progress for $\geq 85\%$ Enrollment Grades 2016/17

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 2 provides descriptive statistics of the Progress distribution. Table 3 shows the number of remaining treatment grades after applying enrollment and progress filters.

	Min.	Max.	Average	S.D.
ST Math % Progress	38.6	89.8	68.5	11.9

Table 2: Descriptive Statistics of ST Math Percent Progress for $\geq 85\%$ Enrollment Grades

Grades with $\geq 85\%$ Enrollment:	49
Grades with in addition $\geq 50\%$ Progress:	46

Table 3: Number of ST Math Grades with $\geq 85\%$ Enrollment and with $\geq 50\%$ percent progress

3.3.2 Filtering Treatment and Controls

Table 4 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	16	16	17	49
ST Math Using Schools	16	16	17	17
ST Math Students	735	735	895	2365
ST Math Grades (Enroll \geq 85%)	16	16	17	49
TRT Grades (Enroll \geq 85% & Prog \geq 50%)	15	15	16	46
TRT Schools (Enroll \geq 85% & Prog \geq 50%)	15	15	16	17
TRT Students (Enroll \geq 85% & Prog \geq 50%)	773	700	790	2263
CTRL Grades	15	15	16	46
CTRL Schools	15	15	16	44
CTRL Students	881	739	709	2329

Table 4: 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 WVGSA Math scale scores (left plot) and baseline percent students at WVGSA Level 3 or 4 (right plot) 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, 2015/16.

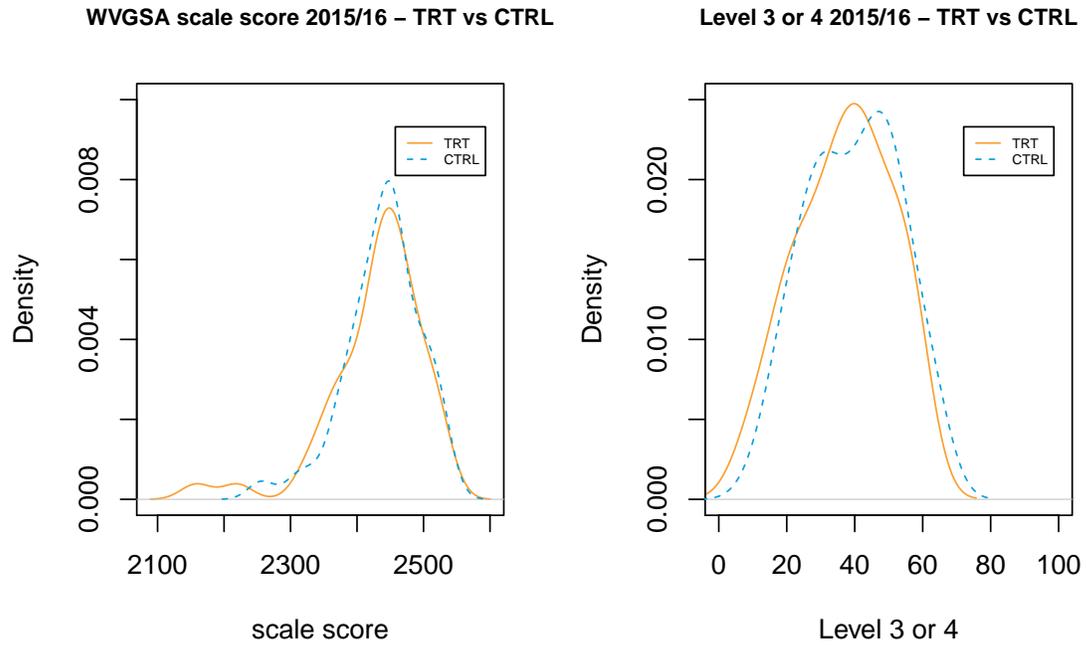


Figure 2: Baseline Year Density Plots Showing Math Scores Match between TRT and CTRL - 2015/16

