

# Massachusetts Math Outcomes Analysis 2018/19

**Grade Levels: 3, 4, 5**  
**ST Math Program: Gen-4**  
**Analysis Type: Z-score of scale score**  
**Treatment-Years: 2018/19**  
**Baseline-Year: 2017/18**  
**Subgroup: All**



**Jessica Guise**

© 2019-11-14

### **Abstract**

This analysis evaluates grades using ST Math in Massachusetts in 2018/19. 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 49 grades, consisting of grades 3, 4, and 5 at 26 schools, with an average baseline of 49% in Meeting or Exceeding Expectations proficiency levels (refer to Figures 2 and 3 to see how your schools compare to those analyzed in this report). They were matched to 49 similar, randomly selected control grades at 48 schools that never used ST Math. Grade-wise growth in math proficiency was evaluated (i.e. growth in same grade, same school, from 2017/18 to 2018/19) 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 5.63 points at Meeting or Exceeding Expectations, 3.41 points at Meeting Expectations, 2.39 points at Exceeding Expectations, and z-score of 0.3.

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>   | <b>5</b>  |
| 1.1      | Background . . . . .  | 5         |
| 1.2      | Program Description . . . . .   | 5         |
| <b>2</b> | <b>Data Collection</b>  | <b>6</b>  |
| 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 . . . . .   | 6         |
| 2.3      | Control Grades Pool and Selection . . . . .                                   | 7         |
| <b>3</b> | <b>Data Analysis</b>  | <b>8</b>  |
| 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 Meeting or Exceeding Expectations     | 18        |
| 3.5.3    | Grade-Level Analysis of Changes in MCAS Math scale scores . . . . .           | 19        |
| 3.5.4    | Grade-Level Analysis of Changes in MCAS Z-scores of scale scores . . . . .    | 20        |
| <b>4</b> | <b>Effect Size</b>  | <b>21</b> |
| <b>5</b> | <b>Findings Summary</b>   | <b>21</b> |
| <b>6</b> | <b>Confounders</b>  | <b>21</b> |
| <b>7</b> | <b>Reference Tables Grouped By School Year</b>                                | <b>22</b> |
| <b>8</b> | <b>Lists of Schools</b>   | <b>23</b> |
| 8.1      | Treatment Schools . . . . .   | 23        |
| 8.2      | Control Schools . . . . .   | 24        |

## List of Figures

|    |   |    |
|----|---|----|
| 1  | Histogram of ST Math Percent Progress for $\geq 85\%$ Enrollment Grades 2018/19 . . . . .   | 9  |
| 2  | Baseline Year Density Plots Showing Math Scores Match between TRT and CTRL - 2017/18 . . . . .  | 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 2017/18 and 2018/19 . . . . .                                 | 13 |
| 5  | Changes in MCAS Math scale scores and Z-scores (See Section 3.1) for Grade-Aggregated TRT and CTRL datasets between 2017/18 and 2018/19 . . . . . | 14 |
| 6  | Changes in Meeting or Exceeding Expectations for Grade-Aggregated TRT and CTRL datasets between 2017/18 and 2018/19 . . . . .                     | 15 |
| 7  | Changes in Percentile Ranking for TRT and CTRL Datasets between 2017/18 and 2018/19 . . . . .   | 16 |
| 8  | Changes in Percent of Students at Meeting or Exceeding Expectations for TRT and CTRL Datasets between 2017/18 and 2018/19 . . . . .               | 18 |
| 9  | Changes in Grade-Mean MCAS Math scale score for TRT and CTRL Datasets between 2017/18 and 2018/19 . . . . .                                       | 19 |
| 10 | Changes in Grade-Mean MCAS Z-score (See Section 3.1) for TRT and CTRL Datasets between 2017/18 and 2018/19 . . . . .                              | 20 |

## List of Tables

|    |   |    |
|----|---|----|
| 1  | Proficiency Level Naming . . . . .  | 6  |
| 2  | Descriptive Statistics of ST Math Percent Progress for $\geq 85$ percent Enrollment Grades . . . . .      | 9  |
| 3  | Number of ST Math Grades with $\geq 85$ percent Enrollment and with $\geq 40$ 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 Meeting or Exceeding Expectations, (TRT - CTRL) . . . . .      | 18 |
| 12 | Statistics for the Differential Changes in MCAS Math scale scores Growth, (TRT - CTRL) . . . . .          | 19 |
| 13 | Statistics for the Differential Changes in MCAS Z-scores (See Section 3.1) Growth, (TRT - CTRL) . . . . . | 20 |
| 14 | Cohen's d Effect Size . . . . .   | 21 |
| 15 | TRT Grades Detail Sorted by Year . . . . .  | 22 |
| 16 | CTRL Grades Detail Sorted by Year . . . . .   | 22 |
| 17 | Treatment Schools (TRT Dataset) . . . . .   | 23 |
| 18 | Matched Control Schools (CTRL Dataset) . . . . .  | 24 |

# 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 MCAS Meeting or Exceeding Expectations percentages. The treatment grades used the ST Math program for 1 year, beginning in the 2018/19 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 (2017/18), and did not use ST Math in 2018/19. The treatment grades’ selection pool was all schools using ST Math in grades 3, 4, and 5 in Massachusetts. The control grades’ pool was all schools not using ST Math in grades 3, 4, and 5 in Massachusetts. This study method measures effectiveness of the ST Math program when nominally implemented.

## 1.2 Program Description

Spatial-Temporal Math (ST Math) is game-based, instructional software for K–12 students, created by the MIND Research Institute (MIND). The purpose of the program is to boost math comprehension through visual learning. The ST Math software games begin without language or symbol abstractions by posing math problems as purely visual puzzles. In this way, three objectives 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. Interactive, animated visual manipulatives provide informative feedback on student solutions. A score of 100 percent on a game level comprised of 4-12 puzzles is required for progression through the levels. Failure requires a re-play of the level, via a new quasi-random set of puzzles. In this way, progression is self-paced.

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.

For students to achieve nominal progress through the program, there is a recommended time-on-task requirement of 90 minutes per week over about 30 weeks. Consistent application of 90 minutes per week throughout the school year is normally sufficient to result in a grade’s average ST Math content coverage exceeding 50% by year-end. In this study, we include grades that have achieved 40% or more content coverage (Progress) by April 15th.

This is a passive study with no experimental setup or extraordinary communications to any schools. All schools in this study therefore received normal program implementation support through the year from MIND support managers. This support includes bundled startup services of approximately 2-4 hours of training either in-person or online, access to live webinars, regular online and push reports on

usage and progress, email/phone helpdesk, and proactive monitoring for gaps or issues by MIND support representatives.

MIND Research Institute initiated, funded, and exercised editorial control over this study.

