

2017-18 Evaluation of Reasoning Mind

Department of Evaluation and Assessment

*Michael Hinojosa, Ed.D.
Superintendent of Schools*



© 2018 Dallas Independent School District (Dallas ISD)

Mail a written request for permission to use or reproduce any part of this document to:

Dallas ISD

Department of Evaluation and Assessment

H. B. Bell School Support Service Center

2909 N. Buckner Blvd.

Box 10

Dallas, TX 75228

2017-18 Evaluation of Reasoning Mind

Approved Report of the Department of Evaluation and Assessment

EA18-512-2

August 2018

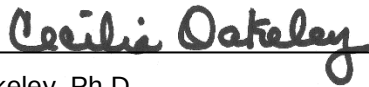
Myoungsook Kim, Ph.D.

Joan Bush, Ph.D.



Joan Bush, Ph.D.

Executive Director – Program Evaluation



Cecilia Oakeley, Ph.D.

Assistant Superintendent – Evaluation and Assessment

*Michael Hinojosa, Ed.D.
Superintendent of Schools*



This page is intentionally left blank.

Table of Contents

ABSTRACT	1
PROGRAM DESCRIPTION	2
Purpose and Background	2
RM Curriculum, Training, and Support	2
PURPOSE AND SCOPE OF THE EVALUATION	3
Limitations of the Evaluation	4
MAJOR EVALUATION QUESTIONS AND RESULTS	4
What were the sources and amounts of funding for Reasoning Mind?	4
Methodology	4
Results	4
What was the context of Reasoning Mind?	5
Methodology	5
Results	5
What were the demographic characteristics of students involved in Reasoning Mind?	7
Methodology	7
Results	7
Was the Reasoning Mind program used by students as planned?	11
Methodology	11
Results	11
Did staff members participate in Reasoning Mind training as planned?	15
Methodology	15
Results	15
What were teacher and administrator perceptions of Reasoning Mind?	16
Methodology	16
Results	16
What were student mathematics achievement outcomes?	24
Methodology	24
Results	26

SUMMARY AND RECOMMENDATIONS.....	36
Summary.....	36
Recommendations	39
REFERENCES.....	41
APPENDICES	41

List of Tables

Table		Page
1	Spring 2018 Campus Staff Survey Respondent Characteristics	17
2	Percentage of Teachers that Used Various Mathematics Intervention Programs by Grade	18
3	Spring 2018 <i>TerraNova</i> Mathematics Results for Students in Grade 2.....	27
4	Spring 2018 <i>STAAR</i> Mathematics Results for Students in Grades 3 to 5.....	27
5	Multiple Regression Major RM Predictors of Spring 2018 Mathematics Scores	33
6	Multiple Regression Results for RM and Non-RM Student Comparisons	34

List of Figures

Figure		Page
1	Framework for the 2017-18 Evaluation of Reasoning Mind.....	3
2	RM Goal, Action Steps, and Implementation Targets.....	5
3	Timeline for Expansion of RM in Dallas ISD	6
4	Student and School Participation in RM (2015-16 to 2017-18)	8
5	2017-18 Number of RM Students by Grade and Feeder Pattern	9
6	2017-18 Demographic Characteristics of RM Students.....	10
7	2017-18 Campus-Level RM Implementation	12
8	2017-18 RM Student Hour and Accuracy Results by Grade Level	13
9	RM Student Mean Hours and Mean Accuracy by Grade Level from 2015-16 to 2017-18	14
10	Percentage of First-Year RM Teachers that Completed the QC by the End of the First Six Weeks	15
11	Percentage of Teachers that Used Other Mathematics Intervention Programs by Grade	18
12	Percentage of Campus Staff Members that “Agreed” or “Strongly Agreed” with Overall Items ...	19
13	Percentage of Campus Staff Members that Were “Satisfied” or “Extremely Satisfied” with RM and District Staff Support	20
14	Percentage of Teachers that Marked “Improved” or “Significantly Improved” for Student Change Items	20
15	Teachers’ Familiarity with RM’s SOC by Training Experience	21
16	Percentage of Campus Staff Members that Chose “Frequently” or “Very Frequently” for General Technology and Computer Scheduling/Access Issues.....	22
17	Campus Staff Survey Responses for RM Successes.....	23
18	Campus Staff Survey Responses for RM Challenges	24
19	Correlations between RM Indicators, Prior Achievement, and Spring 2018 Mathematics Assessment Scores by Grade	28
20	Percentage of Students Performing At or Above Grade Level on Spring 2018 <i>TerraNova</i> by RM Hour and Accuracy Categories.....	29
21	Percentage of Students in Grades 3 to 5 Who Achieved Approaches Grade Level or Above (Approaches +) on Spring 2018 STAAR Mathematics by RM Hour and Accuracy Categories	29

22	Percentage of Students in Grades 3 to 5 Who Achieved Meets Grade Level or Above (Meets +) on Spring 2018 <i>STAAR</i> Mathematics by RM Hour and Accuracy Categories	30
23	Spring 2018 Assessment Mean Scores by RM Hour and Accuracy Categories	31
24	RM and Non-RM Student Comparisons on <i>TerraNova</i> 2018 and <i>STAAR</i> 2018	35
25	RM and Non-RM Student Comparisons on Spring 2018 <i>STAAR</i> Progress Measure	36

List of Appendices

Appendix	Page
A Demographic Characteristics of 2017-18 RM Students and District Students	43
B 2017-18 RM Hour and Accuracy Statistics by Campus and Grade	44
C 2017-18 Campuses with a High Level of RM Implementation.....	47
D 2017-18 Campuses with a Medium Level of RM Implementation	48
E 2017-18 Campuses with a Low Level of RM Implementation.....	49
F RM Student Hour and Accuracy Results by Grade Level from 2015-16 to 2017-18	52
G Number and Percentage of Teachers in 2017-18 that Completed the RM Qualification Course by Feeder Pattern and Time Period	53
H Spring 2018 RM Campus Administrator Survey Results	54
I Spring 2018 RM Teacher Survey Results.....	55
J RM Campus Staff Survey Results for Successes and Challenges	56
K Propensity Score Matching Detailed Methods.....	57
L Spring 2018 <i>TerraNova/STAAR</i> Mathematics Results by RM Hour and Accuracy Categories	60
M Multiple Regression Results for Spring 2018 Mathematics Scores and Various Predictors (RM Students in Grades 2 to 5).....	61

2017-18 EVALUATION OF REASONING MIND

Project Evaluators: Myoungsook Kim, Ph.D. & Joan Bush, Ph.D.

ABSTRACT

2017-18 Reasoning Mind Implementation

In 2017-18, 11,311 students in grades two ($n=4,346$), three ($n=4,464$), four ($n=2,388$), and five ($n=113$) used the Reasoning Mind (RM) curriculum, which was a large decrease from 2016-17 ($N=29,714$). The 2017-18 RM budget was \$847,763.

Implementation fidelity continued to be a concern as the majority of campuses were classified as low implementing for second (87%), third (94%), fourth (88%), and fifth (100%) grades. The continued low level of implementation was surprising since most 2017-18 RM campuses had a medium or high level of implementation in 2016-17. Most students (91% to 100%) did *not* reach the hour target for 2017-18, and a sizeable proportion (79% to 100%) logged less than 80 percent of expected hours. Also, many second-grade students (60%) did *not* meet the accuracy target. Many first-year RM teachers (64%) did not complete the qualification course (QC) by the end of the first six weeks.

Campus Staff Perceptions of RM

Survey results revealed that teachers continued to be more positive toward RM than campus administrators. A noticeable proportion of campus staff members continued to report frequent problems with laptops/Computers on Wheels (COW) carts (34% to 38%), network (22%), and wireless connectivity (21% to 28%). The main successes were increased independent learning and increased mathematics and problem-solving skills. Major challenges were technology issues and meeting time requirements.

Mathematics Achievement of RM Students

With the exception of fifth grade, prior achievement, RM accuracy, and RM objectives completed were major predictors of Spring 2018 assessment scores. Like past years, RM students that met the accuracy target outperformed those who did not. Students that met the hour target showed similar patterns; however, the differences were not as pronounced as those for accuracy. When student performance was compared between RM and non-RM comparison students, results showed no clear pattern of differences between the two groups. For third and fourth grades, when prior achievement and economically disadvantaged status were accounted for, differences in *State of Texas Assessments of Academic Readiness (STAAR)* 2018 scale scores between RM and non-RM comparison students were statistically significant (favoring RM students) but showed no practical significance.

Recommendations

Main recommendations were to (a) require higher levels of implementation for 2018-19 RM campuses, (b) make sure teachers receive adequate guidance on how to meld two different curricula (district and RM) using RM's single objective curriculum, and (c) make certain all campuses have adequate working technology and are reporting issues to Information Technology as problems occur.

PROGRAM DESCRIPTION

Purpose and Background

The mission of Reasoning Mind (RM) is “to provide a first-rate math education for every child.”¹ The RM online mathematics curriculum was provided for district second-grade students in 2011-12, expanded to third grade in 2012-13, added to fourth grade in 2013-14, and expanded to fifth grade in 2014-15. During 2017-18, the RM supplemental mathematics curriculum was used for grades two, three, and four,² and the full (core) mathematics curriculum was used in grade five. However, unlike past years, the Science, Technology, Engineering, and Mathematics (STEM) Department limited the number of schools that could be funded centrally through Title I due to the budget cuts for 2017-18. Criteria to select schools funded centrally were based on 2016-17 campus-level RM implementation for each grade (schools that were categorized as having a high or medium level of RM implementation in 2016-17). For this reason, the number of RM schools funded centrally decreased from 147 in 2016-17 to 57 in 2017-18. There were five schools that purchased RM with their own budget;³ however, this evaluation report is limited to the 57 schools⁴ that were funded centrally because those schools were expected to follow the district’s implementation requirements such as maintaining time and accuracy targets.

RM Curriculum, Training, and Support

Through a blended learning model, RM students (a) used individual computers to work through the RM curriculum at their own pace, (b) took notes and wrote complete problem solutions as they developed mathematical writing skills, and (c) worked together to learn and solve mathematics problems. Using real-time RM data, RM teachers provided one-on-one or small group interventions. In addition, when a student struggled with a problem, a request for the “Genie Solution” could be made. The “Genie Solution” provided an explanation for solving the problem.

Until 2016-17, first-year RM teachers were required to take the RM Qualification Course (QC) before they used RM with students. However, the processing of the purchase orders for RM was delayed in Fall 2017. For this reason, first-year RM teachers in 2017-18 were not required, but were strongly encouraged, to take the QC before using RM with students. Additional training was available to all teachers along with campus-based support from a RM implementation coordinator (IC). In 2017-18, RM ICs held two group sessions at each RM campus based on the needs that were identified at the initial meeting with campus administrators. In addition, RM ICs held two individual meetings with each RM teacher to provide assistance. Since 2015-16, RM has used the Administrator Community Dashboard. The dashboard provided data on a variety of metrics including student online hours and Level A accuracy.⁵ All RM teachers, principals and school leadership executive directors had access to the dashboard.

¹ RM website, <https://www.reasoningmind.org/who-we-are/mission/>. Imagine Learning acquired RM on July 30, 2018.

² Because RM was implemented as a supplemental curriculum in grades two to four, teachers split instructional time between district and RM curricula.

³ Those schools included Carpenter, King, Peeler, Polk (grade five only), and Williams.

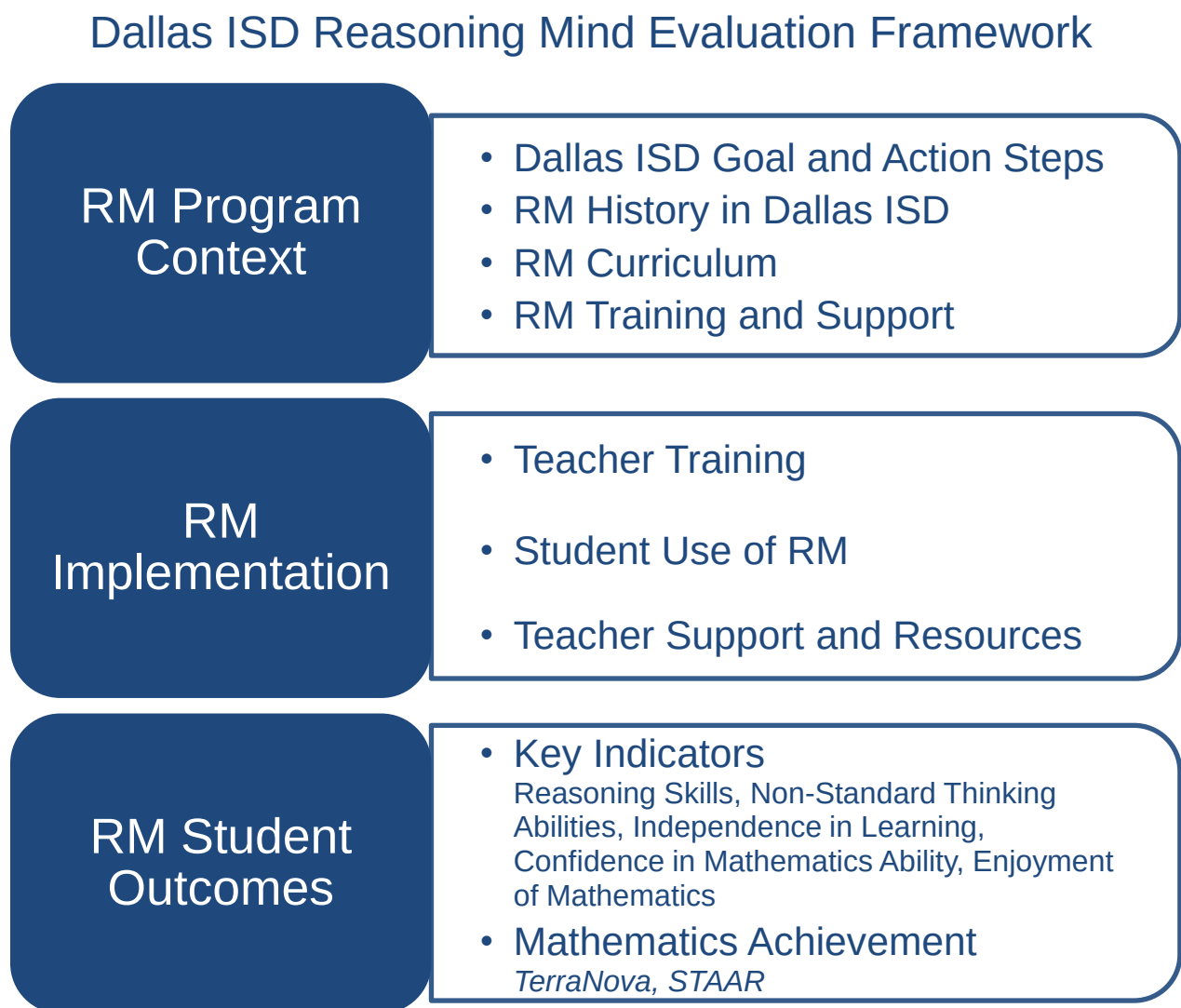
⁴ Source: RM program management as of January 2018.

⁵ In the Reasoning Mind learning and review modes, problems increase in difficulty from Level A (easiest) to Level C (hardest). The district’s accuracy target was based on Guided Study Level A items.

PURPOSE AND SCOPE OF THE EVALUATION

The purpose of this report was to summarize 2017-18 RM implementation findings, campus staff survey findings, and assessment outcomes. Figure 1 shows the Dallas ISD Reasoning Mind Evaluation Framework.⁶ Context information was gathered through review of internal documents, the RM website, and interviews with RM and Dallas ISD program staff. For the process information, the evaluators analyzed RM professional development (PD) data and student implementation data. Also, the evaluators administered and analyzed campus staff surveys. The outcome information consisted of student change data from teacher surveys and mathematics standardized assessment data.

