

Massachusetts Math Outcomes Analysis 2021/22

Grade Levels: 3, 4, 5
ST Math Program: Gen-6
Analysis Type: Z-score of scale score
Treatment-Years: 2021/22
Baseline-Year: 2016/17, 2017/18, or 2018/19
Subgroup: All



Jessica Guise

© 2022-12-20

Abstract

This analysis evaluates grades using ST Math in Massachusetts in 2021/22. 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 64 grades, consisting of grades 3, 4, and 5 at 44 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 64 similar, randomly selected control grades at 59 schools that never used ST Math. Grade-wise growth in math proficiency was evaluated (i.e. growth in same grade, same school, from Baseline to 2021/22) 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 8.35 points at Meeting or Exceeding Expectations, 7.8 points at Meeting Expectations, and z-score of 0.35.

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	6
2.3	Control Grades Pool and Selection	7
3	Data Analysis	8
3.1	Z-scores of Scale Score	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
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 2021/22	9
2	Baseline Year Density Plots Showing Math Scores Match between TRT and CTRL - Baseline	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 Baseline and 2021/22	13
5	Changes in MCAS Math scale scores and Z-scores (See Section 3.1) for Grade- Aggregated TRT and CTRL datasets between Baseline and 2021/22	14
6	Changes in Meeting or Exceeding Expectations for Grade-Aggregated TRT and CTRL datasets between Baseline and 2021/22	15
7	Changes in Percentile Ranking for TRT and CTRL Datasets between Baseline and 2021/22	16
8	Changes in Percent of Students at Meeting or Exceeding Expectations for TRT and CTRL Datasets between Baseline and 2021/22	18
9	Changes in Grade-Mean MCAS Math scale score for TRT and CTRL Datasets be- tween Baseline and 2021/22	19
10	Changes in Grade-Mean MCAS Z-score of Scale Score (See Section 3.1) for TRT and CTRL Datasets between Baseline and 2021/22	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 ≥ 60 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 of Scale Score (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
19	Matched Control Schools (CTRL Dataset)	26

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 multi-year changes in grade-mean MCAS Meeting or Exceeding Expectations percentages. The treatment grades used the ST Math program for 2, 3, 4, or 5 years, beginning in the 2020/21, 2019/20, 2018/19, or 2017/18 school year, respectively. 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 (Baseline), and did not use ST Math in any subsequent year. 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 60% 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 in 2021/22 was identified. They comprise the Treatment grades pool for this evaluation of multi-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 60% 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 60% 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 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).

The method of matching used is propensity score matching, via the "matchit" program in R, with "mahalanobis" as the distance measure.

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 $\geq 60\%$ 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 of Scale Score

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 – 2021/22

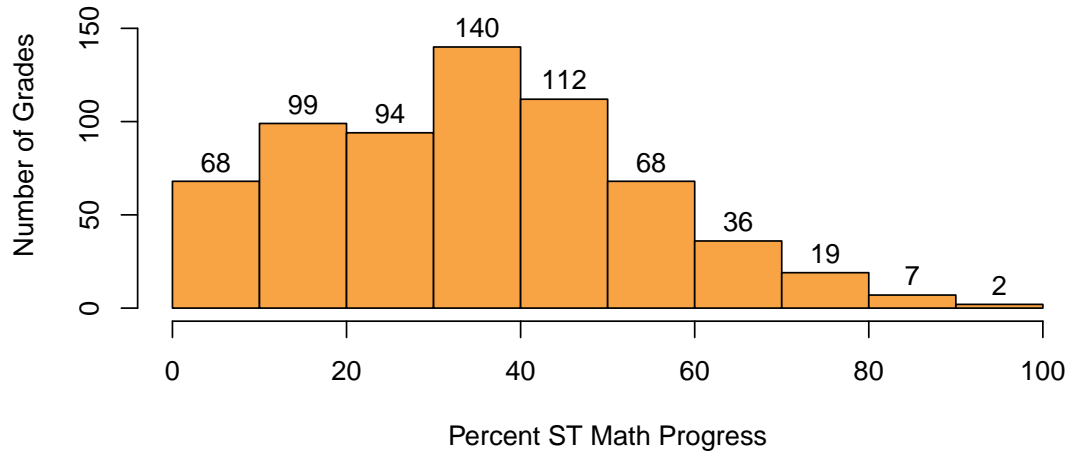


Figure 1: Histogram of ST Math Percent Progress for $\geq 85\%$ Enrollment Grades 2021/22

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 $\geq 60\%$ 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	0.0	94.3	34.6	19.2

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

Grades with $\geq 85\%$ Enrollment:	613
Grades with in addition $\geq 60\%$ Progress:	64

Table 3: Number of ST Math Grades with $\geq 85\%$ Enrollment and with $\geq 60\%$ 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 60% Progress filter. Other rows in the table indicate counts of numbers of students (2021/22 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	233	226	186	645
ST Math Using Schools	233	226	186	260
ST Math Students	14739	14692	13573	43004
ST Math Grades (Enroll \geq 85%)	221	220	172	613
TRT Grades (Enroll \geq 85% & Prog \geq 60%)	26	19	19	64
TRT Schools (Enroll \geq 85% & Prog \geq 60%)	26	19	19	44
TRT Students (Enroll \geq 85% & Prog \geq 60%)	1592	1296	980	3868
CTRL Grades	26	19	19	64
CTRL Schools	26	19	19	59
CTRL Students	1704	1107	1396	4207

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, Baseline.