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

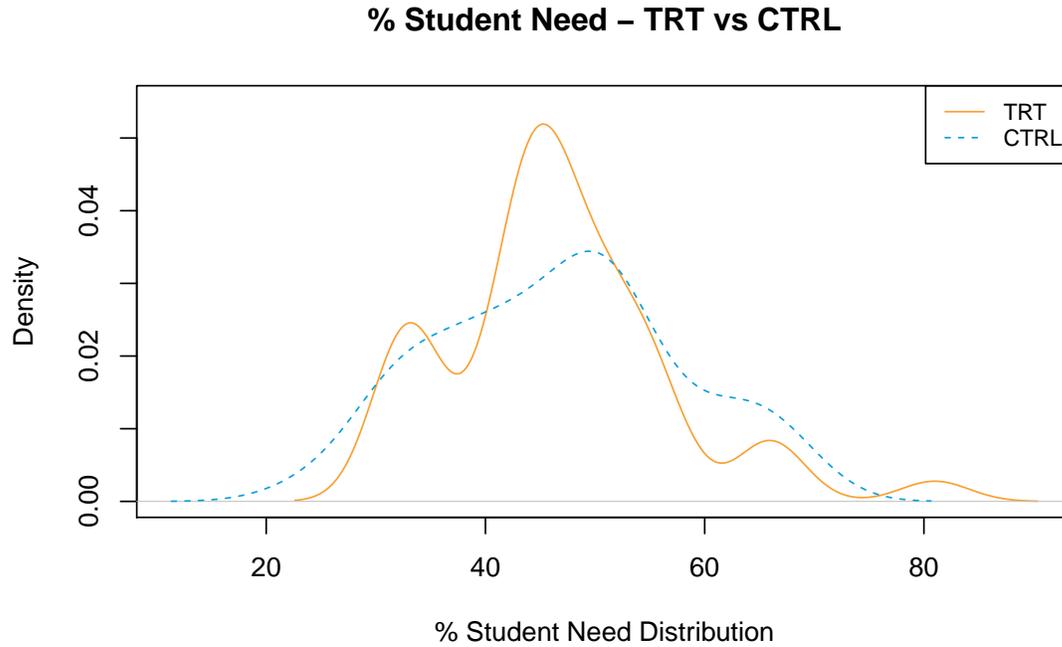


Figure 3: Baseline Year Density Plot Showing Student Need Match between TRT and CTRL

Table 5 shows the difference of the means of Treatment versus Control in the baseline year, with accompanying p-values, for percent Level 3 or 4, for mean scale 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
Level 3 or 4 - 2015/16	37.09	13.90	39.72	13.56	-2.63	0.36
Scale score - 2015/16	2429.19	72.76	2438.53	57.17	-9.35	0.50
Percent Free or Reduced Lunch	46.65	10.21	46.76	11.12	-0.11	0.96

Table 5: Matching TRT and CTRL

3.4 Grade-Aggregated Analysis

Table 6 shows for both Treatment (TRT) and Control (CTRL) aggregation across grades of scale scores, z-scores, and proficiency level distributions. The far right column also shows the average ST Math Progress for the TRT set.

	# Grades	# Schools	# Students	Scale score	Z-score	Percentile	L1	L2	L3	L4	Level 3 or 4	ST Math Per Comp.
TRT.15.16	46	17	2217	2429.2	-0.07	52.07	27.68	35.22	23.66	13.43	37.09	-
TRT.16.17	46	17	2217	2439.2	0.04	57.11	21.90	32.30	26.56	19.23	45.79	70.32
TRT.Delta	-	-	-	10.0	0.11	5.04	-5.78	-2.92	2.90	5.80	8.70	-
CTRL.15.16	46	44	2329	2438.5	0.09	54.78	25.85	34.42	26.11	13.61	39.72	-
CTRL.16.17	46	44	2329	2434.2	-0.03	51.83	27.18	31.35	25.95	15.52	41.47	-
CTRL.Delta	-	-	-	-4.3	-0.11	-2.96	1.32	-3.07	-0.16	1.91	1.75	-

Table 6: Yearly Math Proficiency and Counts for TRT and CTRL Grade-Aggregated Datasets

The following chart (Figure 4) shows the changes in percentage of students at each math proficiency level for the grade-aggregated Treatment and Control sets (TRT.delta and CTRL.delta).

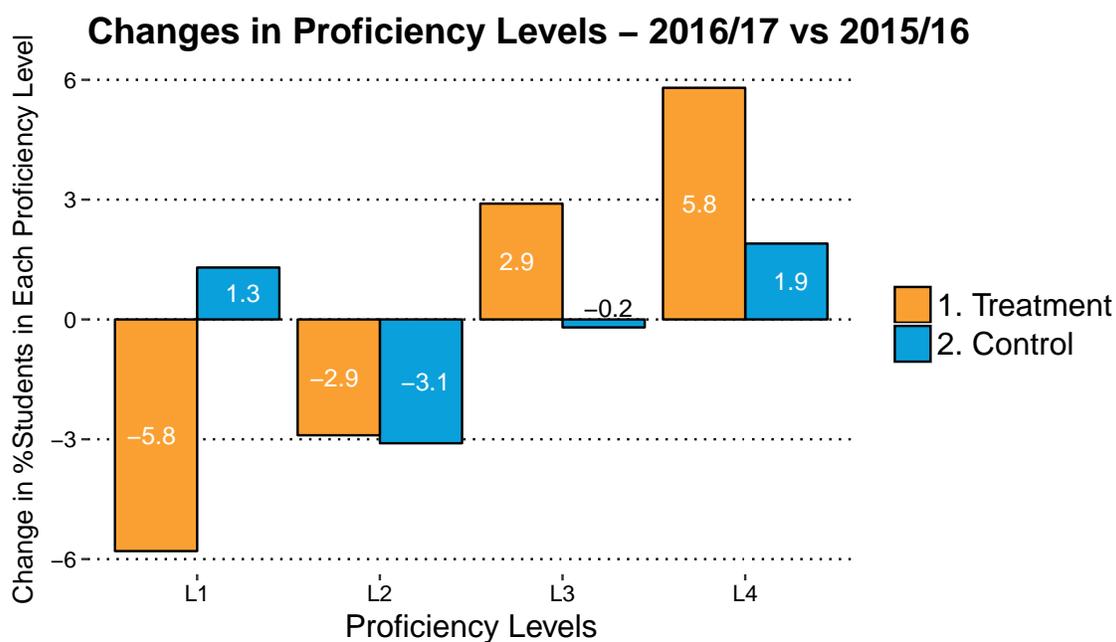


Figure 4: Change at each Proficiency Level for Grade-Aggregated TRT and CTRL Datasets between 2015/16 and 2016/17

Similarly, Figure 5 shows the changes in WVGSA Math scale scores and changes in z-scores for the grade-aggregated Treatment and Control sets.