## 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 Massachusetts's proficiency level descriptions:

| Proficiency Level | State Proficiency Level Name   |
|-------------------|--------------------------------|
| L1                | Not Meeting Expectations       |
| L2                | Partially Meeting Expectations |
| L3                | Meeting Expectations           |
| L4                | Exceeding Expectations         |

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 Massachusetts. From these schools, every grade that had used the ST Math program only for the year 2018/19 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 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. Massachusetts's standardized math assessment (MCAS). 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 Massachusetts 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 40% by April.

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 40% 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 Massachusetts. Though they are randomly selected, they are also matched to be similar to the Treatment grades' math attributes and demographics during the baseline 2017/18 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 Massachusetts is evaluated for Enrollment percentage and Progress percentage parameters. A filtered Treatment set (TRT) of all ST Math grades with  $\geq 85\%$  Enrollment and  $\geq 40\%$  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 Massachusetts 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 – 2018/19**

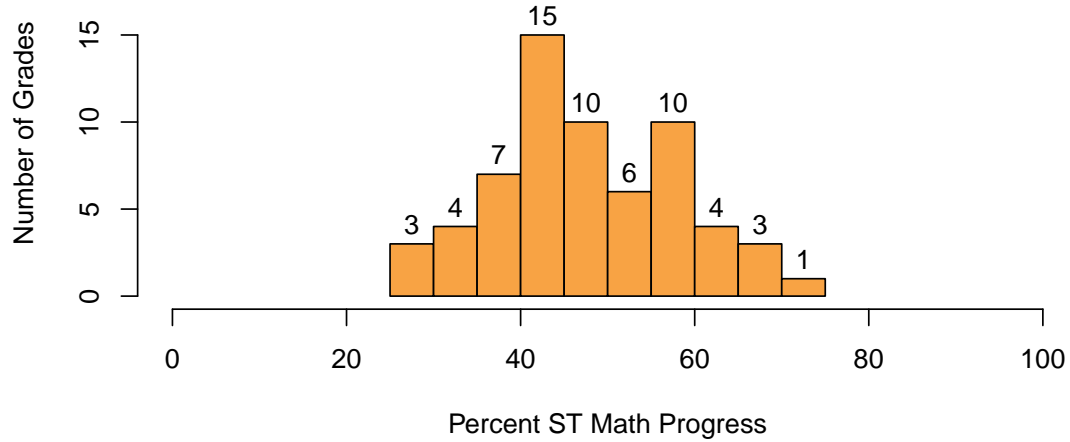


Figure 1: Histogram of ST Math Percent Progress for  $\geq 85\%$  Enrollment Grades 2018/19

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  $\geq 40\%$  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 | 26.2 | 70.4 | 47.8    | 10.5 |

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

|   |    |
|---|----|
| Grades with $\geq 85\%$ Enrollment:           | 63 |
| Grades with in addition $\geq 40\%$ Progress: | 49 |

Table 3: Number of ST Math Grades with  $\geq 85\%$  Enrollment and with  $\geq 40\%$  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 40% Progress filter. Other rows in the table indicate counts of numbers of students (2018/19 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                               | 27      | 29      | 22      | 78    |
| ST Math Using Schools                              | 27      | 29      | 22      | 35    |
| ST Math Students                                   | 2055    | 2320    | 1917    | 6292  |
| ST Math Grades (Enroll $\geq$ 85%)                 | 22      | 24      | 17      | 63    |
| TRT Grades (Enroll $\geq$ 85% & Prog $\geq$ 40%)   | 17      | 20      | 12      | 49    |
| TRT Schools (Enroll $\geq$ 85% & Prog $\geq$ 40%)  | 17      | 20      | 12      | 26    |
| TRT Students (Enroll $\geq$ 85% & Prog $\geq$ 40%) | 1526    | 1661    | 1147    | 4334  |
| CTRL Grades  | 17      | 20      | 12      | 49    |
| CTRL Schools                                       | 17      | 20      | 12      | 48    |
| CTRL Students                                      | 1113    | 1478    | 979     | 3570  |

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 MCAS Math scale scores (left plot) and baseline percent students at MCAS Meeting or Exceeding Expectations (right plot) for treatment grades overlaid on control grades, showing the closeness of the match obtained between Treatment and Control sets of grades in the baseline year, 2017/18.

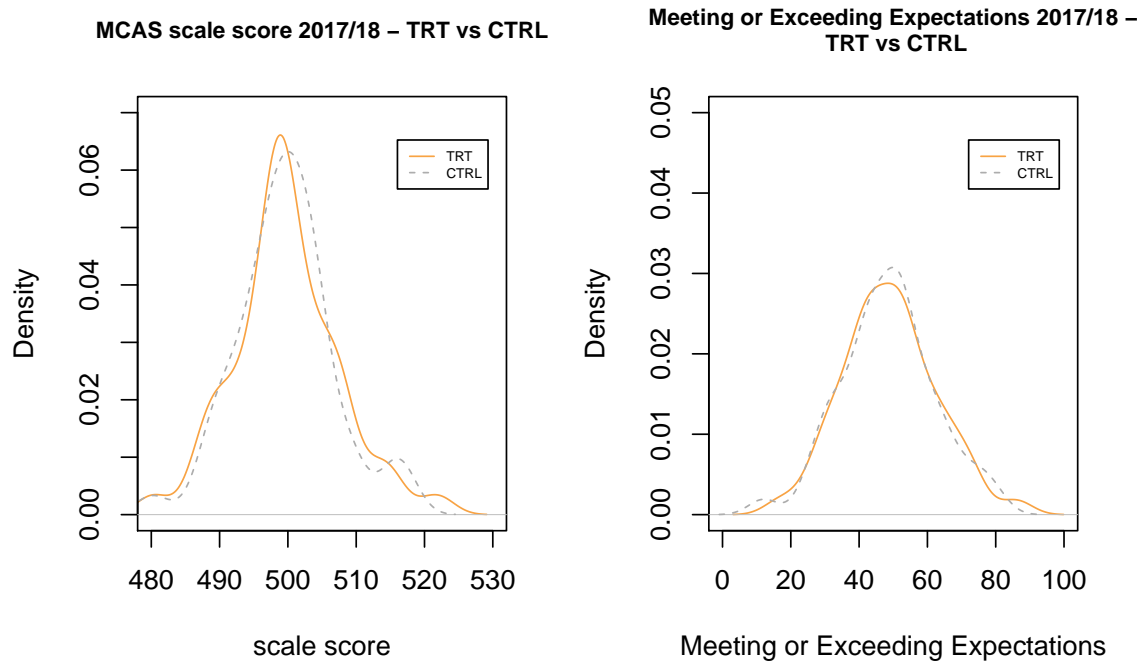


Figure 2: Baseline Year Density Plots Showing Math Scores Match between TRT and CTRL - 2017/18

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.