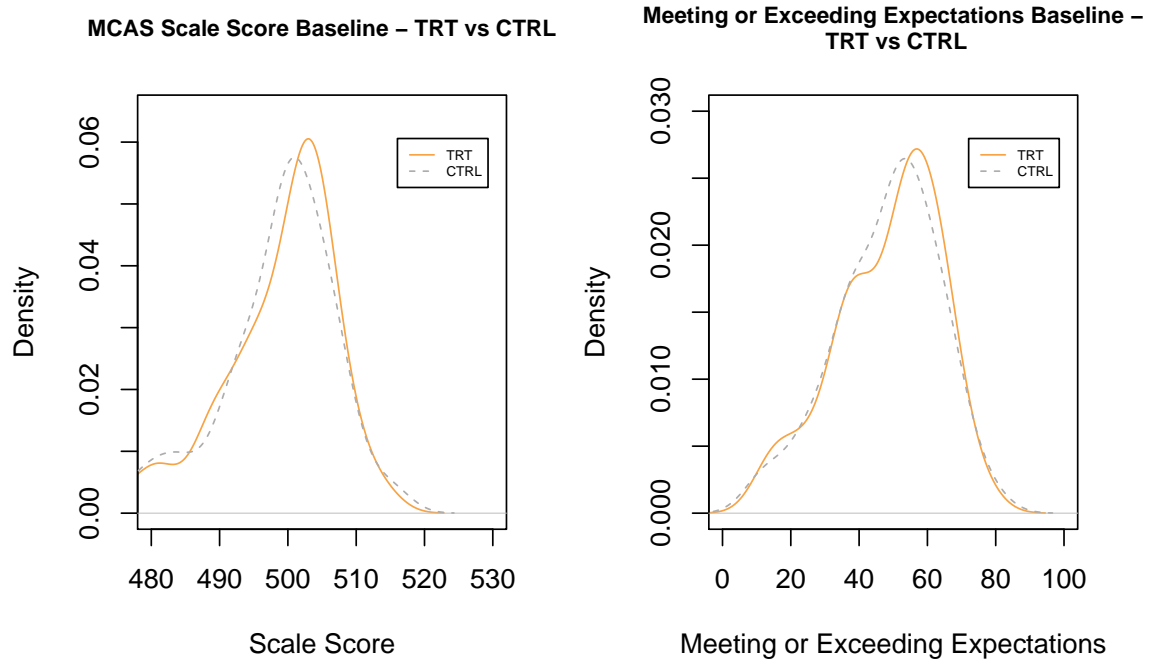


Figure 2: Baseline Year Density Plots Showing Math Scores Match between TRT and CTRL - Baseline

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

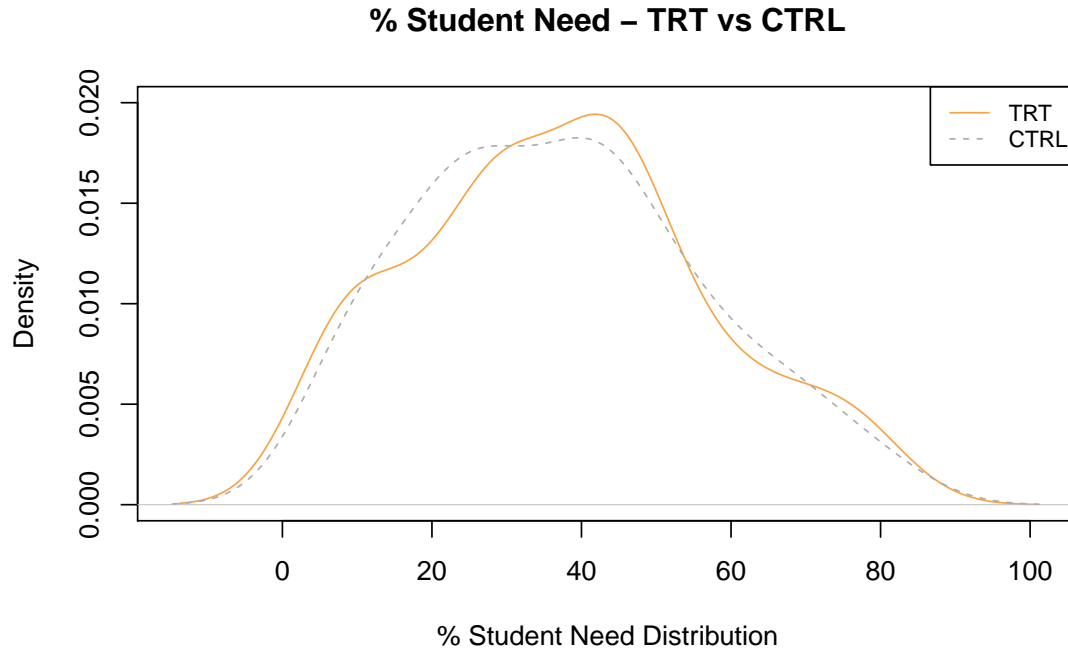


Figure 3: Baseline Year Density Plot Showing Student Need Match between TRT and CTRL

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 - Baseline	48.67	15.06	48.24	14.99	0.43	0.87	0.03
Scale Score - Baseline	498.83	8.34	498.65	8.36	0.18	0.90	0.02
Percent Free or Reduced Lunch	37.31	19.40	37.19	18.94	0.12	0.97	0.01

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 of SS	Percentile	L1	L2	L3	L4	Meeting or Exceeding Expectations	ST Math Per Con
TRT.Baseline	64	44	4001	498.8	-0.04	50.48	13.58	37.83	41.81	6.84	48.67	-
TRT.21.22	64	44	3597	496.9	0.28	59.55	12.33	41.59	41.00	5.08	46.05	70.75
TRT.Delta	-	-	-	-1.9	0.31	9.06	-1.25	3.77	-0.81	-1.77	-2.62	-
CTRL.Baseline	64	59	4452	498.6	-0.06	49.67	12.36	39.43	41.62	6.60	48.24	-
CTRL.21.22	64	59	4207	492.4	-0.09	48.06	16.25	46.55	33.02	4.17	37.27	-
CTRL.Delta	-	-	-	-6.3	-0.03	-1.61	3.89	7.11	-8.61	-2.43	-10.98	-

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).