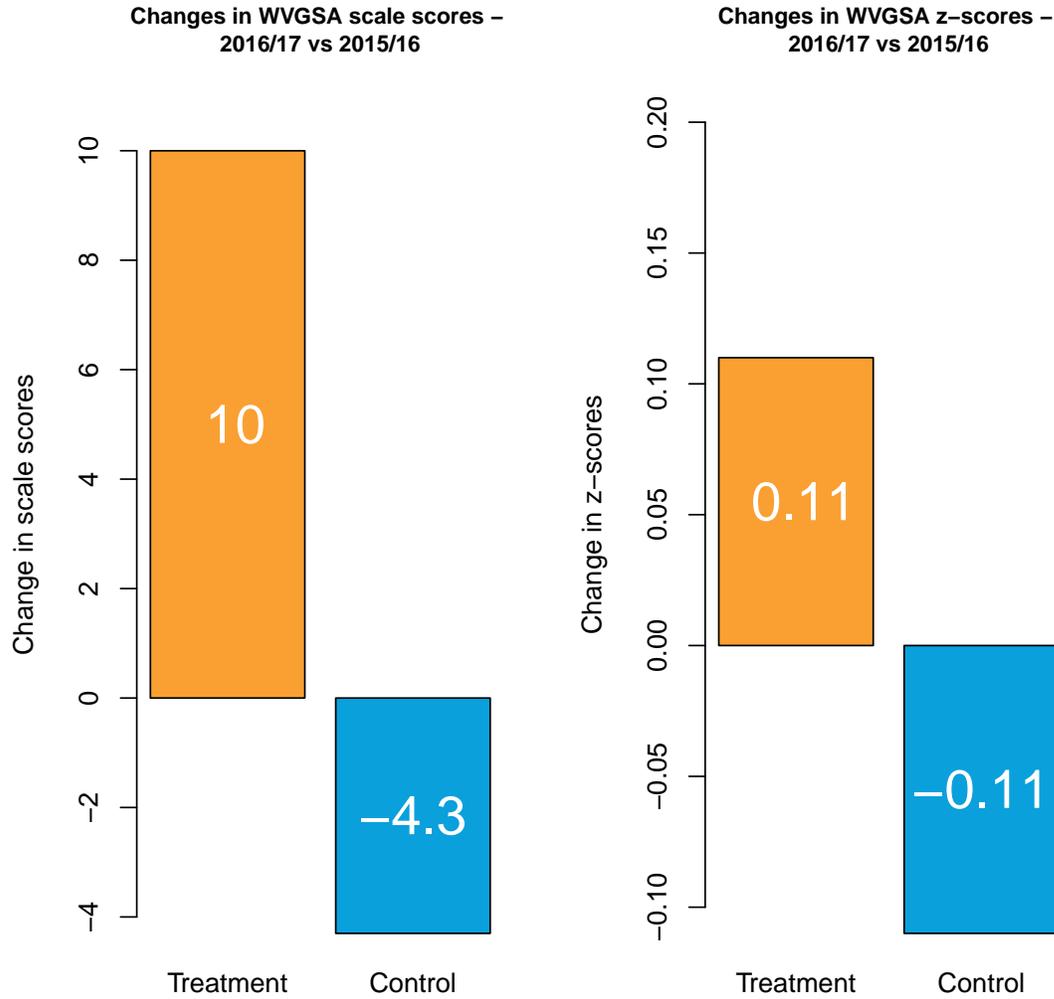


Figure 5: Changes in WVGSA Math scale scores and z-scores (See Section 3.1) for Grade-Aggregated TRT and CTRL datasets between 2015/16 and 2016/17

Further, Figure 6 shows the changes in percent of students at WVGSA Level 3 or 4 for the grade-aggregated Treatment and Control sets.