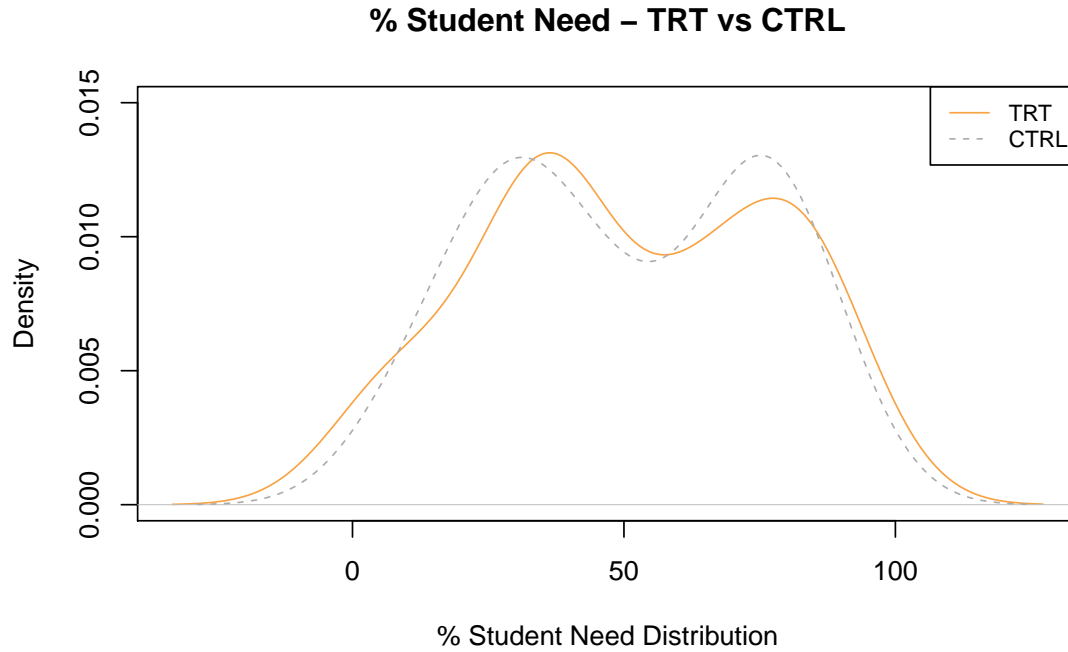


Figure 3: Baseline Year Density Plot Showing Student Need Match between TRT and CTRL

Table 5 shows the difference of the means of Treatment versus Control in the baseline year, with accompanying p-values, for percent Meeting or Exceeding Expectations, 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 | Effect Size |
|---|-----------|---------|------------|----------|----------|---------|-------------|
| Meeting or Exceeding Expectations - 2017/18 | 48.92     | 13.81   | 48.49      | 13.82    | 0.43     | 0.88    | 0.03        |
| Scale score - 2017/18                       | 499.78    | 7.77    | 499.66     | 7.20     | 0.12     | 0.93    | 0.02        |
| Percent Free or Reduced Lunch               | 50.71     | 26.75   | 50.61      | 25.48    | 0.10     | 0.98    | 0.00        |

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 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    | Meeting or Exceeding Expectations | ST Math Per Comp. |
|------------|----------|-----------|------------|-------------|---------|------------|-------|-------|-------|-------|-----------------------------------|-------------------|
| TRT.17.18  | 49       | 26        | 4202       | 499.8       | 0.10    | 52.82      | 10.06 | 41.04 | 41.47 | 7.37  | 48.92                             | -                 |
| TRT.18.19  | 49       | 26        | 4140       | 502.1       | 0.27    | 57.78      | 9.00  | 36.76 | 45.76 | 8.59  | 54.31                             | 51.6              |
| TRT.Delta  | -        | -         | -          | 2.4         | 0.17    | 4.96       | -1.06 | -4.29 | 4.29  | 1.22  | 5.39                              | -                 |
| CTRL.17.18 | 49       | 48        | 3589       | 499.7       | 0.09    | 52.73      | 9.90  | 41.57 | 41.22 | 7.22  | 48.49                             | -                 |
| CTRL.18.19 | 49       | 48        | 3570       | 498.9       | -0.04   | 48.76      | 10.49 | 41.41 | 42.10 | 6.06  | 48.24                             | -                 |
| CTRL.Delta | -        | -         | -          | -0.8        | -0.12   | -3.98      | 0.59  | -0.16 | 0.88  | -1.16 | -0.24                             | -                 |

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).

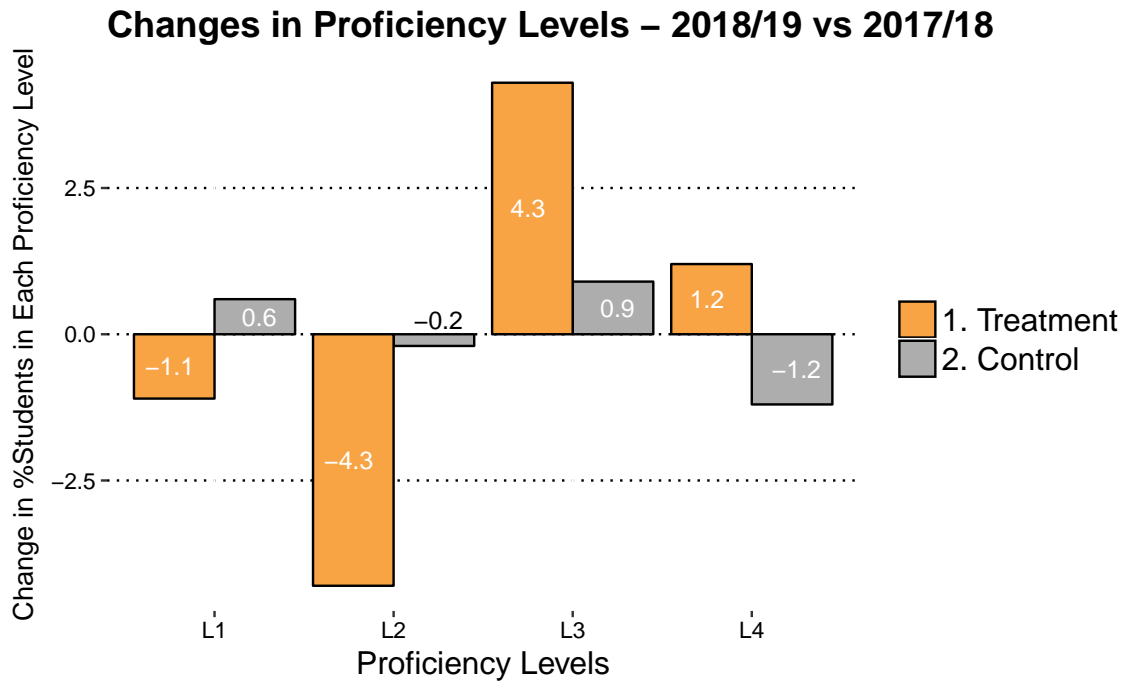


Figure 4: Change at each Proficiency Level for Grade-Aggregated TRT and CTRL Datasets between 2017/18 and 2018/19