Figure 1: Framework for the 2017-18 Evaluation of Reasoning Mind



Note. STAAR=State of Texas Assessments of Academic Readiness.

⁶ The Dallas ISD Reasoning Mind Evaluation Framework was developed by Bush and Kim (2015) to guide the evaluation.

Limitations of the Evaluation

Evaluation findings should be reviewed with caution for four main reasons. First, achievement results could not be attributed solely to RM due to the district's use of RM as supplemental to regular instruction in grades two to four (e.g., it is not possible to know how much achievement results should be attributed to RM and how much they should be credited to teacher instruction based on district curriculum). Second, district implementation fidelity continued to be a confounding factor as a sizeable proportion of students in all grades did not meet RM expectations for implementation. Third, there was possible selection bias between treatment and comparison groups for each grade. That is, the number of RM schools funded centrally decreased in 2017-18 because of budget cuts, and criteria to select schools funded centrally were based on 2016-17 campus level RM implementation. It should be noted also that only one school was selected for fifth grade which may limit the generalizability of the findings. Fourth, per the Spring 2018 RM teacher survey, between 10 percent and 29 percent of teacher respondents used additional mathematics programs/resources such as Think Through Math (TTM) and Houghton Mifflin Harcourt (HMH) Personal Math Trainer.

MAJOR EVALUATION QUESTIONS AND RESULTS

What were the sources and amounts of funding for Reasoning Mind?

Methodology

The evaluators reviewed the Title I workscope as well as information provided through the Science, Technology, Engineering, and Mathematics (STEM) Department. RM supplemental costs were funded through Title I funds, and RM fifth-grade costs were funded through general operating funds.

Results

The total budgeted amount for RM in 2017-18 was \$847,763, which included \$838,962 for the grades two to four supplemental RM program (Title I, Part A funds) and \$8,801 for the grade five core RM program (general operating funds). The total budget amount (\$847,763) was decreased by about half from 2016-17 (\$1,763,660).

The \$838,962 of Title I, Part A funds for supplemental RM included \$462,462 for student accounts and \$376,500 for PD. The \$8,801 of general operating funds for fifth-grade included \$4,181 for student accounts and \$4,620 for PD.⁷

⁷ In 2017-18, the cost per student account was \$37 for both grades two to four and grade five.

What was the context of Reasoning Mind?

Methodology

Following the Dallas ISD Reasoning Mind Evaluation Framework, information was reviewed to provide context for RM use in the district. The evaluators conducted informal interviews with RM and Dallas ISD program staff members. Also, internal documents were reviewed including the Title I grant workscope, district internal documents, and the RM dashboard.

Results

The RM goal and action steps to achieve it were summarized in a November 2012 memorandum from Superintendent Miles to the Board of Trustees and updated based on additional RM documentation. Teacher and student targets were set to ensure a high level of student implementation (see Figure 2).

Figure 2: RM Goal, Action Steps, and Implementation Targets

RM Goal: To increase student mathematics achievement		
RM Action Steps		Dallas ISD Action Steps
• Staff training • On-site coaching/support • Dashboard summaries		• District implements RM with fidelity • Teachers complete training • Schedules include required RM time
Teacher Training Targets	Student Hour Targets	Student Accuracy Target
• First-year teachers will be strongly encouraged to complete Qualification Course before end of first six weeks	• Students in grades two to four will spend at least 62 hours online for year • Students in grade five will spend at least 174 hours online for year	• Students will answer 75% or more of Guided Study Level A items correctly

RM History in Dallas ISD

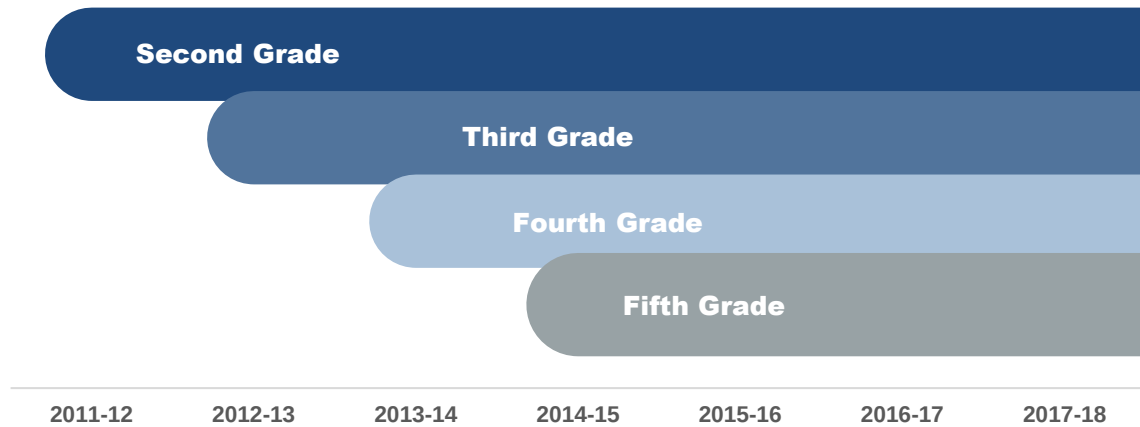
As shown in Figure 3, RM became available to district second-grade students in 2011-12. A new grade level was added each year from 2012-13 to 2014-15. Use of RM was required in grades two and three and optional in grades four and five until 2016-17, while it was optional for all grades in 2017-18. The number of campuses using RM at all grade levels was notably less in 2017-18 than previous years due to budget cuts for the year.

In 2017-18, the district used Title I funds and general operating funds to support a total of 57 elementary schools. This included 55 schools for second grade, 54 schools for third grade, 34 schools for fourth grade, and one school for fifth grade. Five schools purchased RM with their own budget.⁸ Campuses

⁸ Those schools included Carpenter, King, Peeler, Polk (grade five only), and Williams.

funded centrally for the 2017-18 school year were required to commit and adhere to the implementation requirements (i.e., meeting hour/accuracy targets).

Figure 3: Timeline for Expansion of RM in Dallas ISD



RM Curriculum

Teacher use of RM in grades two to four was supplemental to regular classroom instruction, whereas RM served as the complete curriculum in grade five. Per the RM website, “A Reasoning Mind classroom is a hybrid of online and face-to-face instruction, where the teacher gives each child individual help and attention.” Students used individual computers to work through mathematics lessons at their own pace.⁹ When a student struggled with a problem, a request for the “Genie Solution” could be made. The “Genie Solution” provided an explanation for computing the problem. The teacher could view the RM administrator screen to note which students were having difficulties. Based on that information, the teacher could provide one-on-one interventions or small-group tutorials. In Spring 2016, RM piloted online live tutoring¹⁰ for fifth-grade students. It was extended to all grade levels beginning with Fall 2016.

Since RM was introduced in the district, there have been concerns about lack of alignment between RM and district curricula (Kim & Bush, 2017). Even though both RM and district curricula were 100 percent aligned to the state standards, the order in which the standards were taught throughout the year varied. Therefore, it was possible that the instruction a teacher delivered on a particular day might be different than the instruction the students received in the RM program on the same day.¹¹ To alleviate those concerns in 2016-17, RM modified the use of RM curriculum in the district by introducing single objective curriculum

⁹ A personalized path is developed for each child based on an assessment that identifies strengths and weaknesses.

¹⁰ Tutors were RM employees based in the U.S. who were trained on the RM curriculum, RM system, and metrics. Tutors could invite identified students to a tutoring session during regular classroom hours. Tutors provided one-on-one instruction via voice, chat, and interactive blackboard. After a tutoring session, tutors could assign a problem set to assess understanding, or students were able to return directly to Guided Study for the rest of class.

¹¹ Source: Dallas ISD Reasoning Mind Update, prepared for the April 2016 Board meeting presentation by the STEM Mathematics Director.

(SOC), which allowed teachers to customize the RM curriculum to the district curriculum. RM provided SOC webinars for teachers in 2016-17 and SOC PD sessions in 2017-18.

RM Training and Support

Until 2016-17, first-year RM teachers were required to take the RM Qualification Course (QC) before they could use RM with students. However, due to delayed processing of the purchase orders for RM in Fall 2017, first-year RM teachers in 2017-18 were not required, but were strongly encouraged, to take the QC before using RM with students.

Additional training was available to all teachers along with campus-based support from a RM implementation coordinator (IC). There were 14 district PD sessions related to RM in 2017-18. RM ICs held two group sessions at each campus based on the needs that were identified at the initial meeting with campus administrators. In addition, RM ICs held two individual meetings with each RM teacher.

What were the demographic characteristics of students involved in Reasoning Mind?

Methodology

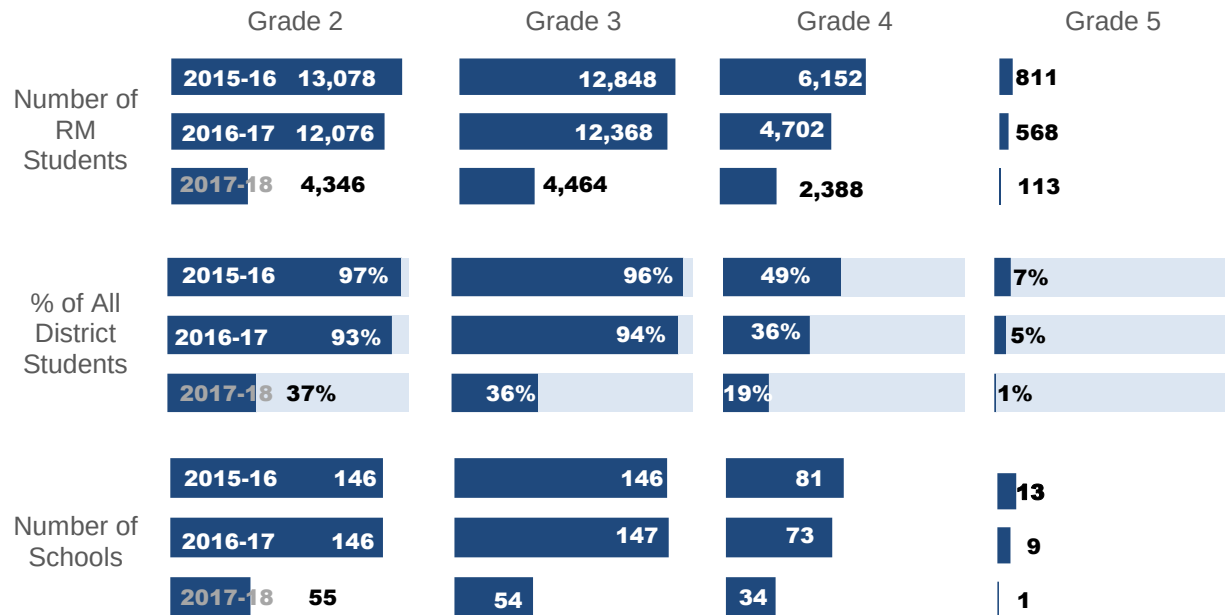
RM student data from the Dallas ISD Public Education Information Management System (PEIMS) database were merged into the RM student implementation data files. Frequency and crosstab analyses were computed for student demographic characteristics overall, by grade, by feeder pattern, and by student group.

Results

A total of 11,311 students participated in RM during 2017-18.¹² This included 4,346 (37%) of the district's second-grade students, 4,464 (36%) of the district's third-grade students, 2,388 (19%) of the district's fourth-grade students, and 113 (1%) of the district's fifth-grade students (see Figure 4). The percentage of district students enrolled in RM decreased from 2015-16 to 2016-17 to 2017-18 for all grades.

¹² RM students were defined as students who were enrolled in grades two to five as of October 30, 2017 (PEIMS snapshot date) and logged into the RM system at least 0.1 hour during 2017-18.

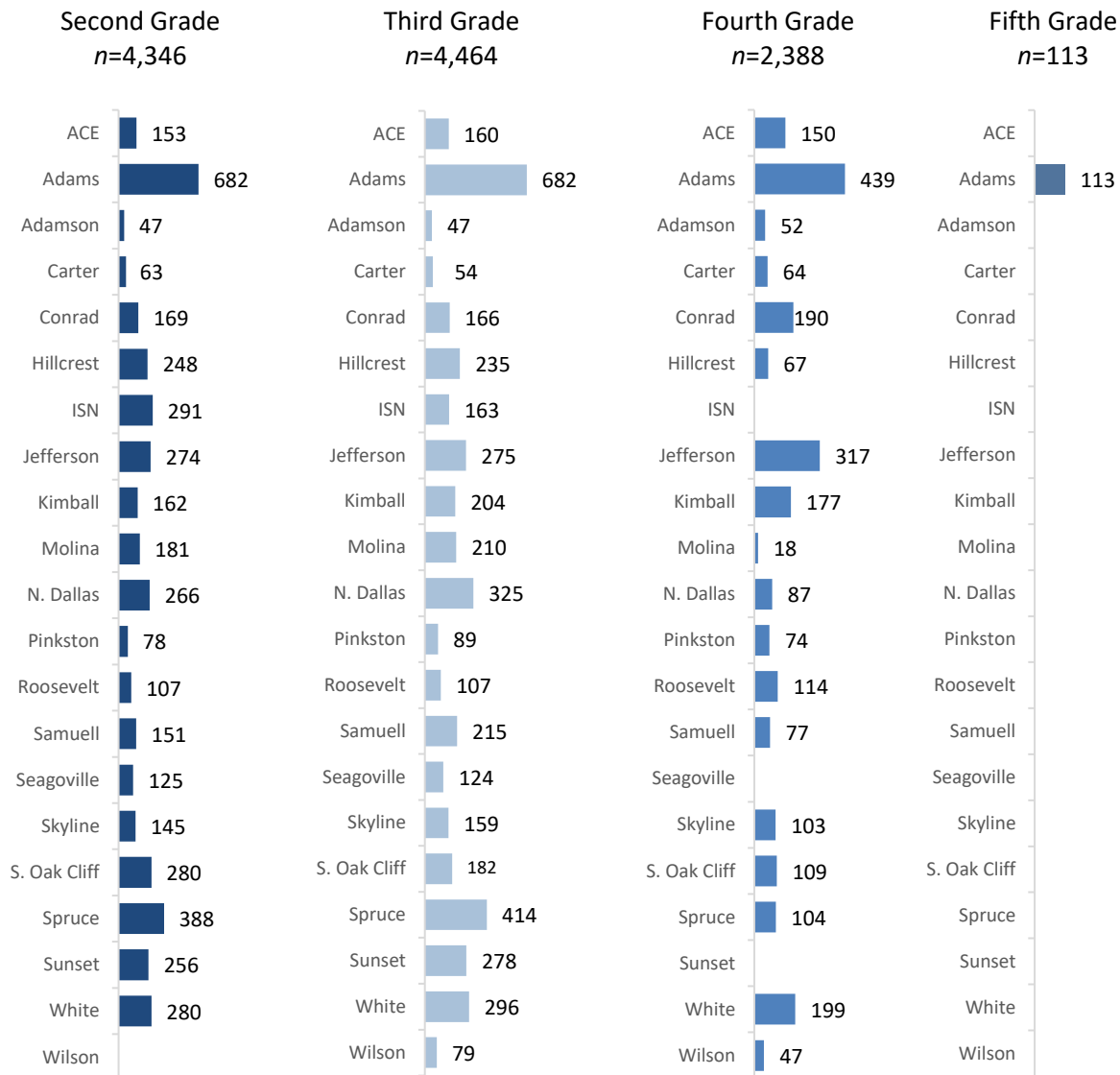
Figure 4: Student and School Participation in RM (2015-16 to 2017-18)



Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018).