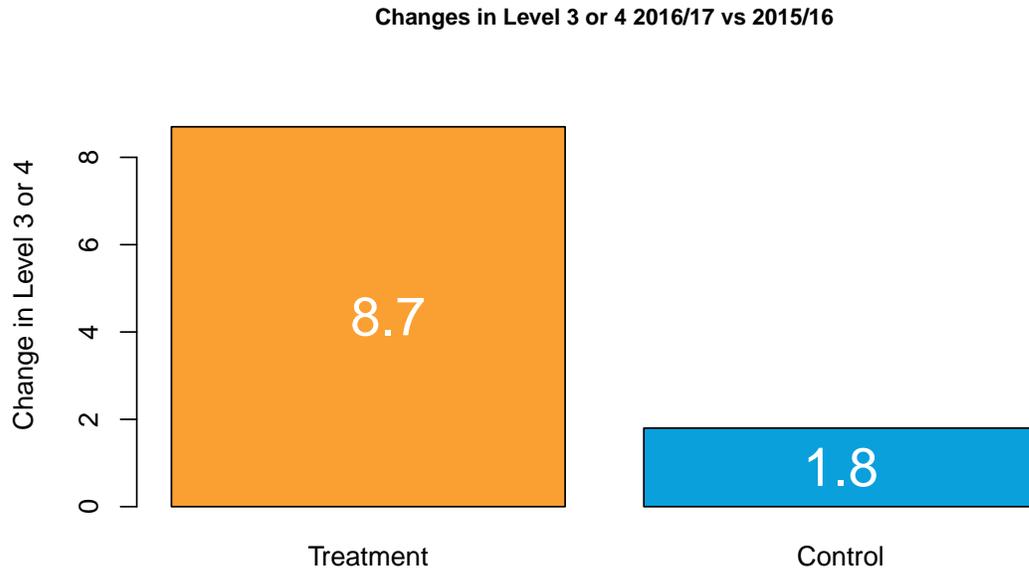


Figure 6: Changes in Level 3 or 4 for Grade-Aggregated TRT and CTRL datasets between 2015/16 and 2016/17

Table 7 shows the statistics for the *differences* in changes between TRT and CTRL (Treatment - Control) for these same WVGSA math proficiency and z-score changes as in the above figures. ¹

	Estimate	P-Value	Int.Low	Int.High
Level 3 or 4	6.95	0.02*	1.32	12.58
scale score	14.31	0.23	-9.05	37.66
Z-score	0.22	0.17	-0.10	0.55
L1	-7.10	0.00*	-11.76	-2.45
L2	0.15	0.95	-4.31	4.61
L3	3.06	0.13	-0.88	7.00
L4	3.89	0.07	-0.30	8.09

Table 7: Statistics for the Differential Changes in Math Scores Growth (TRT - CTRL)

¹* statistically significant $p < 0.05$

Finally, Figure 7 shows the changes in mean percentile ranking between TRT and CTRL.

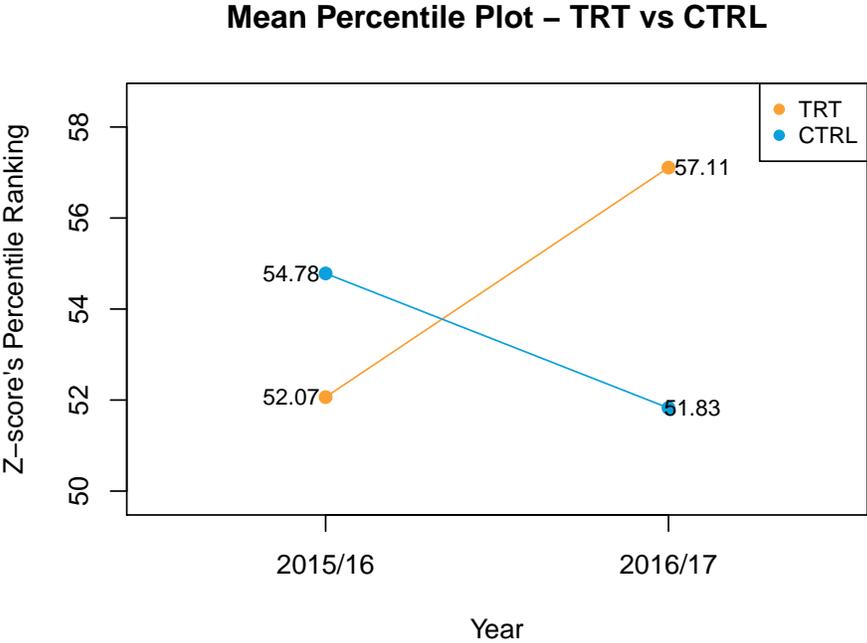


Figure 7: Changes in Percentile Ranking for TRT and CTRL Datasets between 2015/16 and 2016/17

3.5 Grade-Level Analysis

3.5.1 Grade Level Result Tables

The following tables (Table 8, 9, and 10) 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	Scale score	Z-score	Percentile	L1	L2	L3	L4	Level 3 or 4	ST Math Per Prog.
TRT.15.16	15	15	678	2395.8	-0.23	49.87	21.62	32.62	30.42	15.35	45.77	-
TRT.16.17	15	15	678	2413.6	0.17	58.60	19.37	24.41	30.34	25.89	56.23	70.95
TRT.Delta	-	-	-	17.8	0.40	8.73	-2.25	-8.21	-0.08	10.55	10.47	-
CTRL.15.16	15	15	881	2413.5	0.09	54.00	22.01	30.79	32.07	15.11	47.17	-
CTRL.16.17	15	15	881	2395.8	-0.14	46.87	25.42	27.86	30.02	16.68	46.70	-
CTRL.Delta	-	-	-	-17.7	-0.23	-7.13	3.41	-2.93	-2.05	1.57	-0.47	-