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

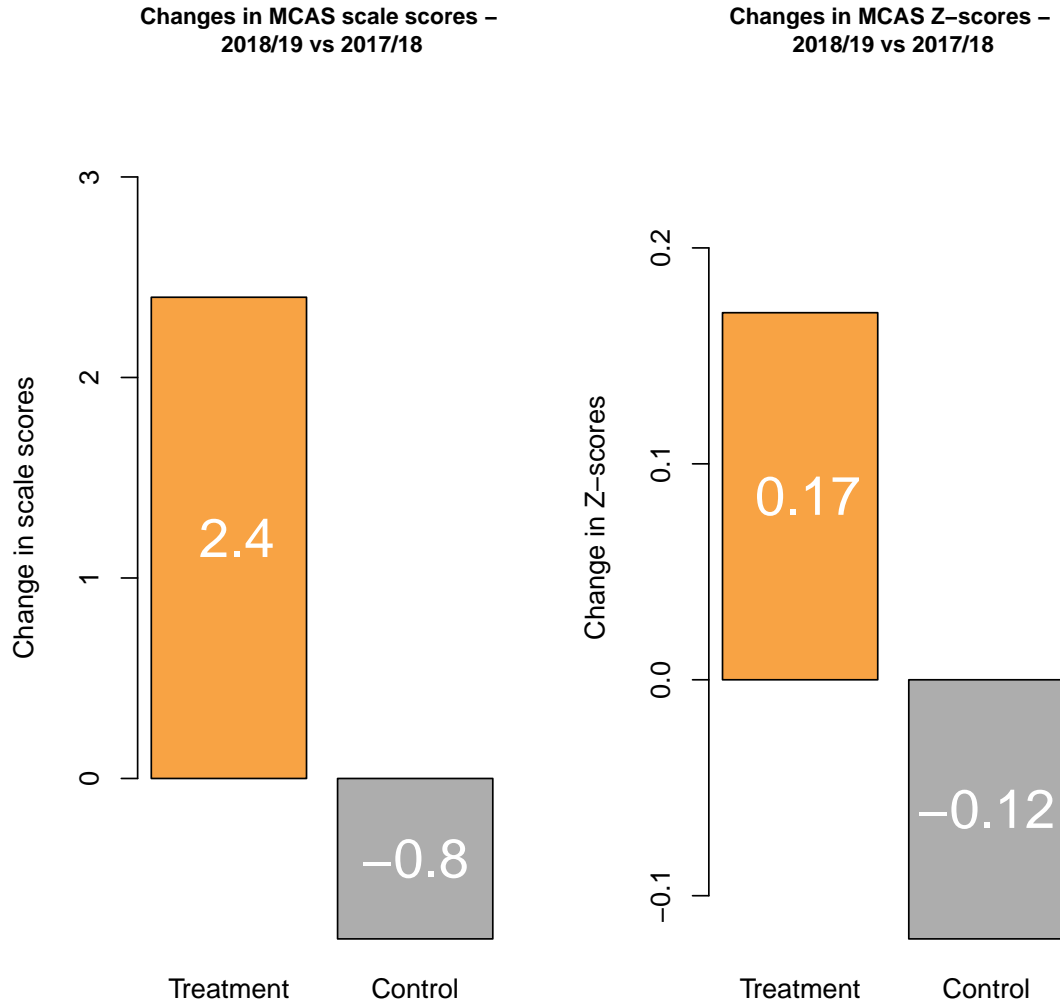


Figure 5: Changes in MCAS Math scale scores and Z-scores (See Section 3.1) for Grade-Aggregated TRT and CTRL datasets between 2017/18 and 2018/19

Further, Figure 6 shows the changes in percent of students at MCAS Meeting or Exceeding Expectations for the grade-aggregated Treatment and Control sets.

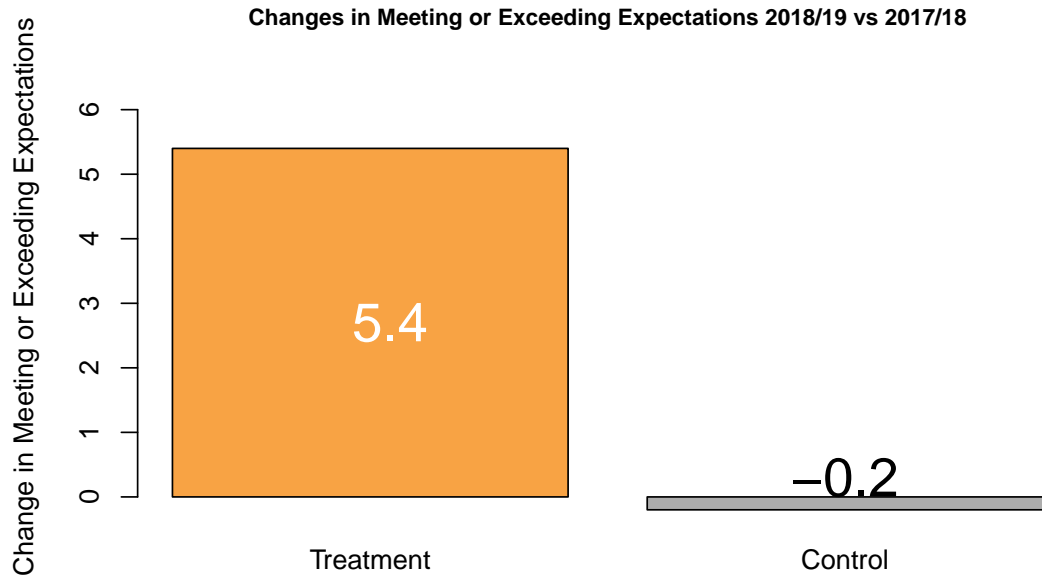


Figure 6: Changes in Meeting or Exceeding Expectations for Grade-Aggregated TRT and CTRL datasets between 2017/18 and 2018/19

Finally, Table 7 shows the statistics for the *differences* in changes between TRT and CTRL (Treatment - Control) for these same MCAS math proficiency and scale score changes as in the above figures.<sup>1</sup>

|                                   | Estimate | P-Value | Int.Low | Int.High |
|-----------------------------------|----------|---------|---------|----------|
| Meeting or Exceeding Expectations | 5.63     | 0.02*   | 0.85    | 10.41    |
| scale score                       | 3.13     | 0.01*   | 0.79    | 5.46     |
| Z-score                           | 0.30     | 0.01*   | 0.07    | 0.52     |
| L1                                | -1.65    | 0.24    | -4.42   | 1.12     |
| L2                                | -4.12    | 0.05*   | -8.18   | -0.06    |
| L3                                | 3.41     | 0.09    | -0.59   | 7.41     |
| L4                                | 2.39     | 0.07    | -0.17   | 4.95     |