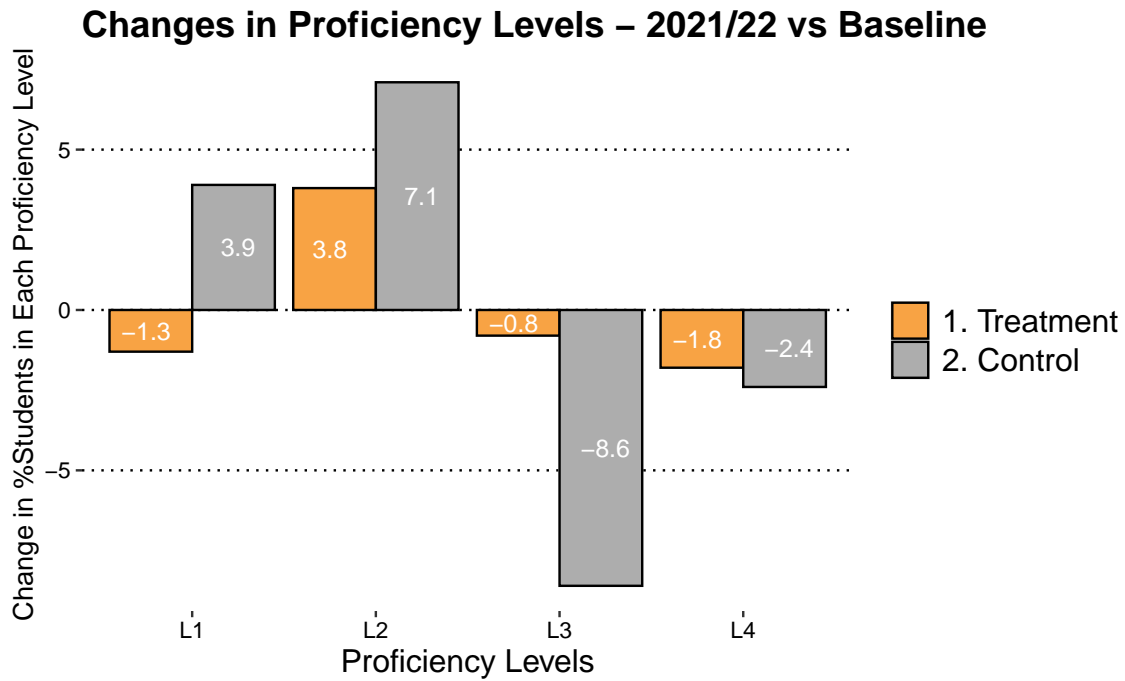


Figure 4: Change at each Proficiency Level for Grade-Aggregated TRT and CTRL Datasets between Baseline and 2021/22