When studied by feeder pattern, RM enrollment varied widely (see Figure 5). Within feeder patterns using RM, the number of second-grade students ranged from 47 in the Adamson feeder to 682 in the Adams Feeder. Similarly, the number of third graders ranged from 47 in Adamson and 682 in Adams. RM participants for fourth grade ranged from 18 (Molina feeder) to 439 (Adams feeder). As for fifth grade, only one school in the Adams feeder participated in 2017-18.

Figure 5: 2017-18 Number of RM Students by Grade and Feeder Pattern

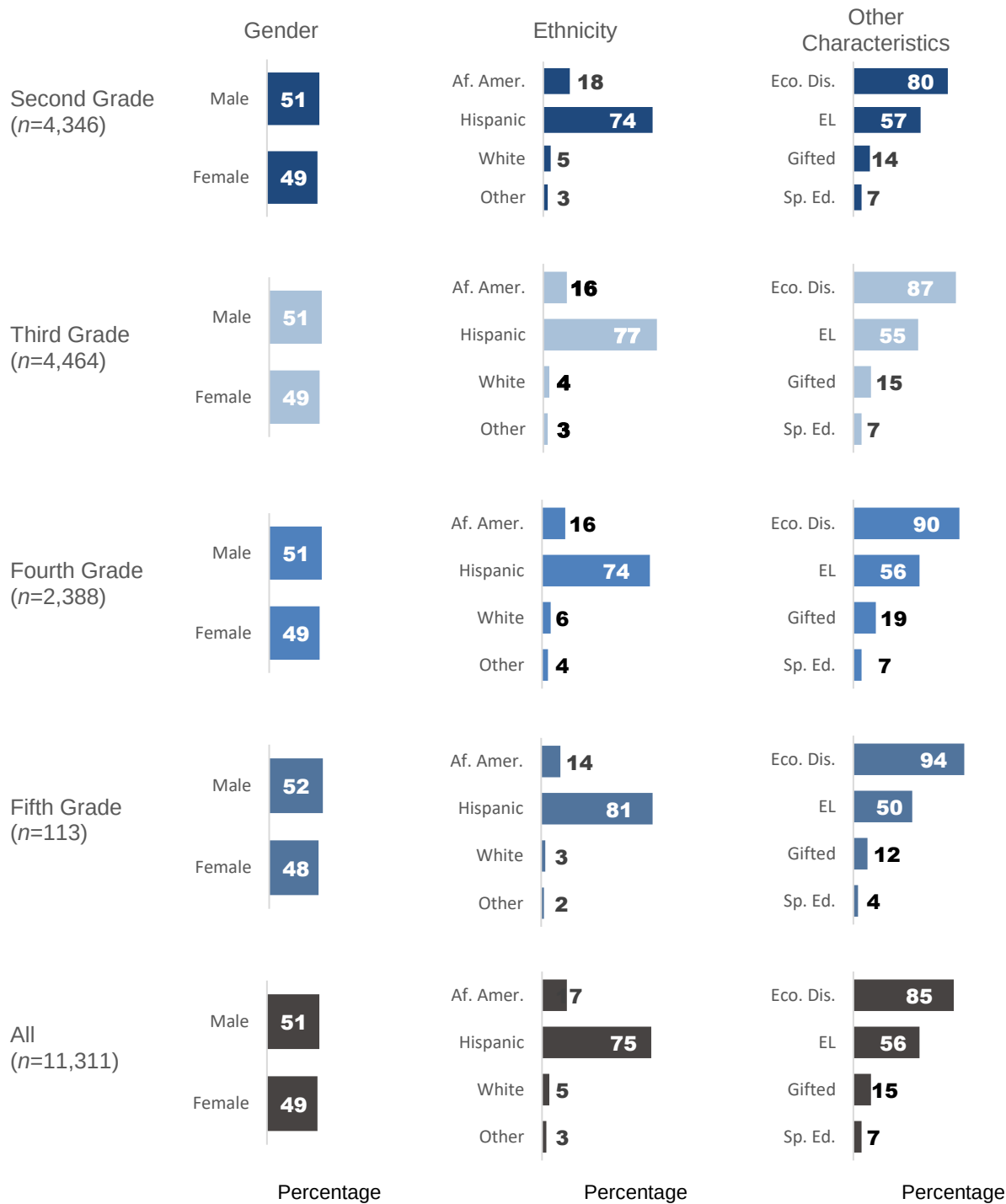


Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018).

Note: ACE=Accelerating Campus Excellence. ISN=Intensive Support Network. N. Dallas=North Dallas. S. Oak Cliff=South Oak Cliff. Analyses were limited to students in the RM database that were in Dallas ISD as of PEIMS snapshot date (10/30/2017) and with RM hours equal to or greater than 0.1 in 2017-18. Five schools that purchased RM with their own budget were excluded from analyses.

As seen in Figure 6, the majority of RM students were economically disadvantaged (85%) and Hispanic (75%). Fifty-six percent were English learners (ELs). When reviewed by grade level, demographic results were similar. Appendix A contains demographic information by grade level and student group.

Figure 6: 2017-18 Demographic Characteristics of RM Students



Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018).

Note: Af. Amer.=African American. Eco. Dis.=economically disadvantaged. EL=English learner. Sp. Ed.=special education. Other ethnicity included Asian, American Indian, Native Hawaiian or Other Pacific Islander, two or more races, and not available.

Was the Reasoning Mind program used by students as planned?

Methodology

Based on RM fidelity expectations, district central administrators set two student targets to ensure a high level of student implementation. Targets related to spending the expected amount of time online to move through RM objectives and answering an adequate percentage of mathematics items correctly.

- 1) Students will meet 2017-18 grade-level hour expectations (62 hours for grades two to four and 174 hours for grade five).
- 2) Students will answer 75 percent of RM Guided Study Level A problems correctly.¹³

RM provided the evaluators with student data files for 2017-18. The files included student time spent online using RM, number of objectives completed, and Guided Study Level A accuracy rates. Frequency and descriptive analyses were conducted to determine district progress on RM hour and accuracy targets overall, by grade level, and by campus and grade level. Next, high, medium, and low implementation campus groups were determined. *High* implementing campuses for each grade were selected based on the following criteria: (a) campus hour mean was at least 80 percent of the expected district hour target (target = 62 hours or higher for grades two to four and 174 or higher for grade five), (b) 50 percent or more of students at the campus met the hour target, (c) campus accuracy mean was 60 percent or higher (target=75% or higher), and (d) 50 percent or more of students at the campus met the district accuracy target. *Low* implementing campuses for each grade were selected based on the following criteria: (a) campus hour mean was less than 80 percent of the hour target, and (b) 49 percent or fewer of students at the campus met the hour target. Remaining campuses were classified as *medium* implementing campuses.

Results

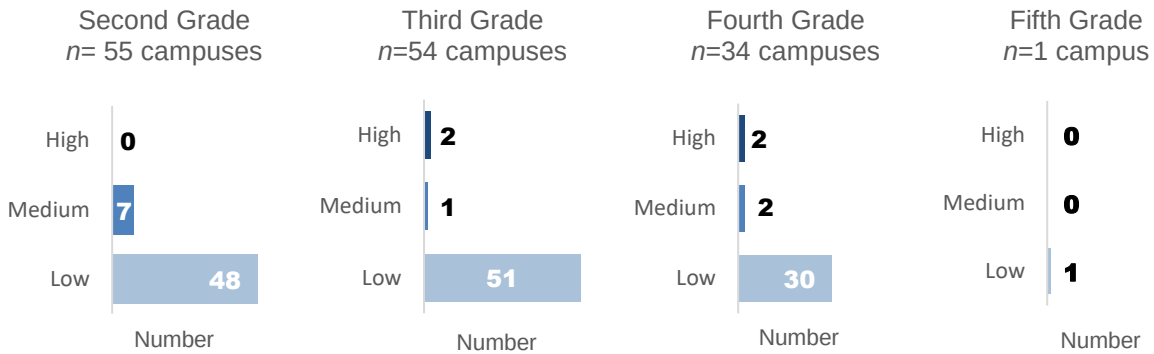
School-Level Implementation

As in previous years, review of RM implementation data by campus revealed great variation within schools and within grade levels at the same schools (see Appendix B). Key findings related to school-level implementation follow (see Figure 7 and Appendices C, D, and E).

- The majority of campuses for second (87%), third (94%), fourth (88%), and fifth (100%) grades were classified as low implementation campuses.
- The large proportion of campuses classified as “low implementation” (87% to 100%) is of concern; this is especially true for fifth grade because RM serves as the core curriculum.

¹³ The district chose 75 percent as the accuracy target because previous evaluation results showed that students with high accuracy rates performed best on standardized assessments. The RM system includes mathematics problems in learning mode (Guided Study), review mode (Wall of Mastery), and test mode (STAAR preparation). In the learning and review modes, problems increase in difficulty from Level A (easiest) to Level C (hardest). The accuracy target was based on Guided Study Level A items.

Figure 7: 2017-18 Campus-Level RM Implementation



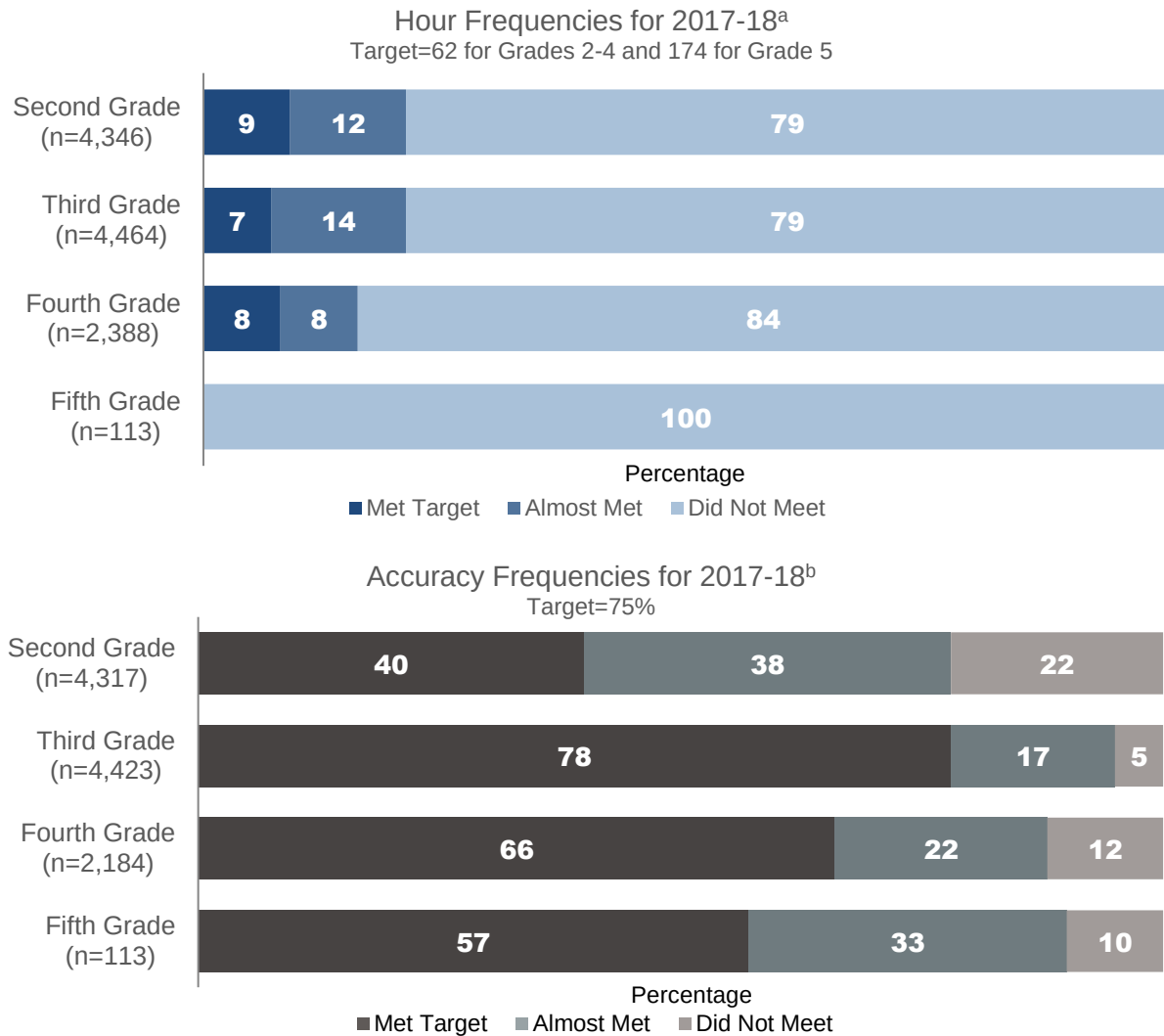
Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018).

Note: High implementing campuses for each grade were selected based on campus mean statistics, specifically based on the following criteria: (a) campus hour mean was at least 80 percent of the expected hour target, (b) 50 percent or more of students at the campus met the hour target, (c) campus accuracy mean was 60 percent or higher, and (d) 50 percent or more of students at the campus met the accuracy target. Low implementing campuses for each grade were also selected based on campus mean statistics, specifically based on the following criteria: (a) campus hour mean was less than 80 percent of the expected hour target, and (b) 49 percent or fewer of students at the campus met the hour target. The campuses that were not categorized as having either high or low level RM implementation were considered medium level RM implementation schools. Hour target=62 for grades two to four and 174 for grade five. Accuracy (Guided Study Level A) target=75 percent or higher.

RM Student-Level Implementation

Consistent with past years, a sizeable percentage of students (91% to 100%) in all grades did not reach the grade-level hour target for 2017-18 (target = 62 hours for grades two to four and 174 hours for grade five). As seen in Figure 8, there were more fourth- and fifth-grade students (84% and 100%, respectively) in the “did not meet” category than second- and third-grade students (79% for both grades). For the accuracy target, more than 50 percent of students in grades three to five (57% to 78%) met the target, while 40 percent of second-grade students met the accuracy target. The continued low level of implementation was surprising since most 2017-18 RM campuses had a medium or high level of implementation in 2016-17.