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

<sup>1</sup>\* statistically significant p<0.05



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

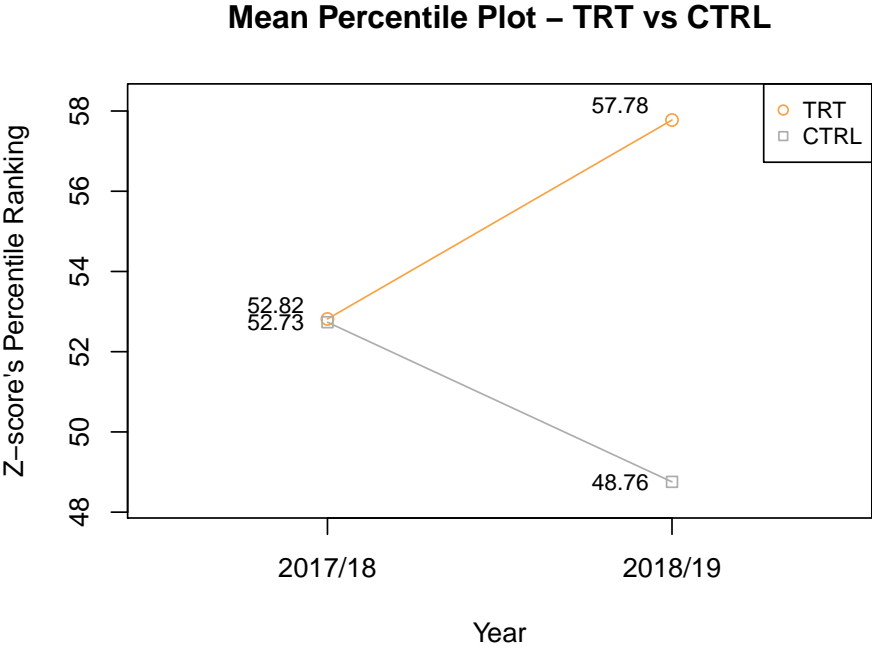


Figure 7: Changes in Percentile Ranking for TRT and CTRL Datasets between 2017/18 and 2018/19

### 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    | Meeting or Exceeding Expectations | ST Math Per Prog. |
|------------|----------|-----------|------------|-------------|---------|------------|-------|-------|-------|-------|-----------------------------------|-------------------|
| TRT.17.18  | 17       | 17        | 1388       | 500.5       | 0.02    | 49.76      | 10.88 | 39.41 | 41.71 | 8.06  | 49.76                             | -                 |
| TRT.18.19  | 17       | 17        | 1411       | 501.8       | 0.21    | 55.41      | 10.24 | 38.29 | 41.82 | 9.71  | 51.47                             | 49.16             |
| TRT.Delta  | -        | -         | -          | 1.3         | 0.19    | 5.65       | -0.65 | -1.12 | 0.12  | 1.65  | 1.71                              | -                 |
| CTRL.17.18 | 17       | 17        | 1162       | 499.9       | -0.04   | 48.35      | 11.41 | 39.53 | 40.76 | 8.12  | 49.00                             | -                 |
| CTRL.18.19 | 17       | 17        | 1113       | 497.7       | -0.17   | 44.71      | 13.18 | 41.12 | 39.12 | 6.71  | 45.94                             | -                 |
| CTRL.Delta | -        | -         | -          | -2.2        | -0.13   | -3.65      | 1.76  | 1.59  | -1.65 | -1.41 | -3.06                             | -                 |

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    | Meeting or Exceeding Expectations | ST Math Per Prog. |
|------------|----------|-----------|------------|-------------|---------|------------|-------|-------|-------|-------|-----------------------------------|-------------------|
| TRT.17.18  | 20       | 20        | 1670       | 499.7       | 0.15    | 54.75      | 11.70 | 38.55 | 40.80 | 8.85  | 49.70                             | -                 |
| TRT.18.19  | 20       | 20        | 1599       | 502.3       | 0.28    | 56.65      | 10.00 | 34.90 | 45.65 | 9.60  | 55.30                             | 53.52             |
| TRT.Delta  | -        | -         | -          | 2.6         | 0.13    | 1.90       | -1.70 | -3.65 | 4.85  | 0.75  | 5.60                              | -                 |
| CTRL.17.18 | 20       | 20        | 1462       | 500.2       | 0.19    | 56.35      | 10.40 | 40.05 | 40.85 | 8.70  | 49.50                             | -                 |
| CTRL.18.19 | 20       | 20        | 1478       | 499.6       | 0.02    | 50.85      | 9.90  | 40.85 | 42.85 | 6.35  | 49.25                             | -                 |
| CTRL.Delta | -        | -         | -          | -0.6        | -0.17   | -5.50      | -0.50 | 0.80  | 2.00  | -2.35 | -0.25                             | -                 |

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   | Meeting or Exceeding Expectations | ST Math Per Prog. |
|------------|----------|-----------|------------|-------------|---------|------------|-------|-------|-------|------|-----------------------------------|-------------------|
| TRT.17.18  | 12       | 12        | 1144       | 498.9       | 0.14    | 53.92      | 6.17  | 47.50 | 42.25 | 3.92 | 46.42                             | -                 |
| TRT.18.19  | 12       | 12        | 1130       | 502.4       | 0.35    | 63.00      | 5.58  | 37.67 | 51.50 | 5.33 | 56.67                             | 51.85             |
| TRT.Delta  | -        | -         | -          | 3.5         | 0.22    | 9.08       | -0.58 | -9.83 | 9.25  | 1.42 | 10.25                             | -                 |
| CTRL.17.18 | 12       | 12        | 965        | 498.5       | 0.09    | 52.92      | 6.92  | 47.00 | 42.50 | 3.50 | 46.08                             | -                 |
| CTRL.18.19 | 12       | 12        | 979        | 499.4       | 0.06    | 51.00      | 7.67  | 42.75 | 45.08 | 4.67 | 49.83                             | -                 |
| CTRL.Delta | -        | -         | -          | 1.0         | -0.04   | -1.92      | 0.75  | -4.25 | 2.58  | 1.17 | 3.75                              | -                 |

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

### 3.5.2 Grade-Level Analysis of Changes in Math Meeting or Exceeding Expectations

Figure 8 shows the difference in the growth of percentages of students at math Meeting or Exceeding Expectations, for the TRT and CTRL datasets, disaggregated by grade:

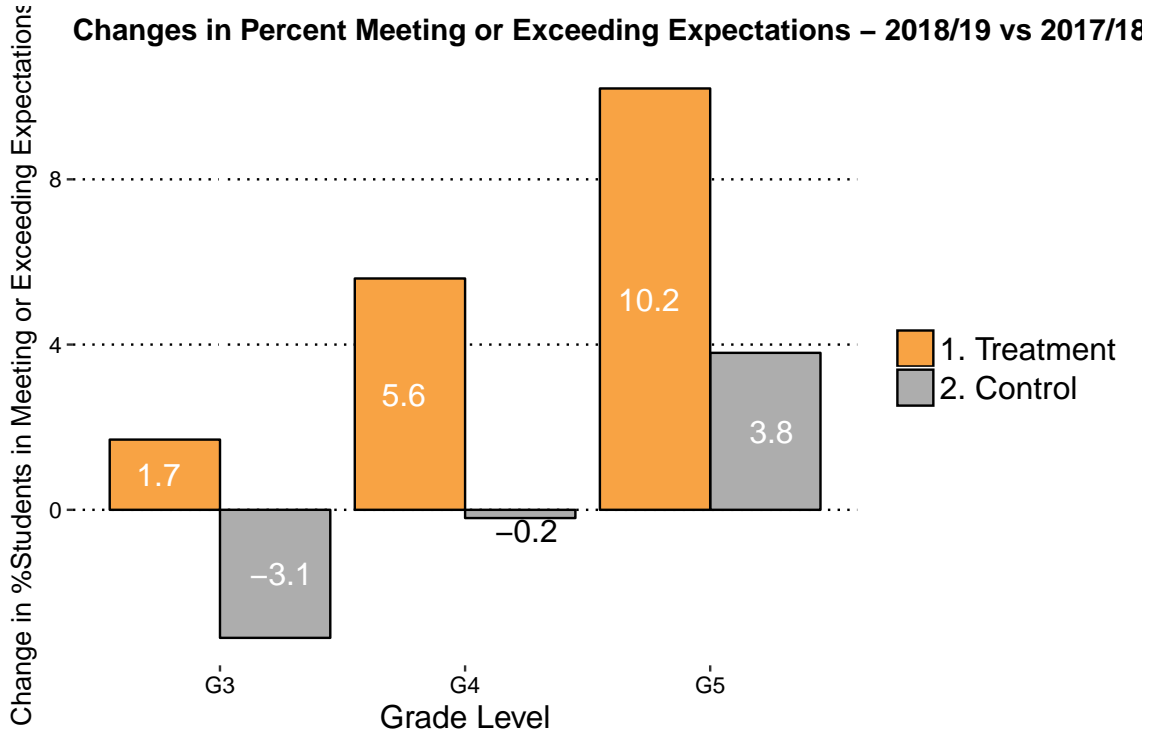


Figure 8: Changes in Percent of Students at Meeting or Exceeding Expectations for TRT and CTRL Datasets between 2017/18 and 2018/19

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

|         | Estimate | P-Value | Int.Low | Int.High |
|---------|----------|---------|---------|----------|
| Grade 3 | 4.76     | 0.13    | -1.57   | 11.10    |
| Grade 4 | 5.85     | 0.15    | -2.19   | 13.89    |
| Grade 5 | 6.50     | 0.26    | -5.25   | 18.25    |

Table 11: Statistics for the Differential Changes in Meeting or Exceeding Expectations, (TRT - CTRL)

### 3.5.3 Grade-Level Analysis of Changes in MCAS 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:

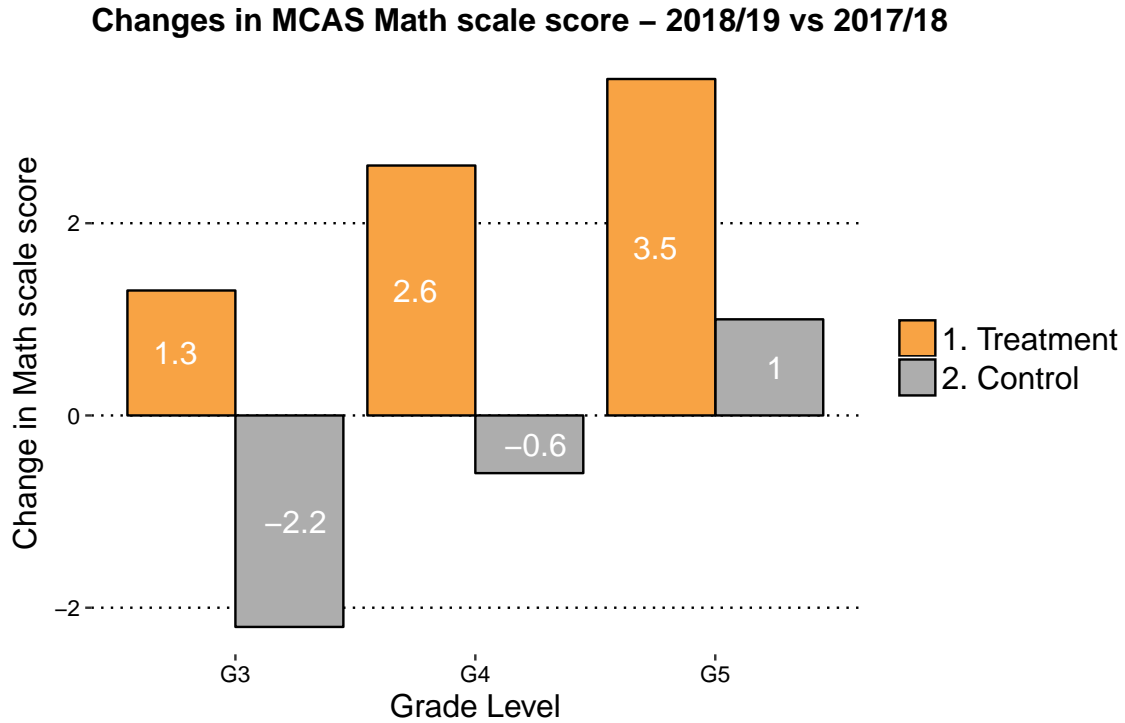


Figure 9: Changes in Grade-Mean MCAS Math scale score for TRT and CTRL Datasets between 2017/18 and 2018/19

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

|         | Estimate | P-Value | Int.Low | Int.High |
|---------|----------|---------|---------|----------|
| Grade 3 | 3.48     | 0.03*   | 0.35    | 6.62     |
| Grade 4 | 3.16     | 0.17    | -1.44   | 7.75     |
| Grade 5 | 2.58     | 0.22    | -1.65   | 6.80     |