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

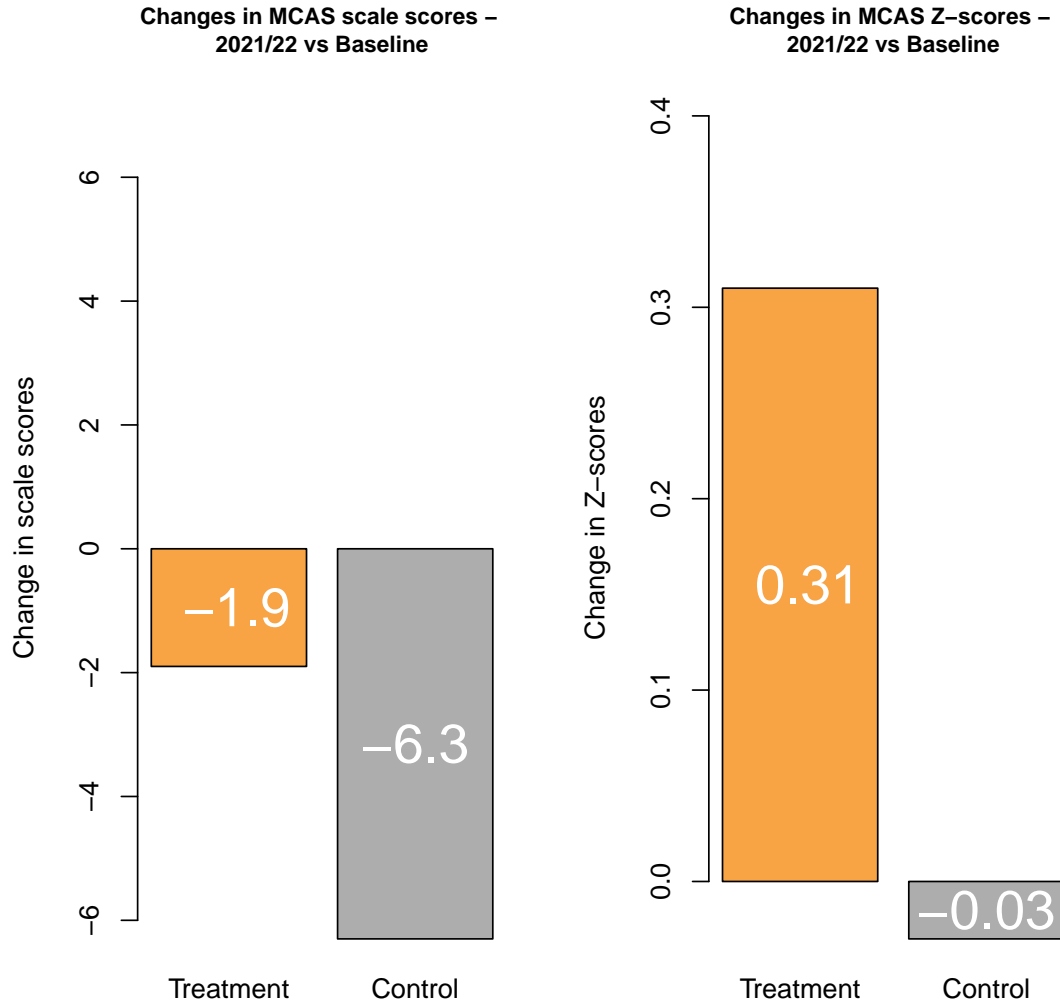


Figure 5: Changes in MCAS Math scale scores and Z-scores (See Section 3.1) for Grade-Aggregated TRT and CTRL datasets between Baseline and 2021/22

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

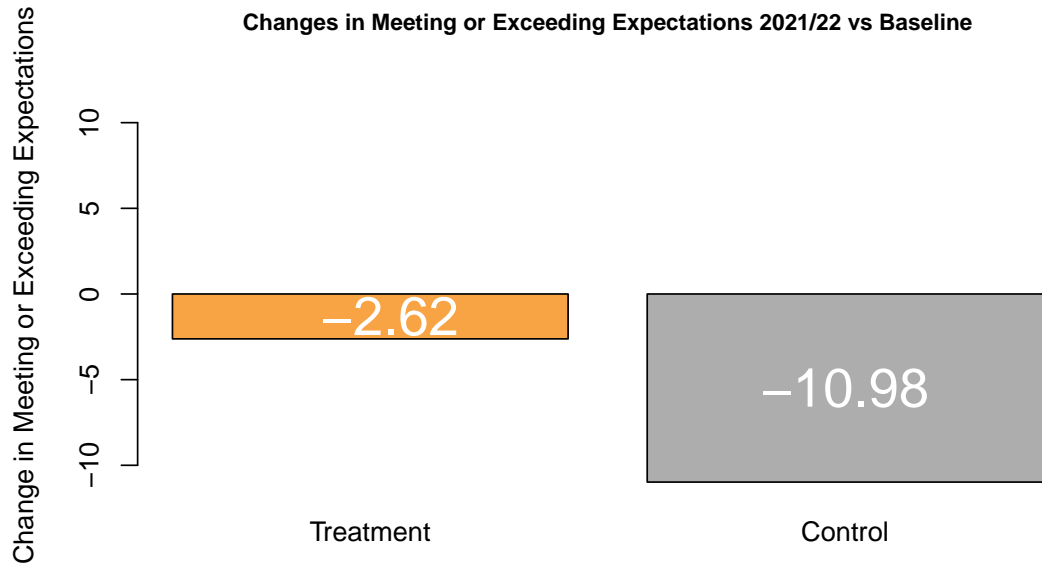


Figure 6: Changes in Meeting or Exceeding Expectations for Grade-Aggregated TRT and CTRL datasets between Baseline and 2021/22

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.¹

	Estimate	P-Value	Int.Low	Int.High
Meeting or Exceeding Expectations	8.35	0.00*	3.54	13.17
Scale Score	4.34	0.00*	1.76	6.93
Z-score of SS	0.35	0.00*	0.12	0.57
L1	-5.15	0.01*	-9.02	-1.27
L2	-3.35	0.19	-8.36	1.66
L3	7.80	0.00*	3.60	11.99
L4	0.67	0.49	-1.22	2.55

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.

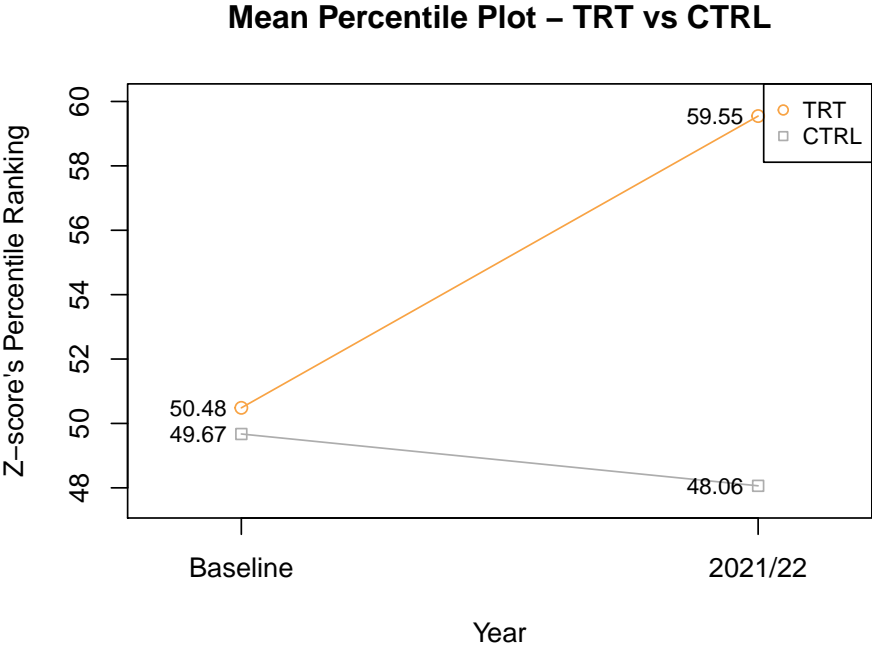


Figure 7: Changes in Percentile Ranking for TRT and CTRL Datasets between Baseline and 2021/22