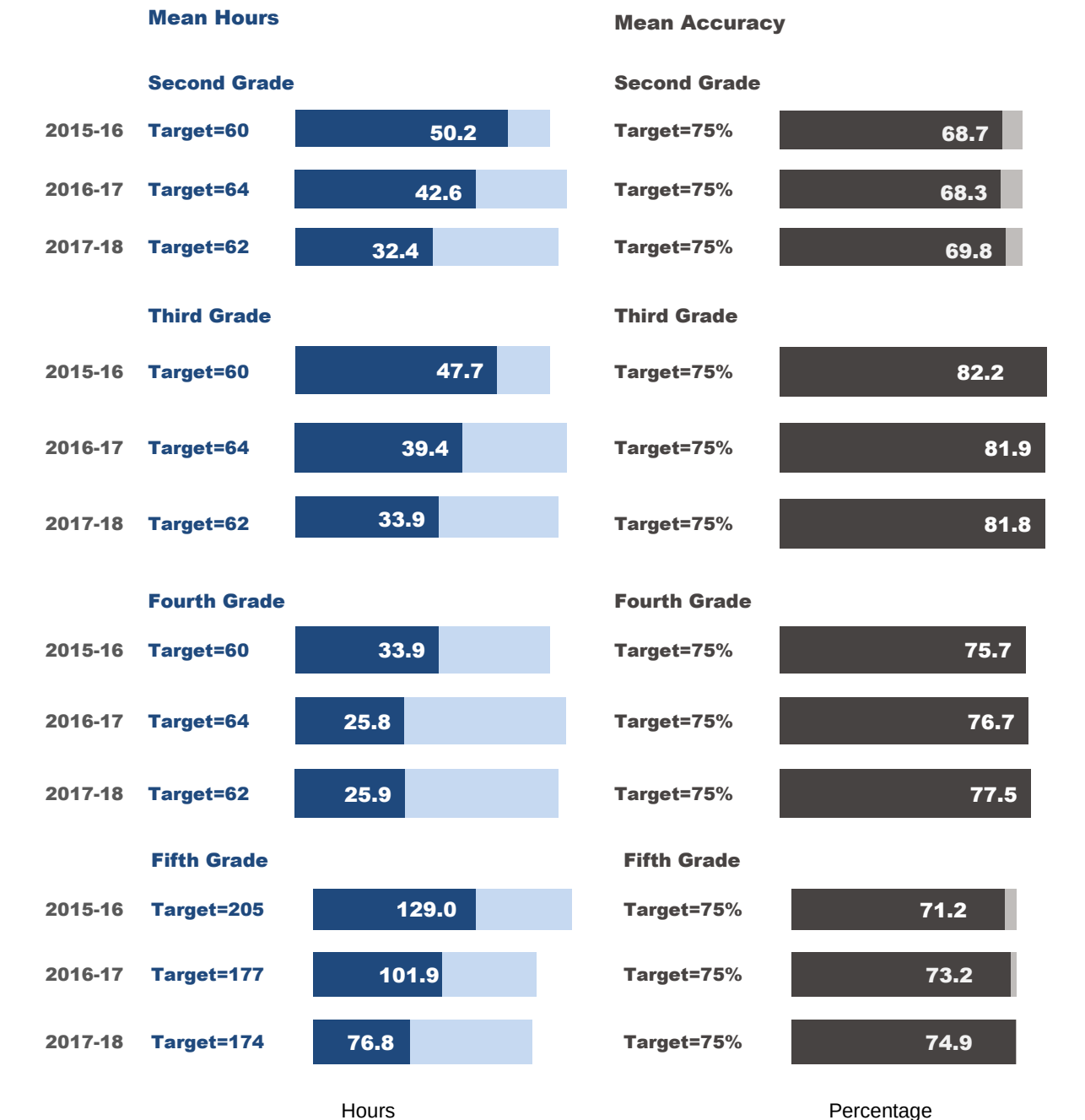
Figure 8: 2017-18 RM Student Hour and Accuracy Results by Grade Level



Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018).

Note: The number of students with hour and accuracy data differ for each grade because some students did not have Level A Guided Study accuracy data. ^a Met=Logged 100 percent of expected hours. Almost Met=Logged 80 to 99 percent of expected hours. Did Not Meet=Logged less than 80 percent of expected hours. ^b Met=Attained Guided Study Level A accuracy rate of 75 percent or higher. Almost Met=Accuracy rate between 60 and 74 percent. Did Not Meet=Accuracy rate of less than 60 percent.

When reviewed over time (2015-16 to 2017-18), student hours spent online for all grades decreased over time except for fourth grade (see Figure 9 and Appendix F). For fourth grade, student hours spent online decreased from 2015-16 to 2016-17, and then remained the same from 2016-17 to 2017-18. Students in grades three to five met the accuracy target with two exceptions (grade five in 2015-16 and 2016-17). However, second-grade mean accuracy rates were below the target for the last three school years.



Source: PEIMS district database (10/30/2015, 10/31/2016, and 10/30/2017). RM implementation data (1/15/2016, 1/12/2017, and 1/04/2018 for fall and 6/16/2016, 6/19/2017, and 6/05/2018 for spring).

Note: The light blue shaded area of the graph shows the deficit in expected time (if any), whereas the light brown shaded area shows the deficit in expected time (if any).

Did staff members participate in Reasoning Mind training as planned?

Methodology

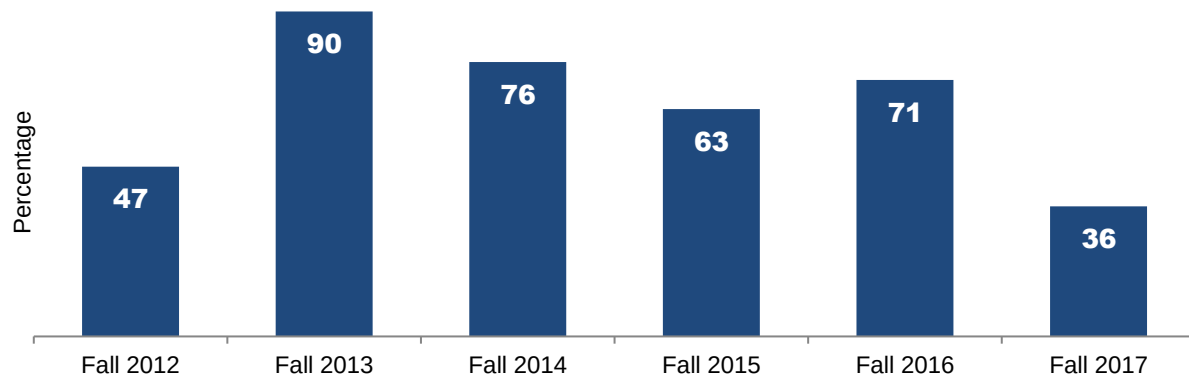
RM provided PD data that included Fall 2017 training information. Frequency analyses were conducted to determine RM training completion rates. Qualification Course (QC) completion data were reviewed overall, by time period, and by district feeder pattern. Overall attendance for other RM PD was also reviewed.

Results

Qualification Course for Teachers

Until 2016-17, first-year RM teachers were required to take the RM Qualification Course (QC) prior to using RM with students. However, due to delayed processing of the purchase orders for RM in Fall 2017, first-year RM teachers in 2017-18 were not required but were strongly encouraged to take the QC before using RM with students. For this reason, only 36 percent of first-year RM teachers completed the QC by the end of the first six weeks, which was a notable decrease from last year (71% in Fall 2016 - see Figure 10). Half (50%) of the first-year RM teachers did not complete the QC by the end of the fall semester (see Appendix G for detailed QC information).

Figure 10: Percentage of First-Year RM Teachers that Completed the QC by the End of the First Six Weeks



Source: RM PD data (01/10/2018). Information for past years were copied from the 2016-17 RM final report (Kim & Bush, 2017).

Note: QC=Qualification Course. These numbers were current as of the end of the fall semester.

Additional Professional Development

RM offered additional training courses for all RM teachers. About eight percent ($n=28$) of the district's 344 RM teachers¹⁴ participated in optional RM PD during the fall semester. All 28 teachers used

¹⁴ The number of RM teachers was based on the RM teacher list from RM program management as of January 10, 2018. The number included only teachers from schools that were funded centrally.

supplemental RM (grades two to four). There were two campus-based mathematics instructional coaches who attended RM's Best Practice Workshop courses.

What were teacher and administrator perceptions of Reasoning Mind?

Methodology

Interview. The lead evaluator conducted interviews with the RM project director in December 2017 and the district mathematics director in January 2018. The purpose of the interviews was to understand communication between RM and district leaders, changes in RM implementation, resources and support provided, successes and challenges.

Campus Staff Survey. The *Reasoning Mind Campus Administrator Survey* and *Reasoning Mind Teacher Survey* were administered during a three-week window between January 16 and February 2, 2018. The purpose of the surveys was to find out staff perceptions of RM. Survey items related to overall perceptions of RM, district/RM support, student change as a result of using RM, technology issues, successes, and challenges. For items related to successes and challenges, the evaluators made a list based on prior years' thematic results¹⁵ of open-ended campus staff survey items; teachers and campus administrators could mark more than one category from the list. Details about survey development including validity and reliability results can be found in the RM 2013-14 final report.¹⁶

Campus administrators and teachers involved in RM during 2017-18 were sent an email with the survey link along with periodic follow-up reminders. Response rates were 61 percent for campus administrators and 64 percent for teachers. Upon removal of invalid responses, 81 campus administrators and 212 teachers were included in the analyses. Frequency and crosstab analyses were conducted for rating-scale items and multiple response items on the two surveys.

As mentioned earlier, the number of RM schools funded centrally decreased from 147 in 2016-17 to 57 in 2017-18. Criteria to select schools funded centrally were based on 2016-17 campus level RM implementation for each grade (schools that were categorized as having a high or medium level of RM implementation in 2016-17). Because of these reasons, the results for 2017-18 should be reviewed with caution when compared to previous years.

Results

Interview Results

Interviewees were asked to describe 2017-18 successes. The RM project director and the district mathematics director viewed more communication and support by RM ICs as a main success. As mentioned previously, in 2017-18, RM ICs held two group sessions at each campus based on need and held two individual meetings with each RM teacher. The RM project director also mentioned the single

¹⁵ Bush and Kim (2015, 2016).

¹⁶ Survey development, reliability, and validity information is described on pages 22 and 68-70 in the 2013-14 Dallas ISD Reasoning Mind final report by Bush and Kim (2014): <http://www.dallasisd.org/cms/lib/TX01001475/Centricity/domain/98/evaluation/13-14/finalrpts/EA14-512-2-Reasoning-Mind-Program-Final.pdf>.

objective curriculum (SOC) as a success. SOC allows teachers to customize the RM curriculum to the district curriculum.

Both interviewees indicated barriers to implementation. The RM project director and the district mathematics director noted that technology problems (i.e., problems with Computers on Wheels (COW) carts, aging of technology, wireless connectivity issues, etc.) continued for some campuses. They also indicated the delayed processing of purchase orders and the district's change in its data management tool to CLEVER.¹⁷ The district mathematics director indicated the reduced budget as another barrier to implementation. The RM project director mentioned fewer ICs as additional barrier to RM implementation. The number of RM ICs decreased from 11 in 2015-16 to six in 2016-17 and then to three in 2017-18.

Campus Survey Results - Overall Perceptions of RM

Responses of 81 campus administrators and 212 teachers were included in analyses. Of the 81 administrators, 39 (48%) were campus instructional coaches, 29 (36%) were principals, 10 (12%) were assistant principals, and 3 (4%) were "Other." Seventy-four percent indicated the primary RM campus administrator contact was a campus instructional coach; others reported the contact as the principal (12%), assistant principal (9%), or other staff member (5%). As seen in Table 1, the majority of teacher respondents were second-grade teachers (44%) or third-grade teachers (35%). Five percent of teacher respondents indicated that they taught more than one grade. About half of teacher respondents had used RM one to two semesters (26%) or nine or more semesters (27%).

Table 1: Spring 2018 Campus Staff Survey Respondent Characteristics

Group	Administrators n (%)	Teachers n (%)
Schools Represented	52 (91)	54 (95)
Grade Level		
Second only	-	94 (44)
Third only	-	73 (35)
Fourth only	-	34 (16)
Fifth only	-	0 (0)
Combination	-	11 (5)
Number of Semesters Used RM		
1-2	-	56 (26)
3-4	-	35 (17)
5-6	-	28 (13)
7-8	-	36 (17)
9 or more	-	57 (27)

Source: RM campus staff (administrator and teacher) surveys administered in Spring 2018.

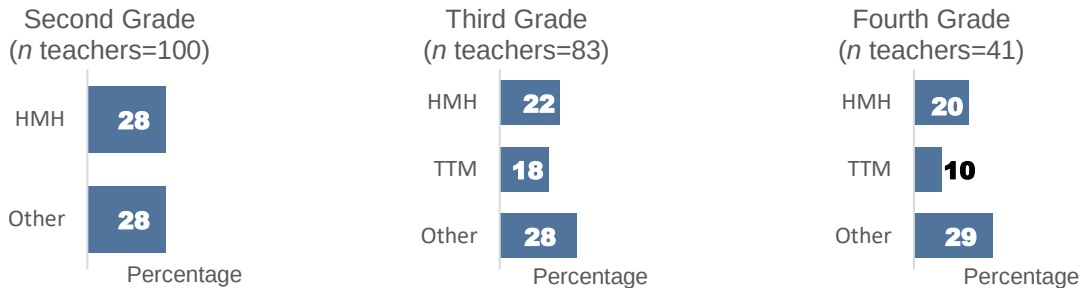
Note: Valid responses were received from 81 campus administrators and 212 teachers. The number of schools that enrolled students in RM during 2017-18 was 57.

Between 10 percent and 29 percent of teacher respondents indicated that they used other mathematics intervention programs in addition to RM (see Figure 11). Between 20 percent and 28 percent of teacher respondents indicated that they used Houghton Mifflin Harcourt (HMH) Personal Math Trainer,

¹⁷ CLEVER is a data management tool that allows single sign-in to all applications including RM. There were sync issues between CLEVER and the RM program.

while 10 percent to 18 percent of teacher respondents indicated that they used Think Through Math (TTM). The “other” category included numerous web resources and intervention programs.¹⁸

Figure 11: Percentage of Teachers that Used Other Mathematics Intervention Programs by Grade



Source: RM teacher survey administered in Spring 2018.

Note: TTM=Think Through Math. HMH=Houghton Mifflin Harcourt (HMH) Personal Math Trainer. Results for fifth grade were not reported in this figure due to the small number of fifth-grade teacher respondents ($n < 5$).

Across grades from second to fourth,¹⁹ more than 95 percent of RM teacher respondents indicated that they used RM at least once a week (see Table 2 and Appendix I). It is noticeable that between 26 percent and 75 percent of teacher respondents also used HMH and/or TTM at least once a week (range: 26% to 61% for HMH and 73% to 75% for TTM). Over 75 percent of teacher respondents indicated that they also used other supplemental mathematics resources/programs (i.e., Istation Math, Motivation Math, etc.) at least once a week.

Table 2: Percentage of Teachers that Used Various Mathematics Intervention Programs by Grade

How often did you use the following?		More than once a week %	Once a week %	At least once a month %	Less than once a month %
Second Grade					
RM	(n = 99)	78	20	1	1
HMH	(n = 28)	29	32	3	36
Other	(n = 28)	89	7	4	0
Third Grade					
RM	(n = 82)	83	15	0	2
TTM	(n = 15)	40	33	7	20
HMH	(n = 18)	39	22	17	22
Other	(n = 23)	52	35	9	4
Fourth Grade					
RM	(n = 40)	78	17	5	0
TTM	(n = 4)	50	25	25	0
HMH	(n = 8)	13	13	50	24
Other	(n = 12)	67	8	25	0

Source: RM teacher survey administered in Spring 2018.