Table 8: Grade 3 - Yearly Math Performance and Counts for TRT and CTRL Datasets

	# Grades	# Schools	# Students	Scale score	Z-score	Percentile	L1	L2	L3	L4	Level 3 or 4	ST Math Per Prog.
TRT.15.16	15	15	701	2431.8	-0.07	53.33	25.86	40.43	22.67	11.02	33.69	-
TRT.16.17	15	15	701	2422.1	-0.33	51.53	15.97	39.73	29.56	14.71	44.27	72.71
TRT.Delta	-	-	-	-9.8	-0.26	-1.80	-9.89	-0.70	6.89	3.69	10.58	-
CTRL.15.16	15	15	739	2442.1	0.09	57.33	24.42	37.57	26.59	11.42	38.01	-
CTRL.16.17	15	15	739	2439.8	-0.03	52.33	24.57	34.47	28.97	12.01	40.98	-
CTRL.Delta	-	-	-	-2.3	-0.12	-5.00	0.15	-3.11	2.38	0.59	2.97	-

Table 9: Grade 4 - Yearly Math Performance and Counts for TRT and CTRL Datasets

	# Grades	# Schools	# Students	Scale score	Z-score	Percentile	L1	L2	L3	L4	Level 3 or 4	ST Math Per Prog.
TRT.15.16	16	16	838	2458.0	0.08	52.94	35.07	32.78	18.24	13.89	32.13	-
TRT.16.17	16	16	838	2479.1	0.26	60.94	29.83	32.74	20.19	17.23	37.42	67.48
TRT.Delta	-	-	-	21.1	0.19	8.00	-5.24	-0.04	1.95	3.34	5.29	-
CTRL.15.16	16	16	709	2458.7	0.08	53.12	30.80	34.88	20.08	14.26	34.34	-
CTRL.16.17	16	16	709	2464.9	0.09	56.00	31.27	31.71	19.31	17.73	37.04	-
CTRL.Delta	-	-	-	6.2	0.01	2.88	0.47	-3.16	-0.77	3.46	2.69	-

Table 10: Grade 5 - Yearly Math Performance and Counts for TRT and CTRL Datasets

3.5.2 Grade-Level Analysis of Changes in Math Level 3 or 4

Figure 8 shows the difference in the growth of percentages of students at math Level 3 or 4, for the TRT and CTRL datasets, disaggregated by grade:

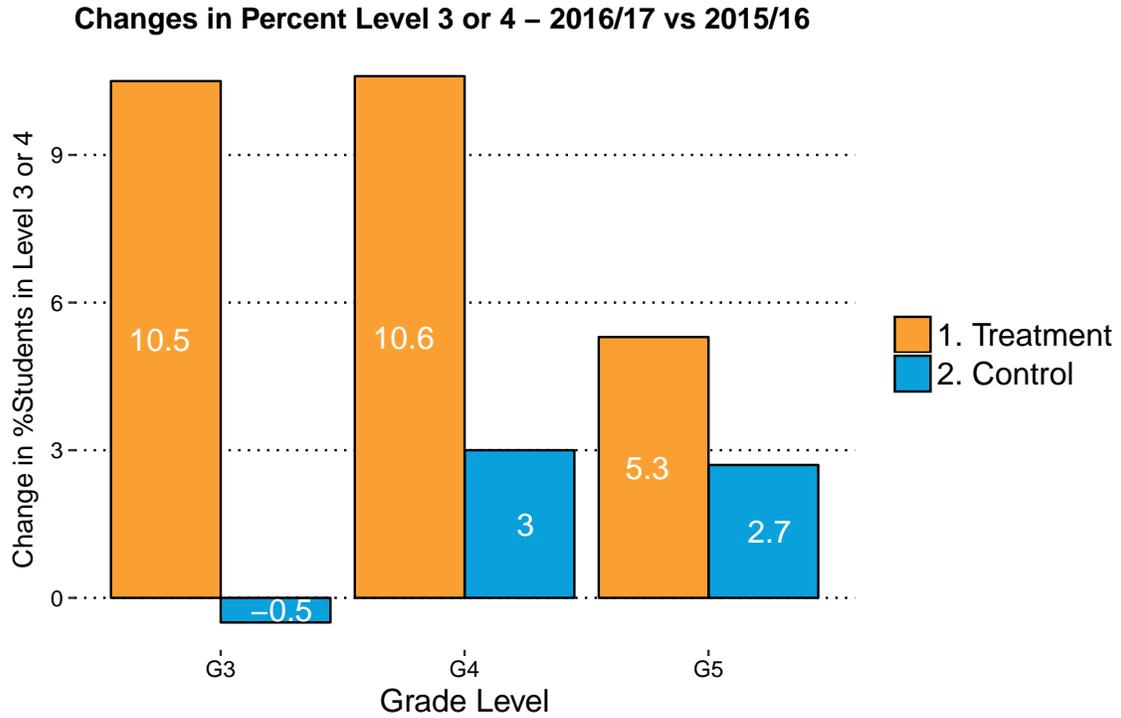


Figure 8: Changes in Percent of Students at Level 3 or 4 for TRT and CTRL Datasets between 2015/16 and 2016/17

Table 11 shows the statistics for the *differences* in changes between TRT and CTRL (Treatment - Control) for these same Level 3 or 4 math proficiency changes as shown in Figure 8.