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 of SS	Percentile	L1	L2	L3	L4	Meeting or Exceeding Expectations	ST Math Per Prog
TRT.Baseline	26	26	1623	500.2	0.07	54.50	14.12	32.96	44.06	9.05	53.07	-
TRT.21.22	26	26	1466	496.9	0.29	60.12	13.81	38.65	41.96	5.65	47.62	68.91
TRT.Delta	-	-	-	-3.3	0.22	5.62	-0.31	5.69	-2.10	-3.39	-5.45	-
CTRL.Baseline	26	26	1778	500.1	0.06	53.73	14.26	33.68	43.15	9.00	52.10	-
CTRL.21.22	26	26	1704	493.5	0.02	52.62	17.38	42.85	34.42	5.35	40.00	-
CTRL.Delta	-	-	-	-6.6	-0.03	-1.12	3.13	9.17	-8.72	-3.65	-12.10	-

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

	# Grades	# Schools	# Students	Scale Score	Z-score of SS	Percentile	L1	L2	L3	L4	Meeting or Exceeding Expectations	ST Math Per Prog
TRT.Baseline	19	19	1341	499.6	0.03	52.63	14.35	37.09	41.22	7.29	48.62	-
TRT.21.22	19	19	1218	498.8	0.35	61.74	10.47	38.68	45.16	5.79	50.79	71.14
TRT.Delta	-	-	-	-0.8	0.32	9.11	-3.88	1.60	3.93	-1.50	2.17	-
CTRL.Baseline	19	19	1098	499.3	0.01	51.53	11.17	40.23	42.00	6.60	48.71	-
CTRL.21.22	19	19	1107	493.2	-0.08	48.68	15.47	43.47	36.00	5.05	40.95	-
CTRL.Delta	-	-	-	-6.1	-0.10	-2.84	4.30	3.25	-6.00	-1.54	-7.76	-

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

	# Grades	# Schools	# Students	Scale Score	Z-score of SS	Percentile	L1	L2	L3	L4	Meeting or Exceeding Expectations	ST Math Per Prog
TRT.Baseline	19	19	1037	496.2	-0.25	42.84	12.07	45.22	39.32	3.39	42.71	-
TRT.21.22	19	19	913	495.0	0.18	56.58	12.16	48.53	35.53	3.58	39.16	72.88
TRT.Delta	-	-	-	-1.2	0.44	13.74	0.09	3.31	-3.79	0.19	-3.55	-
CTRL.Baseline	19	19	1576	496.0	-0.28	42.26	10.94	46.52	39.16	3.34	42.50	-
CTRL.21.22	19	19	1396	490.0	-0.24	41.21	15.47	54.68	28.11	1.68	29.84	-
CTRL.Delta	-	-	-	-6.0	0.03	-1.05	4.54	8.17	-11.05	-1.66	-12.65	-

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 Meeting or Exceeding Expectations, for the TRT and CTRL datasets, disaggregated by grade:

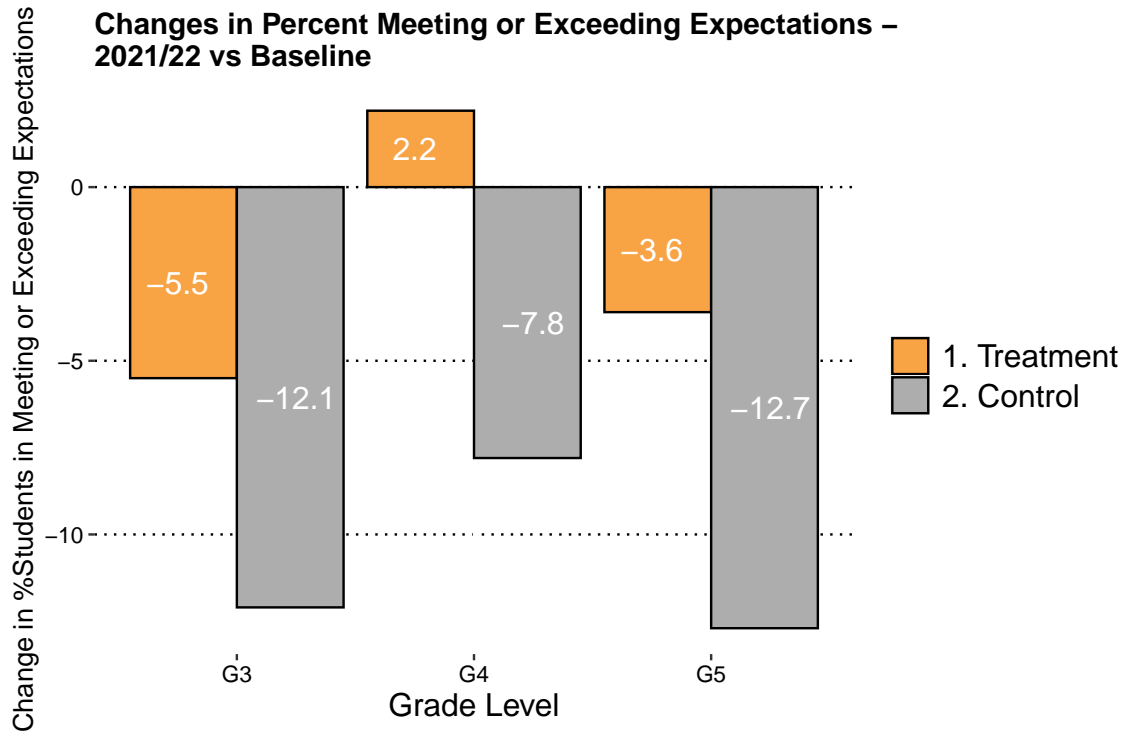


Figure 8: Changes in Percent of Students at Meeting or Exceeding Expectations for TRT and CTRL Datasets between Baseline and 2021/22