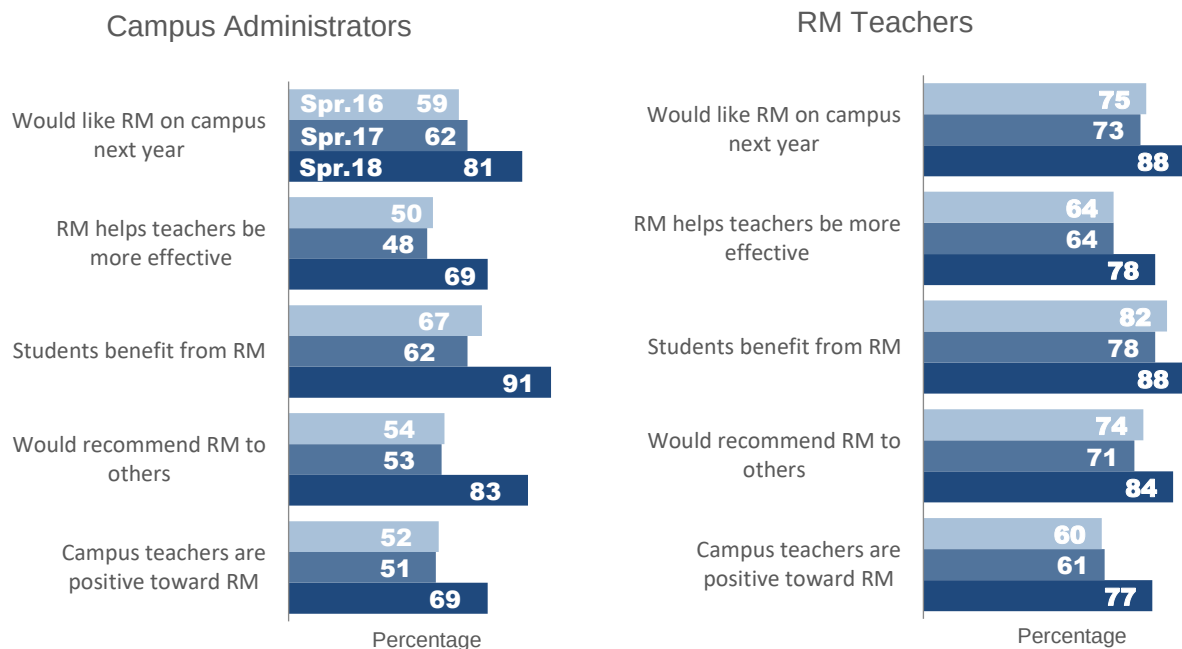
Note: TTM=Think Through Math. HMH=Houghton Mifflin Harcourt (HMH) Personal Math Trainer. Results for fifth grade were not reported in this table due to the small number of fifth-grade teacher respondents ($n < 5$).

¹⁸ “Other” included Abcya, Brain Pop, Education Galaxy, Istation Math, IXL Math, Khan Academy, Learning Farm, MobyMax, Motivation Math – Mentoring Minds, Prodigy, Splash Math, Ten Marks, TEKSING TOWARD STAAR, Thinkcentral, XtraMath, etc.

¹⁹ Fifth grade was not analyzed due to the small number of fifth-grade teacher respondents ($n < 5$).

As in previous years, teachers were more positive than campus administrators on the overall survey items with one exception (the benefit of RM on students: 91% for campus administrators and 88% for teachers - see Figure 12 and Appendices H and I). Whereas 69 percent to 91 percent of administrators gave positive responses to overall items, 77 percent to 88 percent of teachers did so. Administrators were most positive toward the benefit of RM on students (91%), while teachers were most positive toward the benefit of RM on students (88%) and the continuation of RM next year (88%). For both campus administrators and teachers, positive responses to overall items were higher in Spring 2018 than in Spring 2016 or Spring 2017. Because of the reasons mentioned earlier,²⁰ the results for 2017-18 should be reviewed with caution when compared to previous years.

Figure 12: Percentage of Campus Staff Members that “Agreed” or “Strongly Agreed” with Overall Items



Source: RM campus staff (administrator and teacher) surveys administered in Spring 2018.

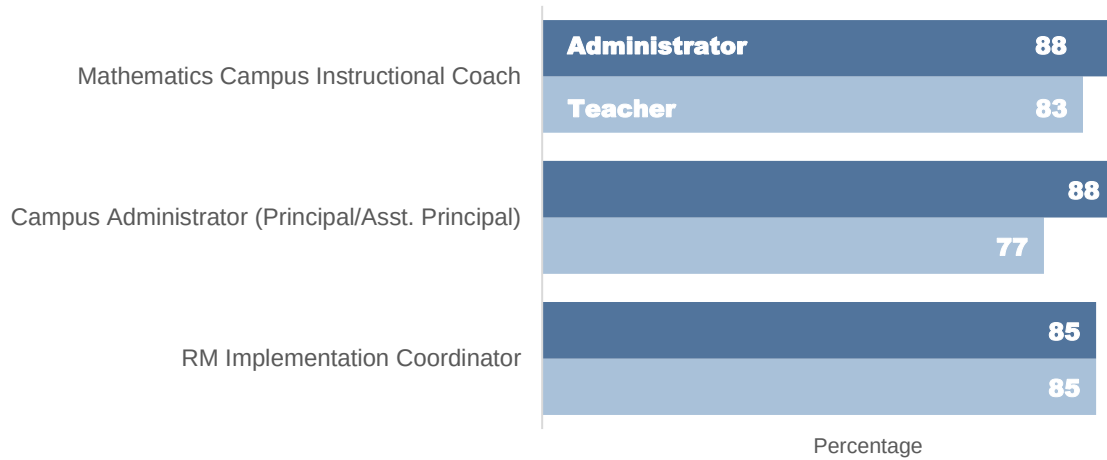
Note: Spr.16=Spring 2016. Spr.17=Spring 2017. Spr.18=Spring 2018.

Campus Survey Results - RM and District Staff Support

As shown in Figure 13, 85 percent of campus administrators and teachers responded that they were “satisfied” or “extremely satisfied” with support from the RM implementation coordinator (see Figure 13). Campus administrators (88%) were more positive than teachers (77% to 83%) about support from district staff members (other campus administrators and mathematics campus instructional coaches).

²⁰ The number of RM schools funded centrally decreased from 147 in 2016-17 to 57 in 2017-18. In addition, criteria to select schools funded centrally were based on 2016-17 campus level RM implementation for each grade (schools that were categorized as having a high or medium level of RM implementation in 2016-17).

Figure 13: Percentage of Campus Staff Members that Were “Satisfied” or “Extremely Satisfied” with RM and District Staff Support



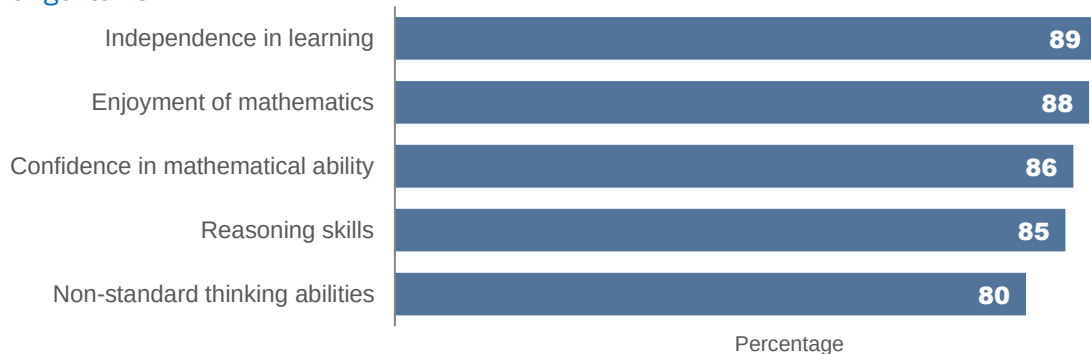
Source: RM campus staff (administrator and teacher) surveys administered in Spring 2018.

Note: Participants that marked “not applicable” were excluded.

Campus Survey Results - Student Change Indicators

The majority of teachers indicated that students had “improved” or “significantly improved” in the five academic-related change areas queried (80% to 89% - see Figure 14).

Figure 14: Percentage of Teachers that Marked “Improved” or “Significantly Improved” for Student Change Items



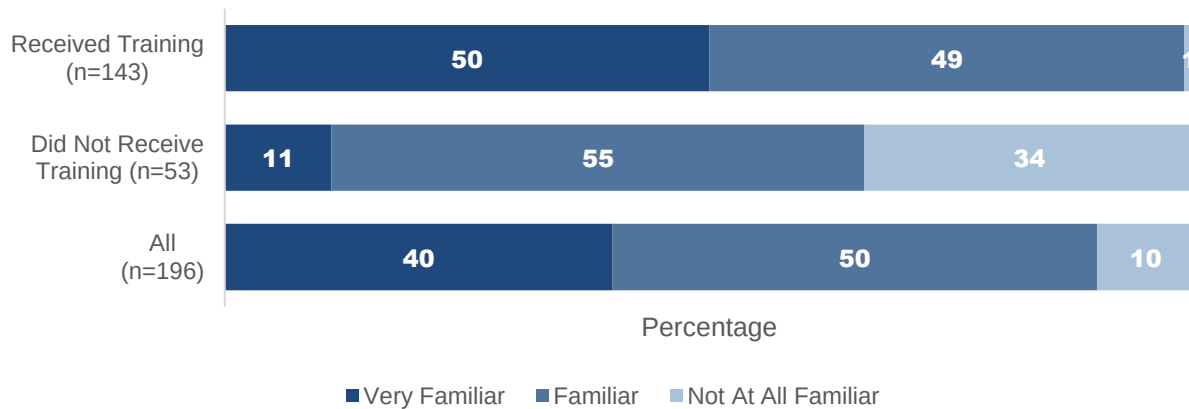
Source: RM teacher survey administered in Spring 2018.

Note: Participants that marked “do not know” were excluded.

Campus Survey Results – RM’s Single Objective Curriculum (SOC)

Of the teachers who responded on the SOC-related questions ($n=196$), 73 percent indicated that they had received the training on RM’s SOC. Overall, 90 percent of teachers responded that they were “somewhat familiar” or “very familiar” with RM’s SOC (see Figure 15). Of the teachers who received the training on RM’s SOC ($n=143$), 50 percent indicated that they were “very familiar” and 49 percent indicated that they were “somewhat familiar” with RM’s SOC.

Figure 15: Teachers' Familiarity with RM's SOC by Training Experience



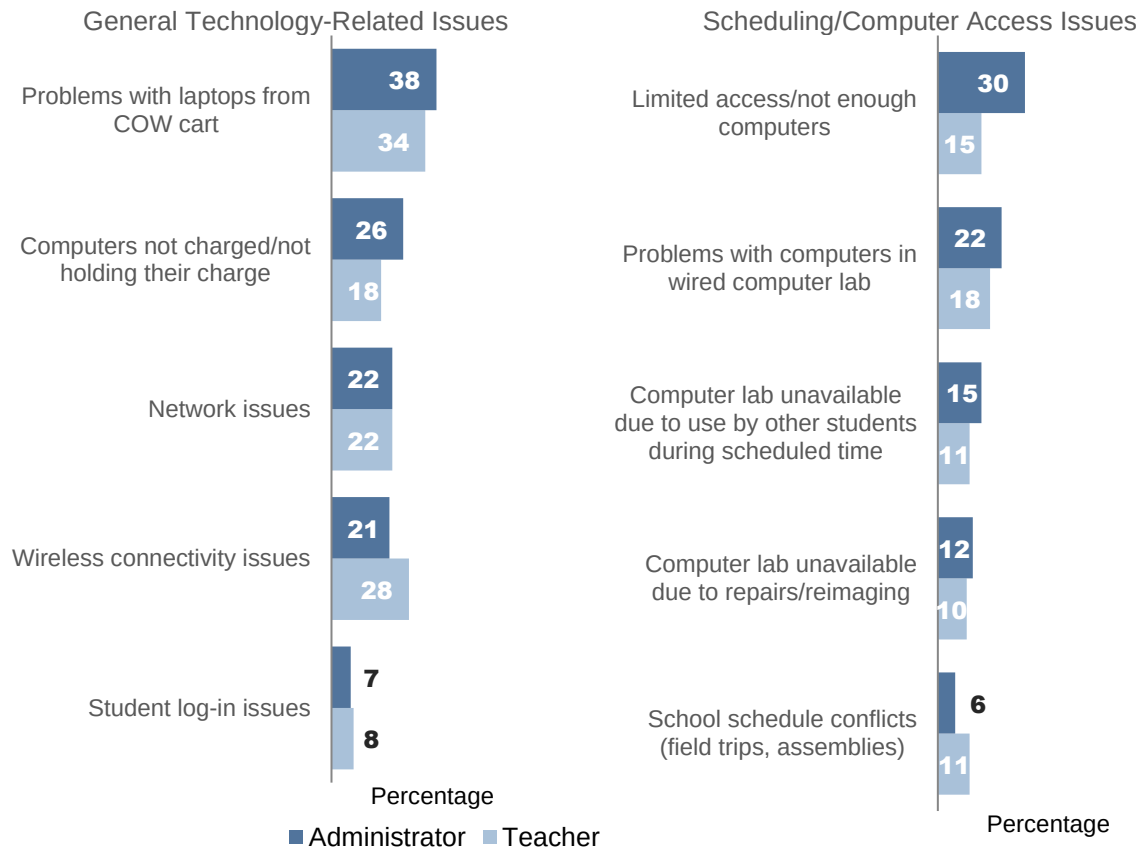
Source: RM teacher survey administered in Spring 2018.

Note: SOC=Single Objective Curriculum

Campus Survey Results - Technology-Related Issues and Computer Scheduling/Access Issues

Some campus administrators and teachers indicated that technology-related issues continued to “frequently” or “very frequently” occur. A fifth or more of campus administrators and teachers reported frequent issues with laptops from Computers on Wheels (COW) carts (38% for campus administrators and 34% for teachers), network (22% for both), and wireless connectivity (21% and 28%, respectively). Between six percent and 30 percent of campus staff members reported other technology issues (6% to 30% for campus administrators and 8% to 18% for teachers - see Figure 16).

Figure 16: Percentage of Campus Staff Members that Chose “Frequently” or “Very Frequently” for General Technology and Computer Scheduling/Access Issues



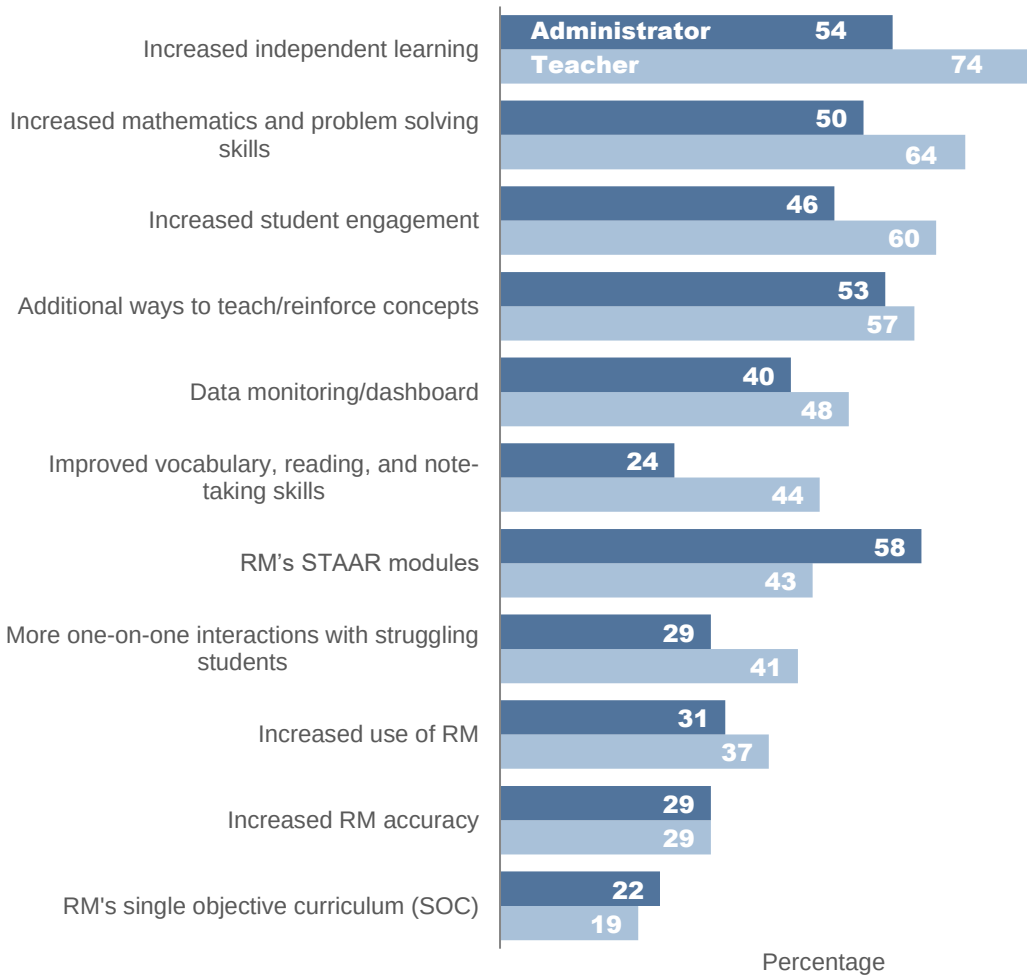
Source: RM campus staff (administrator and teacher) surveys administered in Spring 2018.