	Estimate	P-Value	Int.Low	Int.High
Grade 3	10.94	0.01*	3.31	18.57
Grade 4	7.61	0.07	-0.56	15.77
Grade 5	2.59	0.69	-10.66	15.84

Table 11: Statistics for the Differential Changes in Level 3 or 4, (TRT - CTRL)

3.5.3 Grade-Level Analysis of Changes in WVGSA Math scale scores

Figure 9 shows the changes in the grade-mean math scale scores of students for the TRT and CTRL datasets, disaggregated by grade:

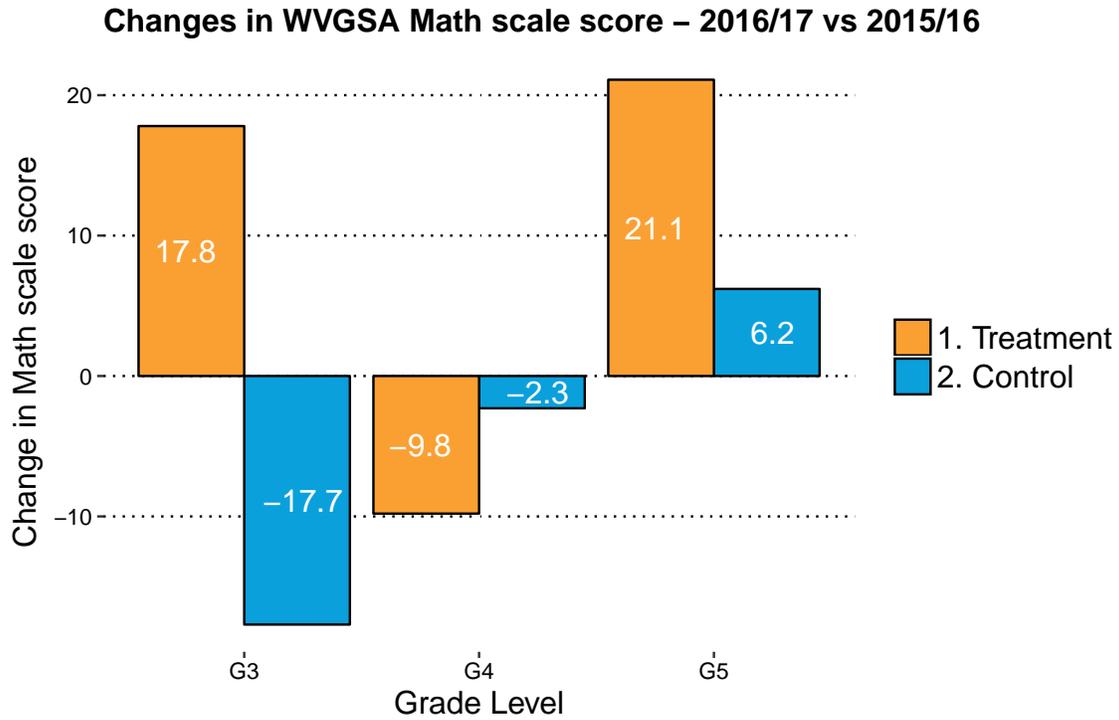


Figure 9: Changes in Grade-Mean WVGSA Math scale score for TRT and CTRL Datasets between 2015/16 and 2016/17

Table 12 shows the statistics for the differences between TRT and CTRL (Treatment - Control) for these same WVGSA math scale score changes as shown in Figure 9.

	Estimate	P-Value	Int.Low	Int.High
Grade 3	35.49	0.04*	2.52	68.47
Grade 4	-7.49	0.66	-41.91	26.93
Grade 5	14.88	0.57	-37.82	67.58

Table 12: Statistics for the Differential Changes in WVGSA Math scale scores Growth, (TRT - CTRL)

3.5.4 Grade-Level Analysis of Changes in WVGSA z-scores of scale scores

Figure 10 shows the changes in the grade-mean z-scores of students for the TRT and CTRL datasets, disaggregated by grade:

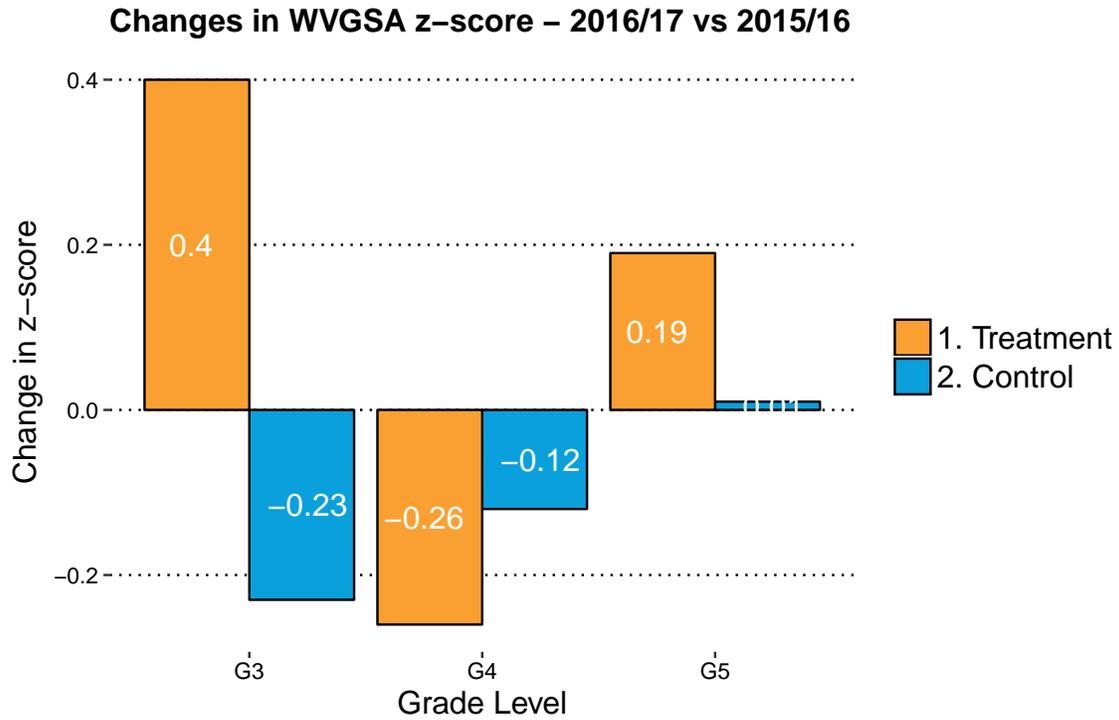


Figure 10: Changes in Grade-Mean WVGSA z-score (See Section 3.1) for TRT and CTRL Datasets between 2015/16 and 2016/17

Table 13 shows the statistics for the differences between TRT and CTRL (Treatment - Control) for these same WVGSA z-score changes as shown in Figure 10.

	Estimate	P-Value	Int.Low	Int.High
Grade 3	0.63	0.04*	0.04	1.23
Grade 4	-0.14	0.63	-0.74	0.46
Grade 5	0.18	0.49	-0.34	0.70

Table 13: Statistics for the Differential Changes in WVGSA z-scores (See Section 3.1) Growth, (TRT - CTRL)

4 Effect Size

The following table shows the effect sizes for Level 3 or 4, WVGSA scale score, and accompanying z-score.

	Scale score Effect Size	Z-score Effect Size	Level 3 or 4 Effect Size
Grade 3	1.16	1.13	0.97
Grade 4	-0.14	-0.16	0.55
Grade 5	0.21	0.57	0.20
All Grades	0.25	0.37	0.51

Table 14: Cohen's d Effect Size

5 Findings Summary

West Virginia grades 3, 4, and 5 using ST Math for the year 2016/17 averaged 68.5% ST Math Progress. 46/49 grades (94%) averaged covering more than 50% of ST Math content. Statistically significant differences were found in this analysis for both grade-aggregated and individual grade levels. Looking at Table 7, a statistically significant difference was found for grade-aggregated Level 3 or 4 proficiency levels, with a 6.95 point favorable differential for the ST Math treatment set. Referring to Table 11, a statistically significant difference was found for grade 3 Level 3 or 4 proficiency levels, with an estimate of 10.94 in favor of the ST Math treatment set. Looking at Table 13, grade 3 ST math treatment set outperformed their matched controls for WVGSA z-scores with a statistically significant difference of 0.63.

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 15) and controls (Table 16) separately.

	# Grades	# Schools	# Students	Scale score	Z-score	Percentile	L1	L2	L3	L4	Level 3 or 4	ST Math Per Comp.
Grade 3 (15.16)	15	15	678	2395.8	-0.23	49.87	21.62	32.62	30.42	15.35	45.77	-
Grade 4 (15.16)	15	15	701	2431.8	-0.07	53.33	25.86	40.43	22.67	11.02	33.69	-
Grade 5 (15.16)	16	16	838	2458.0	0.08	52.94	35.07	32.78	18.24	13.89	32.13	-
All Grades (15.16)	46	17	2217	2429.2	-0.07	52.07	27.68	35.22	23.66	13.43	37.09	-
Grade 3 (16.17)	15	15	678	2413.6	0.17	58.60	19.37	24.41	30.34	25.89	56.23	70.95
Grade 4 (16.17)	15	15	701	2422.1	-0.33	51.53	15.97	39.73	29.56	14.71	44.27	72.71
Grade 5 (16.17)	16	16	838	2479.1	0.26	60.94	29.83	32.74	20.19	17.23	37.42	67.48
All Grades (16.17)	46	17	2217	2439.2	0.04	57.11	21.90	32.30	26.56	19.23	45.79	70.32