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	6.65	0.12	-1.76	15.07
Grade 4	9.93	0.02*	1.60	18.26
Grade 5	9.10	0.03*	0.73	17.48

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:

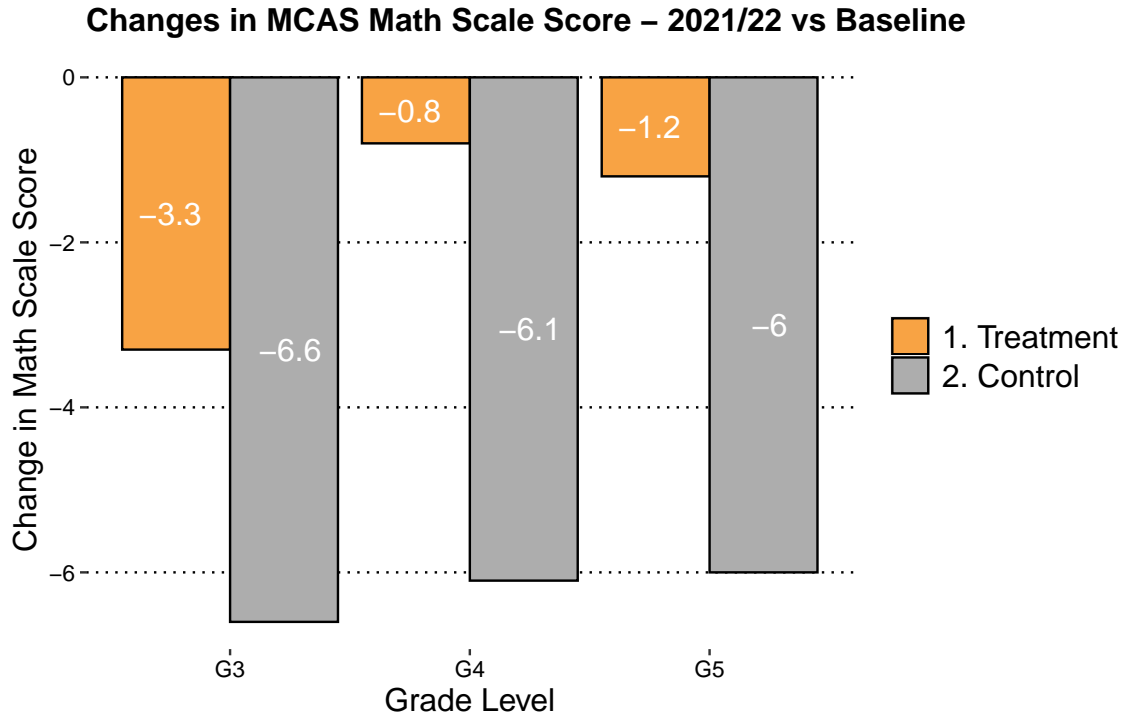


Figure 9: Changes in Grade-Mean MCAS Math scale score for TRT and CTRL Datasets between Baseline and 2021/22

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	3.27	0.2	-1.74	8.28
Grade 4	5.33	0.01*	1.24	9.42
Grade 5	4.82	0.02*	0.71	8.93

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 scale score of students for the TRT and CTRL datasets, disaggregated by grade:

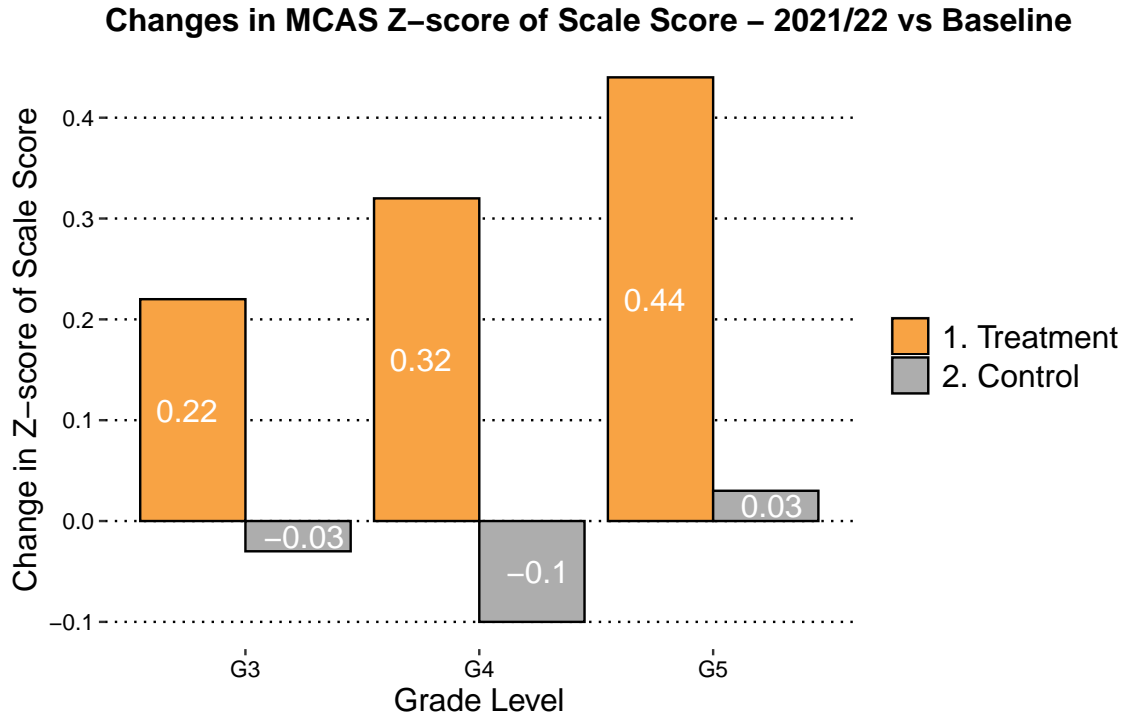


Figure 10: Changes in Grade-Mean MCAS Z-score of Scale Score (See Section 3.1) for TRT and CTRL Datasets between Baseline and 2021/22

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	0.25	0.24	-0.18	0.69
Grade 4	0.41	0.02*	0.06	0.77
Grade 5	0.41	0.04*	0.02	0.79

Table 13: Statistics for the Differential Changes in MCAS Z-scores of Scale Score (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 of Scale Score Effect Size	Meeting or Exceeding Expectations Effect Size
Grade 3	0.34	0.28	0.41
Grade 4	0.77	0.61	0.76
Grade 5	0.64	0.53	0.64
All Grades	0.52	0.43	0.56

Table 14: Cohen's d Effect Size

5 Findings Summary

Massachusetts grades 3, 4, and 5 using ST Math for the year 2021/22 averaged 34.6% ST Math Progress. 64/645 grades (10%) averaged covering more than 60% 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 of scale score, with an estimate of 0.35 points favorable for the ST Math treatment set, as well as for grade-aggregated Meeting or Exceeding Expectations proficiency levels, with a 8.35 point favorable differential for the ST Math treatment set. Further, in Table 7, grade-aggregated ST Math treatment set outperformed their matched controls at the Meeting Expectations level, with a statistically significant difference of 7.8. Referring to Table 11, statistically significant differences were found for grades 4 and 5 Meeting or Exceeding Expectations proficiency levels, with estimates of 9.93 and 9.1 respectively, in favor of the ST Math treatment set. Looking at Table 13, grades 4 and 5 ST math treatment sets outperformed their matched controls for MCAS z-score of scale score with statistically significant differences of 0.41 and 0.41, respectively.

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 60% 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 of SS	Percentile	L1	L2	L3	L4	Meeting or Exceeding Expectations	ST Math Per Comp.
Grade 3 (Baseline)	26	26	1623	500.2	0.07	54.50	14.12	32.96	44.06	9.05	53.07	-
Grade 4 (Baseline)	19	19	1341	499.6	0.03	52.63	14.35	37.09	41.22	7.29	48.62	-
Grade 5 (Baseline)	19	19	1037	496.2	-0.25	42.84	12.07	45.22	39.32	3.39	42.71	-
All Grades (Baseline)	64	44	4001	498.8	-0.04	50.48	13.58	37.83	41.81	6.84	48.67	-
Grade 3 (21.22)	26	26	1466	496.9	0.29	60.12	13.81	38.65	41.96	5.65	47.62	68.91
Grade 4 (21.22)	19	19	1218	498.8	0.35	61.74	10.47	38.68	45.16	5.79	50.79	71.14
Grade 5 (21.22)	19	19	913	495.0	0.18	56.58	12.16	48.53	35.53	3.58	39.16	72.88
All Grades (21.22)	64	44	3597	496.9	0.28	59.55	12.33	41.59	41.00	5.08	46.05	70.75

Table 15: TRT Grades Detail Sorted by Year

	# Grades	# Schools	# Students	Scale Score	Z-Score of SS	Percentile	L1	L2	L3	L4	Meeting or Exceeding Expectations	ST Math Per Comp.
Grade 3 (Baseline)	26	26	1778	500.1	0.06	53.73	14.26	33.68	43.15	9.00	52.10	-
Grade 4 (Baseline)	19	19	1098	499.3	0.01	51.53	11.17	40.23	42.00	6.60	48.71	-
Grade 5 (Baseline)	19	19	1576	496.0	-0.28	42.26	10.94	46.52	39.16	3.34	42.50	-
All Grades (Baseline)	64	59	4452	498.6	-0.06	49.67	12.36	39.43	41.62	6.60	48.24	-
Grade 3 (21.22)	26	26	1704	493.5	0.02	52.62	17.38	42.85	34.42	5.35	40.00	-
Grade 4 (21.22)	19	19	1107	493.2	-0.08	48.68	15.47	43.47	36.00	5.05	40.95	-
Grade 5 (21.22)	19	19	1396	490.0	-0.24	41.21	15.47	54.68	28.11	1.68	29.84	-
All Grades (21.22)	64	59	4207	492.4	-0.09	48.06	16.25	46.55	33.02	4.17	37.27	-

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 60% progress filtering) used in the analysis.