Note: COW cart= Computers on Wheels cart.

Campus Survey Results – Successes

In terms of RM successes, teachers (19% to 74%) were more positive than campus administrators (22% to 58%) with two exceptions (Figure 17 and Appendix J). Campus administrators were more positive than teachers on RM's *State of Texas Assessments of Academic Readiness (STAAR)* modules (58% and 43%, respectively) and RM's single objective curriculum (SOC – 22% and 19%, respectively). Twenty-nine percent of campus administrators and teachers indicated the increased RM accuracy as a success.

Figure 17: Campus Staff Survey Responses for RM Successes



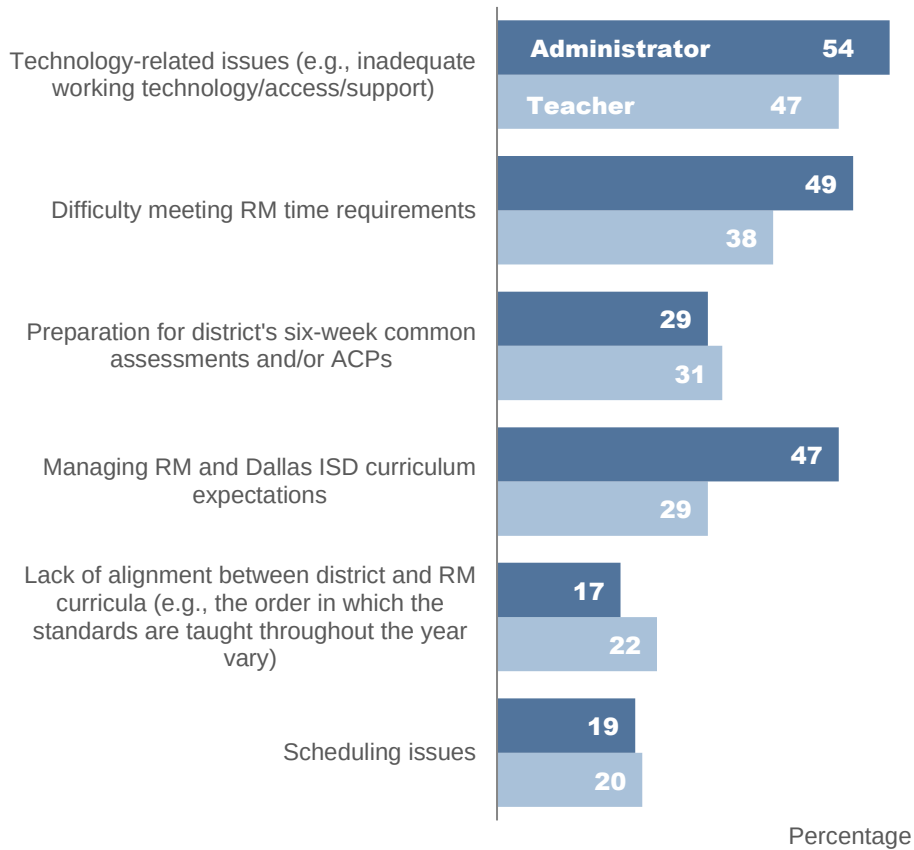
Source: RM campus staff (administrator and teacher) surveys administered in Spring 2018.

Note: Respondents could mark more than one category. This figure was limited to categories with 15% or more of respondents for each group ($n=194$ for teachers and $n=72$ for campus administrators).

Campus Survey Results – Challenges

As seen in Figure 18 and Appendix J, the most frequent challenges that campus administrators and teachers indicated were technology-related issues (54% for campus administrators and 47% for teachers) and difficulty meeting RM time requirements (49% and 38%, respectively). In addition, between 17 percent and 47 percent of respondents listed challenges such as preparation for district's six-week common assessments and/or *Assessments of Course Performance (ACPs)* (29% and 31%, respectively), managing RM and Dallas ISD curriculum expectations (47% and 29%, respectively), lack of alignment between district and RM curricula (17% and 22%, respectively), and scheduling issues (19% and 20%, respectively).

Figure 18: Campus Staff Survey Responses for RM Challenges



Source: RM campus staff (administrator and teacher) surveys administered in Spring 2018.

Note: Respondents could mark more than one category. This figure was limited to categories with 15% or more of respondents for each group ($n=198$ for teachers and $n=79$ for campus administrators).

What were student mathematics achievement outcomes?

Methodology

Outcome analyses were conducted to determine how RM students performed on spring mathematics assessments. Data from RM files, PEIMS files, and mathematics assessment files were merged to prepare for analyses. Assessments included *TerraNova*²¹ and *State of Texas Assessments of Academic Readiness (STAAR)*.²² Descriptions follow.

- *TerraNova* is a nationally normed achievement battery that is administered to district students in grades kindergarten to two. Normal Curve Equivalent (NCE) scores were used in descriptive, correlation, and regression analyses. National Percentile scores were used to determine students that met the grade-level standards; students at or above the fortieth percentile were considered performing at or above grade level.

²¹ Data reported as of 6/11/2018.

²² Data reported as of 6/20/2018.

- The state's criterion-referenced assessment, *STAAR*, is administered each spring. Crosstab analyses were used to determine the percentage of students in grades three to five that achieved the Approaches Grade Level or Above (Approaches +) and Meets Grade Level or Above (Meets +) performance standards. Scale scores were used in descriptive, correlation, and regression analyses. *STAAR* results included first administration data, and *STAAR Alternate 2* scores were excluded from the analyses. There were online testing issues that occurred in April and May 2018. Although all scores were included, some students' test scores might have been adversely affected by the online testing issues.

Descriptive, Correlation, and Crosstab Analyses

Descriptive analyses were carried out to determine mean scores (*TerraNova* NCE scores for second grade and *STAAR* scale scores for grades three to five), and frequency analyses were conducted to determine students that met the performance standards; analyses were conducted overall and by subgroup for each grade level. Pearson product correlation coefficients (r) were computed to note the strength of relationships between students' assessment scores and RM indicators (hours online, objectives completed, and Guided Study Level A accuracy). Crosstab analyses were used to further examine the relationship between time, accuracy, and student achievement.

Regression Analyses

Multiple regression analyses were conducted for each grade level to determine which RM variables best predicted students' spring achievement scores (*TerraNova* NCE scores for second grade and *STAAR* scale scores for grades three to five). Variables included: (1) number of objectives completed, (2) accuracy rates of Guided Study Level A problems, (3) prior achievement (2017 *TerraNova* NCE scores for second and third grade students and 2017 *STAAR* scale scores for fourth and fifth grade students), (4) economically disadvantaged status, (5) gender, and (6) ethnicity.²³

RM and Non-RM Student Comparisons

Propensity score matching (PSM) was used to create matched comparison groups of non-RM students (see Appendix K for detailed methodology).²⁴ After that, multiple regression analyses were carried out for each grade to determine if there were statistically and practically significant differences in mathematics achievement scores between RM and non-RM comparison group students.²⁵ For further analyses, frequencies and crosstab analyses were computed to examine differences between the two

²³ Evaluators replicated the regression model used by Waxman and Houston (2012). Time was excluded due to the high correlation between total hours and number of objectives completed (potential multicollinearity and low statistical power -Bobko, 2001). Dummy variables were used for economically disadvantaged status (with non-economically disadvantaged as the reference group), gender (with female as the reference group), and ethnicity (with White as the reference group).

²⁴ The matched comparison group for each grade was comprised of non-RM students in the district that had the same probability of being in RM as RM students. This procedure mimics random assignment as used in an experimental design.

²⁵ Prior achievement (2017 *TerraNova* NCE scores for second and third grade students and 2017 *STAAR* scale scores for fourth and fifth grade students), economically disadvantaged status, and RM/non-RM status were included as predictor variables in the regression analyses.

groups in percentages of students at or above grade level on *TerraNova* and that achieved Approaches +, Meets +, and Masters Grade Level (Masters) on *STAAR*. *STAAR* progress measures were also analyzed to examine how much fourth- and fifth-grade students improved this year.²⁶

Limitations of Outcome Analyses

As mentioned previously, there were four main limitations of the analyses. First, achievement results could not be attributed solely to RM due to the district's use of RM as supplemental to regular instruction in grades two to four (e.g., it is not possible to know how much achievement results should be attributed to RM and how much they should be credited to teacher instruction based on district curriculum). Second, district implementation fidelity continued to be a confounding factor as a sizeable proportion of students in all grades did not meet RM expectations for implementation. Third, there was possible selection bias between treatment and comparison groups for each grade. That is, the number of RM schools funded centrally decreased in 2017-18 because of budget cuts, and criteria to select supported schools were based on 2016-17 campus level RM implementation. It should be noted also that only one school was selected for fifth grade which may limit the generalizability of the findings. Fourth, per the Spring 2018 RM teacher survey, between 10 percent and 29 percent of teacher respondents used additional mathematics programs/resources such as TTM and HMH Personal Math Trainer.

Results

Descriptive Statistics for RM Student Spring Mathematics Performance

As seen in Tables 3 and 4, overall mean assessment scores for RM students were close to the district mean for grade two (mean difference=0.8), slightly higher for grades three and five (mean differences=7.6 and 2.8, respectively) and much higher for grade four (mean difference=23.6). Although the mean difference between RM and district students for grade four was quite large, the results should be reviewed with caution because RM students also had higher prior year (2017) *STAAR* mean scale scores (1472.4) than the district as a whole (1446.2).²⁷ When reviewed by subgroup, mean scale scores for African American students were lower than other ethnic groups, and mean scale scores for White students were higher than other groups.

²⁶ The progress measure is based on a comparison of a student's test score last year with his or her score this year. This measure informs how much the student has improved in a subject area.

²⁷ District information for last year was obtained from the 2016-17 RM evaluation report (Kim and Bush, 2017).

Table 3: Spring 2018 TerraNova Mathematics Results for Students in Grade 2

Group	RM Students				District			
	<i>n</i>	Mean	SD	% At or Above 40 th Percentile	<i>n</i>	Mean	SD	% At or Above 40 th Percentile
All	4,115	49.1	20.3	60.9	10,974	48.3	20.7	59.5
African American	700	43.0	21.2	49.4	2,317	42.0	20.8	47.9
Hispanic	3,101	49.5	19.6	62.2	7,712	48.9	19.9	60.9
White	177	60.9	21.1	77.4	660	61.2	21.0	80.6
Eco. Dis.	3,297	48.0	20.2	58.8	8,541	46.9	20.3	57.0
Special Education	294	33.7	22.1	33.0	784	32.4	22.2	31.3
English Learner	2,358	49.0	19.4	61.5	5,689	48.0	19.7	59.5

Source: PEIMS district data (10/30/2017). RM data (1/04/2018 and 6/05/2018). TerraNova (6/11/2018).

Note: SD=standard deviation. Eco. Dis.=economically disadvantaged. Normal Curve Equivalent (NCE) scores were used for TerraNova descriptive statistics, while National Percentile scores were used to determine students that met the grade-level standard (At or Above the 40th percentile).

Table 4: Spring 2018 STAAR Mathematics Results for Students in Grades 3 to 5

Group	RM Students					District Students				
	<i>n</i>	Mean	SD	% Appr.+	% Meets+	<i>n</i>	Mean	SD	% Appr.+	% Meets+
Third Grade STAAR										
All	4,269	1467.8	144.7	77.3	44.2	11,663	1460.2	145.6	75.2	42.4
African American	642	1409.1	138.4	62.0	27.6	2,422	1406.1	134.0	61.9	28.0
Hispanic	3,338	1474.6	140.2	79.9	45.9	8,386	1469.1	141.5	78.2	44.6
White	183	1520.0	175.1	82.0	61.2	578	1539.8	167.0	85.6	66.3
Eco. Dis.	3,717	1461.7	142.0	76.5	42.3	10,110	1453.0	141.9	74.1	40.4
Spe. Ed.	287	1348.6	130.6	41.8	16.0	785	1347.7	136.8	41.0	15.9
English Learner	2,402	1473.5	142.2	79.2	45.8	5,930	1466.5	142.9	77.4	44.1
Fourth Grade STAAR										
All	2,284	1583.4	151.4	79.2	51.0	11,854	1559.8	144.6	75.5	44.7
African American	356	1515.1	127.9	67.1	30.6	2,497	1500.8	126.0	62.9	27.4
Hispanic	1,708	1594.6	150.1	81.4	54.5	8,516	1571.4	143.1	78.4	48.1
White	124	1626.6	168.8	86.3	62.9	579	1635.0	155.3	87.0	67.0
Eco. Dis.	2,047	1579.2	150.6	78.4	49.7	10,684	1554.7	142.7	74.7	43.3
Spe. Ed.	147	1451.3	123.7	42.9	17.7	850	1440.1	130.7	38.8	16.0
English Learner	1,296	1586.9	156.3	78.0	52.2	6,018	1568.1	145.0	77.4	47.0
Fifth Grade STAAR										
All	111	1627.9	130.0	83.8	56.8	11,785	1625.1	142.3	81.5	54.2
African American	16	1581.8	98.9	75.0	43.8	2,441	1564.0	132.2	67.7	36.3
Hispanic	91	1636.2	133.5	84.6	60.4	8,552	1637.9	139.4	84.8	57.9
White	*	*	*	*	*	537	1684.3	139.4	90.5	73.2
Eco. Dis.	104	1632.3	132.7	83.7	59.6	10,677	1621.4	141.4	80.8	53.2
Spe. Ed.	*	*	*	*	*	886	1517.6	135.0	50.9	24.4
English Learner	55	1658.9	143.8	85.5	65.5	6,019	1637.1	141.0	84.4	57.7

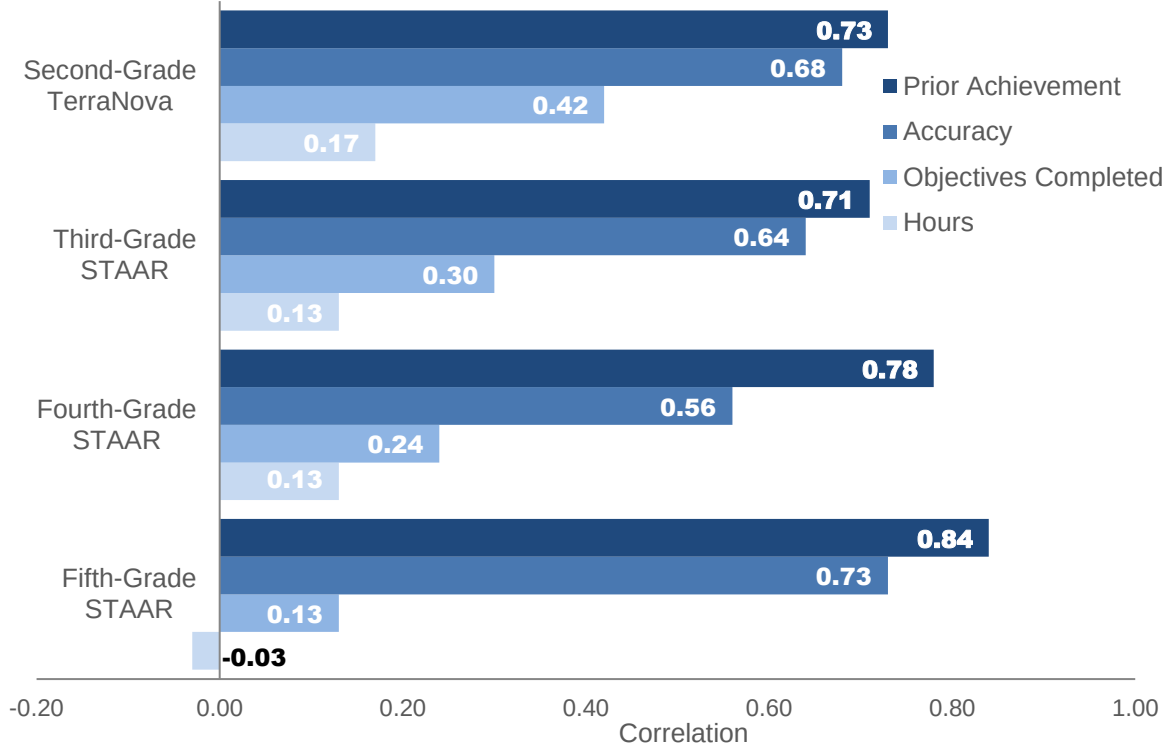
Source: PEIMS district data (10/30/2017). RM data (1/04/2018 and 6/05/2018). STAAR (6/20/2018).