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

### 3.5.4 Grade-Level Analysis of Changes in MCAS 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:

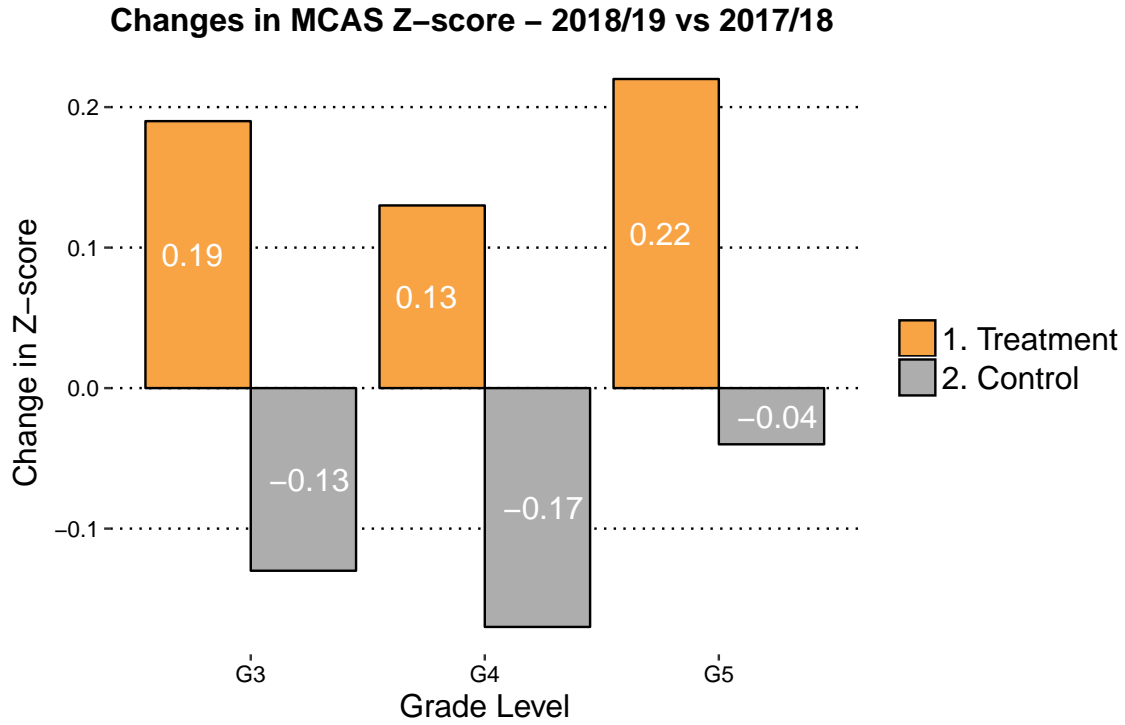


Figure 10: Changes in Grade-Mean MCAS Z-score (See Section 3.1) for TRT and CTRL Datasets between 2017/18 and 2018/19

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

|         | Estimate | P-Value | Int.Low | Int.High |
|---------|----------|---------|---------|----------|
| Grade 3 | 0.32     | 0.03*   | 0.03    | 0.61     |
| Grade 4 | 0.30     | 0.18    | -0.14   | 0.75     |
| Grade 5 | 0.25     | 0.24    | -0.18   | 0.69     |

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

## 4 Effect Size

The following table shows the effect sizes for Meeting or Exceeding Expectations, MCAS scale score, and accompanying Z-score.

|            | Scale score Effect Size | Z-score Effect Size | Meeting or Exceeding Expectations Effect Size |
|------------|-------------------------|---------------------|---|
| Grade 3    | 0.47                    | 0.45                | 0.37  |
| Grade 4    | 0.38                    | 0.37                | 0.38  |
| Grade 5    | 0.52                    | 0.48                | 0.51  |
| All Grades | 0.43                    | 0.42                | 0.41  |

Table 14: Cohen’s d Effect Size

## 5 Findings Summary

Massachusetts grades 3, 4, and 5 using ST Math for the year 2018/19 averaged 42.3% ST Math Progress. 52/78 grades (67%) averaged covering more than 40% of ST Math content. Statistically significant differences were found in this analysis for both grade-aggregated and individual grade levels. Looking at Table 7, statistically significant differences were found for grade-aggregated z-score, with an estimate of 0.3 points favorable for the ST Math treatment set, as well as for grade-aggregated Meeting or Exceeding Expectations proficiency levels, with a 5.63 point favorable differential for the ST Math treatment set. Looking at Table 13, grade 3 ST math treatment set outperformed their matched controls for MCAS z-scores with a statistically significant difference of 0.32.

## 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 40% 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   | Meeting or Exceeding Expectations | ST Math Per Comp. |
|--------------------|----------|-----------|------------|-------------|---------|------------|-------|-------|-------|------|-----------------------------------|-------------------|
| Grade 3 (17.18)    | 17       | 17        | 1388       | 500.5       | 0.02    | 49.76      | 10.88 | 39.41 | 41.71 | 8.06 | 49.76                             | -                 |
| Grade 4 (17.18)    | 20       | 20        | 1670       | 499.7       | 0.15    | 54.75      | 11.7  | 38.55 | 40.80 | 8.85 | 49.70                             | -                 |
| Grade 5 (17.18)    | 12       | 12        | 1144       | 498.9       | 0.14    | 53.92      | 6.17  | 47.50 | 42.25 | 3.92 | 46.42                             | -                 |
| All Grades (17.18) | 49       | 26        | 4202       | 499.8       | 0.10    | 52.82      | 10.06 | 41.04 | 41.47 | 7.37 | 48.92                             | -                 |
| Grade 3 (18.19)    | 17       | 17        | 1411       | 501.8       | 0.21    | 55.41      | 10.24 | 38.29 | 41.82 | 9.71 | 51.47                             | 49.16             |
| Grade 4 (18.19)    | 20       | 20        | 1599       | 502.3       | 0.28    | 56.65      | 10.0  | 34.90 | 45.65 | 9.60 | 55.30                             | 53.52             |
| Grade 5 (18.19)    | 12       | 12        | 1130       | 502.4       | 0.35    | 63.00      | 5.58  | 37.67 | 51.50 | 5.33 | 56.67                             | 51.85             |
| All Grades (18.19) | 49       | 26        | 4140       | 502.1       | 0.27    | 57.78      | 9.00  | 36.76 | 45.76 | 8.59 | 54.31                             | 51.6              |

Table 15: TRT Grades Detail Sorted by Year

|                    | # Grades | # Schools | # Students | Scale score | Z-Score | Percentile | L1    | L2    | L3    | L4   | Meeting or Exceeding Expectations | ST Math Per Comp. |
|--------------------|----------|-----------|------------|-------------|---------|------------|-------|-------|-------|------|-----------------------------------|-------------------|
| Grade 3 (17.18)    | 17       | 17        | 1162       | 499.9       | -0.04   | 48.35      | 11.41 | 39.53 | 40.76 | 8.12 | 49.00                             | -                 |
| Grade 4 (17.18)    | 20       | 20        | 1462       | 500.2       | 0.19    | 56.35      | 10.4  | 40.05 | 40.85 | 8.70 | 49.50                             | -                 |
| Grade 5 (17.18)    | 12       | 12        | 965        | 498.5       | 0.09    | 52.92      | 6.92  | 47.00 | 42.50 | 3.50 | 46.08                             | -                 |
| All Grades (17.18) | 49       | 48        | 3589       | 499.7       | 0.09    | 52.73      | 9.90  | 41.57 | 41.22 | 7.22 | 48.49                             | -                 |
| Grade 3 (18.19)    | 17       | 17        | 1113       | 497.7       | -0.17   | 44.71      | 13.18 | 41.12 | 39.12 | 6.71 | 45.94                             | -                 |
| Grade 4 (18.19)    | 20       | 20        | 1478       | 499.6       | 0.02    | 50.85      | 9.9   | 40.85 | 42.85 | 6.35 | 49.25                             | -                 |
| Grade 5 (18.19)    | 12       | 12        | 979        | 499.4       | 0.06    | 51.00      | 7.67  | 42.75 | 45.08 | 4.67 | 49.83                             | -                 |
| All Grades (18.19) | 49       | 48        | 3570       | 498.9       | -0.04   | 48.76      | 10.49 | 41.41 | 42.10 | 6.06 | 48.24                             | -                 |