PID	District	School Name	GRADE
420838	Amesbury	Amesbury Elementary	3, 4
1398496	Amesbury	Charles C Cashman Elementary	3
418093	Attleboro	A. Irvin Studley Elementary School	3, 4
1415472	Attleboro	Hill-Roberts Elementary School	3
1415484	Attleboro	Hyman Fine Elementary School	4
441167	Boston	Bates Elementary School	3
440979	Boston	Mozart Elementary School	5
441466	Boston	Warren-Prescott K-8 School	3
428775	Burlington	Francis Wyman Elementary	5
428751	Burlington	Memorial	3
418316	Dartmouth	George H Potter	5
418897	Fall River	Spencer Borden	3
416564	Falmouth	Teaticket	3
11848598	Lowell	Rogers STEM Academy	3
4868701	Mendon-Upton	Henry P Clough	3, 4
446325	Mendon-Upton	Memorial School	3, 4
438952	Middleborough	Mary K. Goode Elementary School	4
417271	North Adams	Colegrove Park Elementary	3, 5
4362484	North Andover	Annie L Sargent School	5
419669	Seekonk	George R Martin	4
4755881	Taunton	East Taunton Elementary	3
446284	Wachusett	Thomas Prince	3, 5
4841624	Walpole	Elm Street School	4, 3
439499	West Bridgewater	Rose L Macdonald	3
437192	Weymouth	Academy Avenue	4
437348	Weymouth	Lawrence W Pingree	3
437398	Weymouth	Thomas W. Hamilton Primary School	4
447551	Worcester	Belmont Street Community	5
447599	Worcester	Burncoat Street	3, 5
447721	Worcester	Flagg Street	3
1540766	Worcester	Francis J McGrath Elementary	3
447824	Worcester	Heard Street	5
3333466	Worcester	Jacob Hiatt Magnet	4, 5
447848	Worcester	Lake View	3, 4, 5
447898	Worcester	May Street	4, 5
447915	Worcester	Midland Street	4, 5
447941	Worcester	Nelson Place	4, 5
447965	Worcester	Norrback Avenue	5
447939	Worcester	Quinsigamond	3, 4
448036	Worcester	Roosevelt	4, 5
448074	Worcester	Thorndyke Road	3, 4, 5
448086	Worcester	Union Hill School	5
448115	Worcester	West Tatnuck	3
448050	Worcester	Worcester Arts Magnet School	3, 4, 5

Table 17: Treatment Schools (TRT Dataset)

8.2 Control Schools

The following tables list 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
424779	Agawam	Clifford M Granger	3
424808	Agawam	Robinson Park	3, 4
434669	Avon	Ralph D Butler	3
416241	Barnstable	West Villages Elementary School	3
416851	Central Berkshire	Becket Washington School	3
4282793	Chelsea	George F. Kelly Elementary	3
424937	Chicopee	Bowe	5
3251820	Chicopee	Fairview Elementary	5
425058	Chicopee	Streiber Memorial School	4
1523108	Clinton	Clinton Middle School	5
429365	Dracut	George H. Englesby Elementary School	3
429418	Dracut	Joseph A Campbell Elementary	5
445084	Dudley-Charlton Reg	Heritage School	4
429559	Everett	George Keverian School	3
418603	Fall River	Mary Fonseca Elementary School	3, 5
11435517	Fitchburg	McKay Elementary School	4
429652	Framingham	Charlotte A Dunning	3
5092979	Franklin	Helen Keller Elementary	4
418146	Freetown-Lakeville	Freetown Elementary School	3
421624	Gloucester	Veterans Memorial	3
424640	Hawlemont	Hawlemont Regional	4
438768	Hull	Lillian M Jacobs	3
10911702	Leominster	Frances Drake School	3
431069	Malden	Linden	5
431459	Medford	John J McGlynn Elementary School	5
431320	Medford	Missituk Elementary School	4, 5
3049392	Milton	Cunningham School	3
424171	Mohawk Trail	Buckland-Shelburne Regional	5
416643	Monomoy Regional School District	Harwich Elementary School	4
2044090	Mount Greylock	Lanesborough Elementary	4
1413515	New Bedford	Casimir Pulaski	5
419164	New Bedford	Elizabeth Carter Brooks	3
419530	North Attleborough	Community	5
419475	North Attleborough	Joseph W Martin Jr Elementary	3
1171195	Northampton	Jackson Street	4
427355	Pelham	Pelham Elementary	5
2044595	Pembroke	Hobomock Elementary	5
1822277	Plymouth	West Elementary	4
3266851	Quabbin	Hardwick Elementary	3
445864	Quabbin	Hubbardston Center	4
4369690	Quincy	Beechwood Knoll Elementary	4
436368	Quincy	Montclair	3
436538	Randolph	Elizabeth G Lyons Elementary	4, 3
432295	Reading	Alice M Barrows	3
10001773	Reading	Wood End Elementary School	3
438380	Rochester	Rochester Memorial	4
4887783	Roxbury Preparatory Charter (District)	Roxbury Preparatory Charter School	5
419671	Seekonk	Mildred Aitken School	3
419774	Somerset	South	5
446959	Southbridge	West Street	3
436801	Stoughton	Joseph R Dawe Jr Elementary	5
436849	Stoughton	Richard L. Wilkins Elementary School	3, 4
12101903	TEC Connections Academy Commonwealth Virtual School District	TEC Connections Academy Commonwealth Virtual School	5
420785	Tisbury	Tisbury Elementary	4

Table 18: Matched Control Schools (CTRL Dataset)

PID	District	School Name	GRADE
433196	Waltham	Henry Whittemore Elementary School	5
433421	Watertown	Hosmer	4
426519	West Springfield	Mittineague	4
447343	Westborough	Elsie A Hastings Elementary	3

Table 19: Matched Control Schools (CTRL Dataset)