Note: STAAR=State of Texas Assessments of Academic Readiness. SD=standard deviation. Eco. Dis.=economically disadvantaged. Spe. Ed.=Special Education. % Appr.+ = % of students who achieved Approaches Grade Level or Above. % Meets+=% of students who achieved Meets Grade Level or Above. Results included STAAR first administration data. STAAR Alternate 2 scores were excluded from the analyses for both RM students and district students. Scale scores were used for STAAR descriptive statistics. Although all STAAR scores were included, some students' test scores might have been adversely affected by the online testing issues that occurred in April and May 2018. Cells marked with an asterisk (*) indicate five or fewer students were tested.

Correlations between RM Components and Spring Assessment Scores

Like past years, students' mathematics scores were more highly correlated with prior achievement ($r=0.71$ to 0.84) and accuracy ($r=0.56$ to 0.73) than with objectives completed ($r=0.13$ to 0.42) or hours online ($r=-0.03$ to 0.17 - see Figure 19).²⁸

Figure 19: Correlations between RM Indicators, Prior Achievement, and Spring 2018 Mathematics Assessment Scores by Grade



Source: PEIMS district data (10/30/2017). RM data (1/04/2018 and 6/05/2018). TerraNova (7/19/2017 and 6/11/2018). STAAR (6/15/2017 and 6/20/2018).

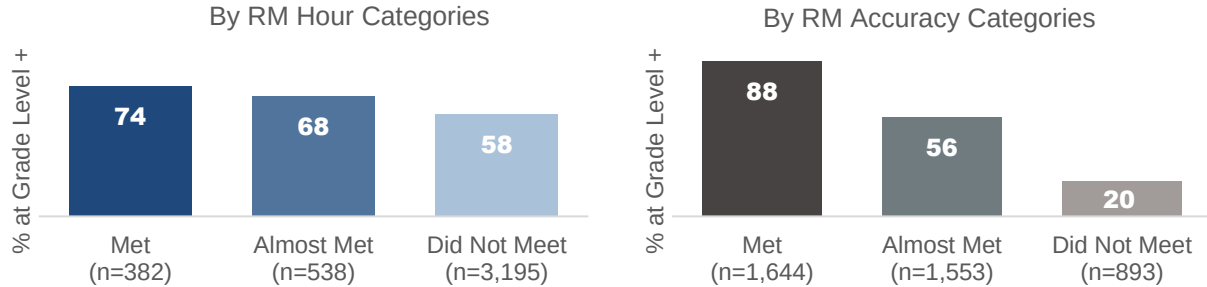
Note: STAAR=State of Texas Assessments of Academic Readiness. Accuracy= RM Guided Study Level A Accuracy.

Comparisons of Mathematics Achievement by RM Hour and Accuracy Target Categories

Across grade levels and assessments, RM students that “met” the accuracy target noticeably outperformed students that “almost met” or “did not meet” the target (see Figures 20 to 23 and Appendix L). Students that “met” the hour target tended to perform slightly better than students in the “almost met” and “did not meet” categories with the exception of third grade; however, the differences were not as pronounced as those for accuracy. Results were consistent with 2016-17 evaluation findings (Kim & Bush, 2017).

²⁸ n for second grade=4,115, 4,029, 4,090, and 3,673 for hours, objectives completed, accuracy, and prior achievement, respectively; n for third grade=4,269, 4,159, 4,231, and 3,835, respectively; n for fourth grade=2,284, 2,064, 2,082 and 2,105, respectively; n for fifth grade=111, 111, 111, and 104, respectively. For the correlation coefficient (r), absolute values of .1, .3, and .5 are considered to be small, medium, and large, respectively (Cohen, 1988).

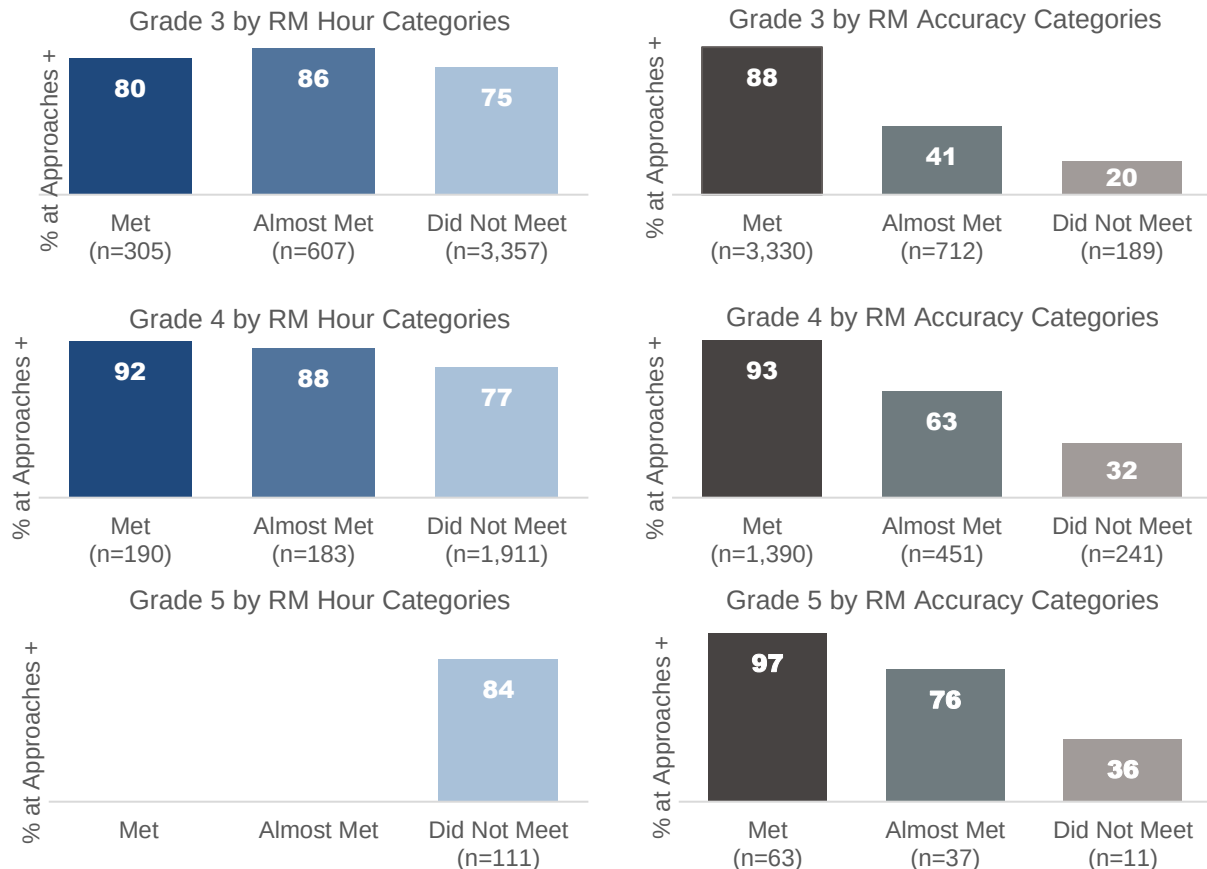
Figure 20: Percentage of Students Performing At or Above Grade Level on Spring 2018 *TerraNova* by RM Hour^a and Accuracy^b Categories



Source: PEIMS district data (10/30/2017). RM data (1/04/2018 and 6/05/2018). *TerraNova* (6/11/2018).

Note: % at Grade Level+ = % At or Above Grade Level. Scores were considered at or above grade level if they were at or above the 40th percentile. ^a Met=Logged 100% of expected hours. Almost Met=Logged 80 to 99% of expected hours. Did Not Meet=Logged less than 80% of expected hours. ^b Met=Attained Guided Study Level A accuracy rate of 75% or higher. Almost Met=Accuracy rate between 60 and 74%. Did Not Meet=Accuracy rate of less than 60%.

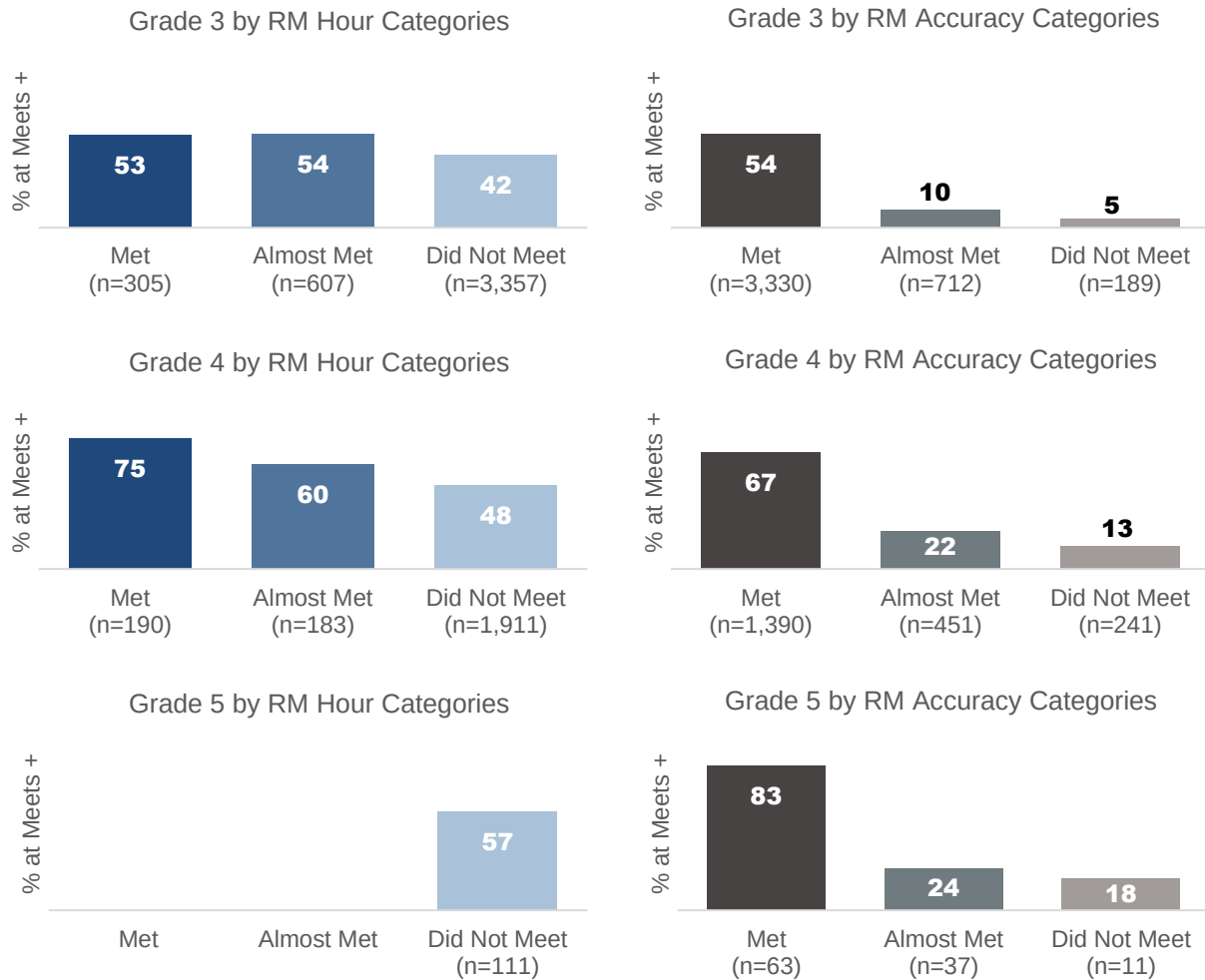
Figure 21: Percentage of Students in Grades 3 to 5 Who Achieved Approaches Grade Level or Above (Approaches +) on Spring 2018 *STAAR* Mathematics by RM Hour^a and Accuracy^b Categories



Source: PEIMS district data (10/30/2017). RM data (1/04/2018 and 6/05/2018). *STAAR* (6/20/2018).

Note: *STAAR*=State of Texas Assessments of Academic Readiness. Results included *STAAR* first administration data. *STAAR Alternate 2* scores were excluded from the analyses. Although all *STAAR* scores were included, some students' test scores might have been adversely affected by the online testing issues that occurred in April and May 2018. % at Approaches+=% at Approaches Grade Level or Above. ^aMet=Logged 100% of expected hours. Almost Met=Logged 80 to 99% of expected hours. Did Not Meet=Logged less than 80% of expected hours. ^bMet=Attained Guided Study Level A accuracy rate of 75% or higher. Almost Met=Accuracy rate between 60 and 74%. Did Not Meet=Accuracy rate of less than 60%.