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 40% progress filtering) used in the analysis.

| PID     | IID    | District                             | School Name                                   | GRADE   |
|---------|--------|--------------------------------------|---|---------|
| 1415472 | HIL07I | Attleboro                            | Hill-Roberts Elementary School                | 3, 4    |
| 418031  | THA0RT | Attleboro                            | Peter Thacher Elementary School               | 3, 4    |
| 440216  | JOHORX | Boston                               | Eliot Elementary                              | 3, 4    |
| 441167  | PHI0RT | Boston                               | Phineas Bates                                 | 3       |
| 4814566 | BOS0RS | Boston Collegiate Charter (District) | Boston Collegiate Charter School              | 5       |
| 425151  | MOU0RU | East Longmeadow                      | Mountain View                                 | 3       |
| 4284193 | GOL053 | Haverhill                            | Golden Hill                                   | 3, 4    |
| 430613  | BAR0RV | Lowell                               | Bartlett Community Partnership                | 4       |
| 430730  | JGP055 | Lowell                               | Pyne Arts                                     | 3, 4, 5 |
| 438940  | HEN0RS | Middleborough                        | Henry B. Burkland Elementary School           | 3, 4, 5 |
| 438952  | MAR0RX | Middleborough                        | Mary K. Goode Elementary School               | 3, 4, 5 |
| 1398537 | RES0RS | Millbury                             | Raymond E. Shaw Elementary                    | 4, 5    |
| 4362484 | ANN0RS | North Andover                        | Annie L Sargent School                        | 3, 4, 5 |
| 423191  | KIT0RS | North Andover                        | Kittredge                                     | 4, 5    |
| 423220  | THO0RT | North Andover                        | Thomson                                       | 4       |
| 441818  | WHE0RS | Revere                               | A. C. Whelan Elementary School                | 3       |
| 441820  | LIN0RV | Revere                               | Abraham Lincoln                               | 3, 4    |
| 2044624 | BEA0RU | Revere                               | Beachmont Veterans Memorial School            | 4, 5    |
| 441870  | JAM0RS | Revere                               | Garfield Elementary School                    | 3, 4    |
| 441935  | PAU0RS | Revere                               | Paul Revere                                   | 3, 4, 5 |
| 441923  | HIL0RX | Revere                               | Staff Sargent James J. Hill Elementary School | 5       |
| 420046  | HAR0RZ | Taunton                              | H H Galligan                                  | 3, 4    |
| 447173  | WHI0RS | Uxbridge                             | Whitin Intermediate                           | 4, 5    |
| 426478  | JOHORX | West Springfield                     | John R Fausey                                 | 3, 4    |
| 426313  | ABN0RS | Westfield                            | Abner Gibbs                                   | 4       |
| 448086  | UNI0RU | Worcester                            | Union Hill School                             | 3, 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 |
|----------|----------------------------------|--|-------|
| 428127   | Acton-Boxborough                 | McCarthy-Towne School                      | 5     |
| 2044741  | Ashburnham-Westminster           | Westminster Elementary                     | 4     |
| 416265   | Barnstable                       | Hyannis West Elementary                    | 3     |
| 417013   | Berkshire Hills                  | Muddy Brook Regional Elementary School     | 4     |
| 428634   | Billerica                        | Parker                                     | 5     |
| 441193   | Boston                           | Josiah Quincy                              | 5, 3  |
| 4320357  | Cambridge                        | King Open                                  | 4     |
| 416899   | Central Berkshire                | Craneville                                 | 5     |
| 5271814  | Dartmouth                        | James M. Quinn School                      | 4     |
| 418330   | Dartmouth                        | Joseph Demello                             | 4     |
| 429559   | Everett                          | George Keverian School                     | 4     |
| 5325025  | Everett                          | Lafayette School                           | 4     |
| 10905260 | Everett                          | Webster School                             | 5     |
| 418550   | Fall River                       | Letourneau Elementary School               | 3     |
| 3333258  | Franklin                         | Parmenter                                  | 3     |
| 424341   | Greenfield                       | Discovery School at Four Corners           | 3     |
| 430182   | Hudson                           | Mulready Elementary                        | 4     |
| 430792   | Lowell                           | Charles W Morey                            | 3     |
| 4015368  | Lowell                           | Dr Gertrude Bailey                         | 4     |
| 422331   | Lynn                             | Aborn                                      | 5     |
| 422599   | Lynn                             | Sewell-Anderson                            | 3     |
| 1846508  | Lynn                             | Washington Elementary School               | 3     |
| 422654   | Lynnfield                        | Huckleberry Hill                           | 4     |
| 431019   | Malden                           | Forestdale                                 | 3     |
| 431021   | Malden                           | Salemwood                                  | 4     |
| 419061   | Mansfield                        | Jordan/Jackson Elementary                  | 5     |
| 424171   | Mohawk Trail                     | Buckland-Shelburne Regional                | 3     |
| 416643   | Monomoy Regional School District | Harwich Elementary School                  | 4     |
| 419530   | North Attleborough               | Community                                  | 5     |
| 432233   | North Reading                    | J Turner Hood                              | 3     |
| 419554   | Norton                           | L G Nourse Elementary                      | 3     |
| 1168538  | Pittsfield                       | Williams                                   | 4     |
| 438691   | Plymouth                         | Nathaniel Morton Elementary                | 3     |
| 436459   | Quincy                           | Snug Harbor Community School               | 5     |
| 436461   | Quincy                           | Squantum                                   | 3     |
| 439229   | Rockland                         | Memorial Park                              | 3     |
| 432518   | Somerville                       | John F Kennedy                             | 4     |
| 425967   | Springfield                      | Alfred G. Zanetti Montessori Magnet School | 4     |
| 425931   | Springfield                      | Frank H Freedman                           | 5     |
| 426155   | Springfield                      | Samuel Bowles                              | 4     |
| 426260   | Springfield                      | Warner                                     | 4     |
| 426284   | Springfield                      | White Street                               | 3     |
| 419889   | Swansea                          | Joseph G Luther                            | 3     |
| 437037   | Wellesley                        | Katharine Lee Bates                        | 5     |
| 5342475  | Winthrop                         | Arthur T. Cummings Elementary School       | 4     |
| 447628   | Worcester                        | Canterbury                                 | 4     |
| 3333466  | Worcester                        | Jacob Hiatt Magnet                         | 5     |
| 447848   | Worcester                        | Lake View                                  | 4     |

Table 18: Matched Control Schools (CTRL Dataset)