Table 15: TRT Grades Detail Sorted by Year

	# Grades	# Schools	# Students	Scale score	Z-score	Percentile	L1	L2	L3	L4	Level 3 or 4	ST Math Per Comp.
Grade 3 (15.16)	15	15	881	2413.5	0.09	54.00	22.01	30.79	32.07	15.11	47.17	-
Grade 4 (15.16)	15	15	739	2442.1	0.09	57.33	24.42	37.57	26.59	11.42	38.01	-
Grade 5 (15.16)	16	16	709	2458.7	0.08	53.12	30.80	34.88	20.08	14.26	34.34	-
All Grades (15.16)	46	44	2329	2438.5	0.09	54.78	25.85	34.42	26.11	13.61	39.72	-
Grade 3 (16.17)	15	15	881	2395.8	-0.14	46.87	25.42	27.86	30.02	16.68	46.70	-
Grade 4 (16.17)	15	15	739	2439.8	-0.03	52.33	24.57	34.47	28.97	12.01	40.98	-
Grade 5 (16.17)	16	16	709	2464.9	0.09	56.00	31.27	31.71	19.31	17.73	37.04	-
All Grades (16.17)	46	44	2329	2434.2	-0.03	51.83	27.18	31.35	25.95	15.52	41.47	-

Table 16: 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.

PID	IID	District	School Name	GRADE
1153789	ALB1XX	Kanawha	Alban Elementary School	3, 4, 5
1153791	ALU1YA	Kanawha	Alum Creek Elementary School	3, 4, 5
1153818	AND1Y0	Kanawha	Andrews Heights Elementary School	3, 4, 5
1153820	ANN1XX	Kanawha	Anne Bailey Elementary School	3, 4, 5
1153935	CED1XJ	Kanawha	Cedar Grove Elementary School	3, 4, 5
1153947	CEN1XX	Kanawha	Central Elementary School	3, 4, 5
1153959	CHA1YA	Kanawha	Chamberlain Elementary School	3, 4, 5
1154020	CRO1YB	Kanawha	Cross Lanes Elementary School	3, 4, 5
1153870	ELK1YA	Kanawha	Elk Elementary Center	3, 4, 5
1153961	WES1YA	Kanawha	Mary C. Snow West Side Elementary	4
1154501	MID1XH	Kanawha	Midland Trail Elementary School	3, 4, 5
1154654	POI1YB	Kanawha	Point Harmony Elementary School	3, 4, 5
1154642	PRA1XW	Kanawha	Pratt Elementary School	3, 5
1154707	RUF1YB	Kanawha	Ruffner Elementary School	3, 4, 5
1154719	RUT1YA	Kanawha	Ruthlawn Elementary School	3, 4, 5
1154745	SHA1XT	Kanawha	Sharon Dawes Elementary School	3, 4, 5
1154800	SIS1YB	Kanawha	Sissonville Middle School	5

Table 17: 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.

PID	District	School Name	GRADE
1150141	Berkeley	Rosemont Elementary School	3
1150232	Boone	Brookview Elementary	3
1150335	Boone	Ramage Elementary	5
1150438	Boone	Whitesville Elementary School	4
1150517	Braxton	Frametown Elementary School	4
1150543	Braxton	Sutton Elementary School	5
1150634	Brooke	Colliers Primary School	3
1150907	Cabell	Davis Creek Elementary School	5
1150971	Cabell	Explorer Academy	4
1151224	Cabell	Ona Elementary School	4
1151145	Cabell	Southside Elementary	3
1150775	Cabell	Village Of Barboursville Elementary Sch	5
1151822	Fayette	Divide Elementary	5
1151884	Fayette	Gatewood Elementary	3
1151896	Fayette	Gaulley Bridge Elementary	3
1152606	Hampshire	Augusta Elementary School	3, 4
1153234	Harrison	Norwood Elementary School	5
1153325	Harrison	Simpson Elementary School	3
1153454	Jackson	Cottageville Elementary School	4
1155311	Lincoln	Hamlin Pk-8	3
1155218	Lincoln	Harts Pk-8	5
1155397	Lincoln	Midway Elementary School	5
1155828	Logan	West Chapmanville Elementary School	4
1156420	Marion	East Dale Elementary School	4
1157228	Mason	Pt. Pleasant Intermediate	4
1157319	Mercer	Athens Elementary School	5
1157541	Mercer	Oakvale School	5
2179358	Mercer	Straley Elementary School	3
1157682	Mineral	Burlington Primary	4
1758561	Mineral	Keyser Primary School	3
1158870	Monroe	Mountain View Elementary & Middle School	4
1158909	Monroe	Peterstown Elementary School	4
5342932	Morgan	Warm Springs Intermediate School	3
1160122	Pendleton	Franklin Elementary School	5
1160172	Pleasants	Belmont Elementary School	3
1160847	Raleigh	Crab Orchard Elementary	4
1161097	Raleigh	Marsh Fork Elementary	5
1161372	Randolph	Coalton Elementary School	5
1523885	Randolph	North Elementary School	4
1162053	Taylor	West Taylor Elementary School	4
1162405	Upshur	Tennerton Elementary	5
1162467	Wayne	Ceredo Elementary School	3
1173868	Wood	Lubeck Elementary School	5
1163459	Wood	Williamstown Elementary School	5, 3

Table 18: Matched Control Schools (CTRL Dataset)