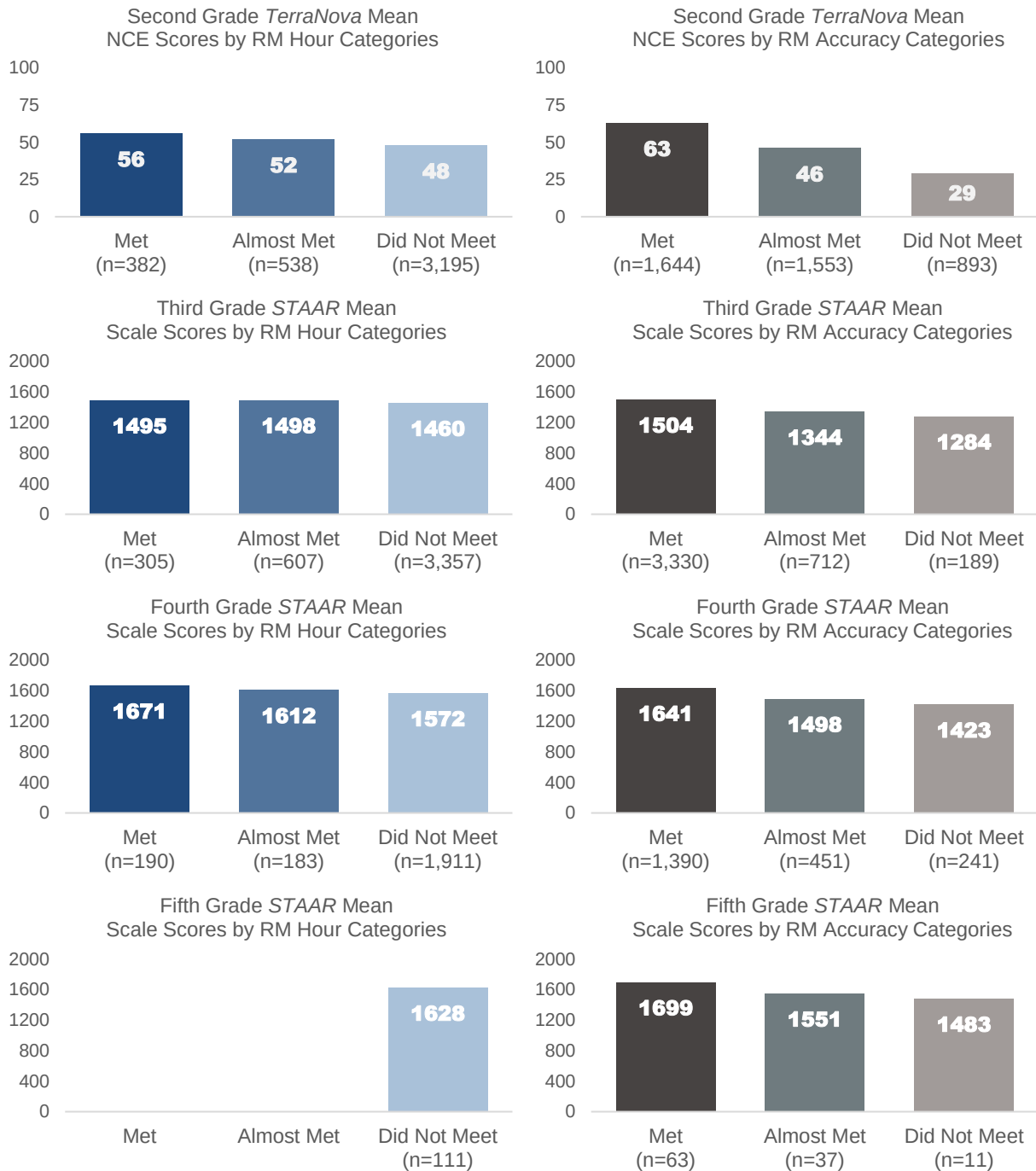
Figure 22: Percentage of Students in Grades 3 to 5 Who Achieved Meets Grade Level or Above (Meets +) on Spring 2018 STAAR Mathematics by RM Hour^a and Accuracy^b Categories



Source: PEIMS district data (10/30/2017). RM data (1/04/2018 and 6/05/2018). STAAR (6/20/2018).

Note: STAAR=State of Texas Assessments of Academic Readiness. Results included STAAR first administration data. STAAR Alternate 2 scores were excluded from the analyses. Although all STAAR scores were included, some students' test scores might have been adversely affected by the online testing issues that occurred in April and May 2018. % at Meets+ = % at Meets Grade Level or Above. ^a Met=Logged 100 percent of expected hours. Almost Met=Logged 80 to 99 percent of expected hours. Did Not Meet=Logged less than 80 percent of expected hours. ^b Met=Attained Guided Study Level A accuracy rate of 75 percent or higher. Almost Met=Accuracy rate between 60 and 74 percent. Did Not Meet=Accuracy rate of less than 60 percent.

Figure 23: Spring 2018 Assessment Mean Scores by RM Hour^a and Accuracy^b Categories



Source: PEIMS district data (10/30/2017). RM data (1/04/2018 and 6/05/2018). TerraNova (6/11/2018). STAAR (6/20/2018).

Note: STAAR=State of Texas Assessments of Academic Readiness. For STAAR, results included STAAR first administration data. STAAR Alternate 2 scores were excluded from the analyses. Although all STAAR scores were included, some students' test scores might have been adversely affected by the online testing issues that occurred in April and May 2018. NCE scores=Normal Curve Equivalent (NCE) scores. ^aMet=Logged 100 percent of expected hours. Almost Met=Logged 80 to 99 percent of expected hours. Did Not Meet=Logged less than 80 percent of expected hours. ^bMet=Attained Guided Study Level A accuracy rate of 75 percent or higher. Almost Met=Accuracy rate between 60 and 74 percent. Did Not Meet=Accuracy rate of less than 60 percent.

Multiple Regression Results

Predictors of Spring Assessment Scores. Multiple regression results showed that with the exception of fifth grade, prior achievement, RM accuracy, and RM objectives completed were major predictors of students' *TerraNova/STAAR* scores (see Table 5 and Appendix M). Grade-level regression results follow.

- For second grade, there were three major predictors: prior achievement ($\beta=.41$, $p<.001$), accuracy ($\beta=.40$, $p<.001$), and objectives completed ($\beta=.19$, $p<.001$). When all variables were accounted for in the model, gender was a statistically significant predictor.²⁹ The eight predictors explained about 65 percent of the variance of the *TerraNova* scores [$R^2=.651$, $F(8, 3,580)=836.01$, $p<.001$].
- Third-grade results showed three major predictors: prior achievement ($\beta=.48$, $p<.001$), accuracy ($\beta=.31$, $p<.001$), and number of objectives completed ($\beta=.09$, $p<.001$). When all variables were accounted for, ethnicity was a statistically significant predictor.³⁰ The eight predictors explained about 59 percent of the variance of the *STAAR* scores [$R^2=.589$, $F(8, 3,728)=666.61$, $p<.001$].
- Fourth-grade regression analyses also revealed prior achievement, accuracy, and number of objectives completed as major predictors ($\beta=.64$, $p<.001$; $\beta=.18$, $p<.001$; and $\beta=.11$, $p<.001$; respectively). When all variables were accounted for, economically disadvantaged status and ethnicity were statistically significant predictors.³¹ The eight predictors explained about 64 percent of the variance of the *STAAR* scores [$R^2=.640$, $F(8, 1,808)=401.71$, $p<.001$].
- Unlike other grade levels, the number of objectives completed variable was not a significant predictor for fifth grade. The regression analyses indicated prior achievement and accuracy as major predictors ($\beta=.64$, $p<.001$ and $\beta=.27$, $p<.001$, respectively). When all variables were accounted for, there were no statistically significant differences in *STAAR* scores between male and female students, or between economically and non-economically disadvantaged students. Ethnic groups were not included in the fifth-grade regression analyses due to small number of students in White and other ethnic groups. The five predictors explained about 73 percent of the variance of the *STAAR* scores [$R^2=.731$, $F(5, 98)=53.19$, $p<.001$].

²⁹ Male students tended to have higher scores than female students. There were no statistically significant differences in *TerraNova* scores between economically and non-economically disadvantaged students or among ethnic group students.

³⁰ White students tended to have higher *STAAR* scores than African American students. There were no statistically significant differences between male and female students or between economically and non-economically disadvantaged students or between White and Hispanic students or between White and other ethnic group students.

³¹ Students who were *not* economically disadvantaged tended to have higher *STAAR* scores than economically disadvantaged students, and White students tended to have higher *STAAR* scores than African American students. There were no statistically significant differences between male and female students or between White and Hispanic students or between White and other ethnic group students.

Table 5: Multiple Regression Major RM Predictors of Spring 2018 Mathematics Scores

Grade	Major Predictors	β	p	Variance Explained
2 (n=3,589)	Prior Achievement	0.41	<.001	65%
	Accuracy	0.40	<.001	
	Objectives Completed	0.19	<.001	
3 (n=3,737)	Prior Achievement	0.48	<.001	59%
	Accuracy	0.31	<.001	
	Objectives Completed	0.09	<.001	
4 (n=1,817)	Prior Achievement	0.64	<.001	64%
	Accuracy	0.18	<.001	
	Objectives Completed	0.11	<.001	
5 (n=104)	Prior Achievement	0.64	<.001	73%
	Accuracy	0.27	<.001	

Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018). TerraNova (7/19/2017 and 6/11/2018). STAAR (6/15/2017 and 6/20/2018).

Note: TerraNova Normal Curve Equivalent (NCE) scores were used for second grade. STAAR scale scores were used for grades three to five.

Comparisons of Mathematics Achievement between RM and Non-RM Students

Multiple Regression Results for Grades 2 to 5. As shown in Table 6, multiple regression results showed that when prior achievement and economically disadvantaged status were accounted for, differences in STAAR 2018 mathematics scale scores between RM and non-RM comparison students were statistically significant (favoring RM students) for third and fourth grades ($\beta=0.02$, $p<.01$ and $\beta=0.04$, $p<.001$, respectively)³² but showed no practical significance for either grade ($f=0.02$ and $f=0.04$, respectively).³³

As for second and fifth grades, there were no statistically or practically significant differences in Spring 2018 mathematics scores between RM and non-RM comparison students ($\beta=0.01$, $p>.05$ and $f=0.01$ for second grade; $\beta=-0.07$, $p>.05$ and $f=0.07$ for fifth grade).³⁴

³² $F(3, 7,666)=2,609.61$, $p<.001$ for third grade and $F(3, 4,206)=2,187.33$, $p<.001$ for fourth grade. For third grade, the average STAAR 2018 scale score was higher for RM students than for non-RM students (1471.31 versus 1463.96). Review of prior achievement (TerraNova 2017) showed that RM students and non-RM students had similar TerraNova 2017 NCE scores (49.11 and 48.91, respectively) but had slightly higher TerraNova 2017 NCE scores than the district as a whole (47.71). For fourth grade, the average STAAR 2018 scale score was higher for RM students than for non-RM students (1585.93 versus 1579.62). Review of prior achievement showed that non-RM students had slightly higher STAAR 2017 mean scale scores than RM students (1479.13 versus 1472.37) and the district as a whole (1446.20).

³³ Cohen's f values of .10, .25, and .40 are considered small, medium, and large, respectively (Cohen, 1988).

³⁴ $F(3, 7,338)=2,736.65$, $p<.001$ for second grade and $F(3, 204)=78.15$, $p<.001$ for fifth grade. For second grade, the average TerraNova NCE scores between RM and non-RM students were about the same (49.71 and 49.24, respectively). Review of prior achievement (TerraNova 2017) showed that RM students and non-RM students had similar TerraNova 2017 NCE scores (47.50 and 47.40, respectively) but had slightly higher TerraNova 2017 NCE scores than the district as a whole (46.87). For fifth grade, the average STAAR 2018 scale scores was higher for non-RM students than for RM students (1663.23 versus 1631.26). Per review of prior achievement, non-RM students also had higher STAAR 2017 mean scale scores than RM students (1611.09 versus 1593.04) and the district as a whole (1548.20).

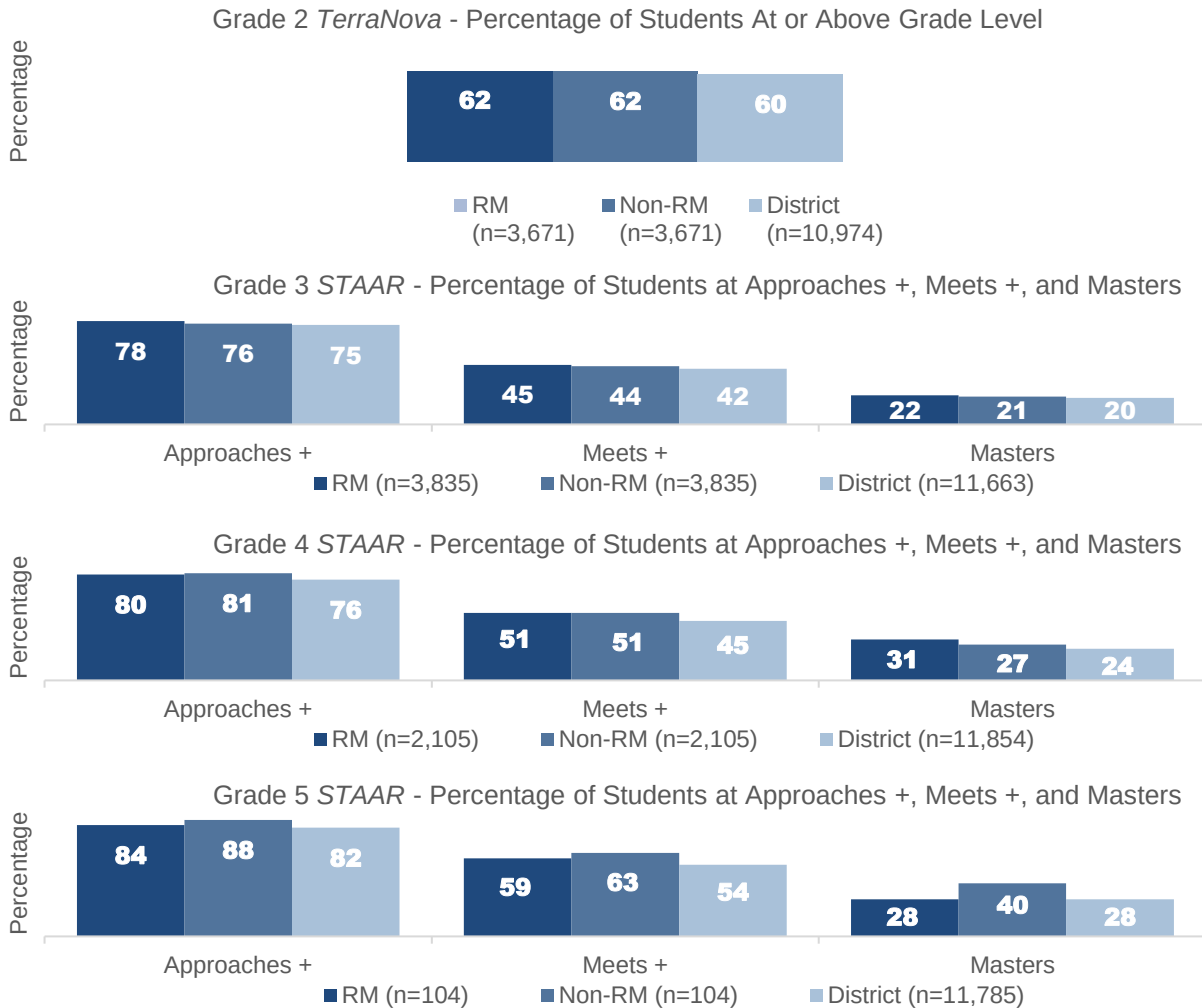
Table 6: Multiple Regression Results for RM and Non-RM Student Comparisons

	Correlations among Variables				β	t-value
	1	2	3	4		
Second Grade ($R^2=0.528$; Adjusted $R^2=0.528$)						
1. TerraNova 2018 NCE Scores	1.00					
2. Prior Year TerraNova NCE Scores	0.73	1.00			0.73	89.60***
3. Economically Disadvantaged	-0.11	-0.14	1.00		-0.01	-1.04
4. RM vs. Non-RM Students	0.01	0.00	-0.01	1.00	0.01	1.20
Mean	49.48	47.45	0.80	0.50		
Standard Deviation	20.14	19.11	0.40	0.50		
Third Grade ($R^2=0.505$; Adjusted $R^2=0.505$)						
1. STAAR 2018 Scale Scores	1.00					
2. Prior Year TerraNova NCE Scores	0.71	1.00			0.71	87.58***
3. Economically Disadvantaged	-0.10	-0.11	1.00		-0.02	-2.52*
4. RM vs. Non-RM Students	0.03	0.01	-0.02	1.00	0.02	2.66**
Mean	1467.63	49.01	0.88	0.50		
Standard Deviation	144.84	19.21	0.32	0.50		
Fourth Grade ($R^2=0.609$; Adjusted $R^2=0.609$)						
1. STAAR 2018 Scale Scores	1.00					
2. Prior Year STAAR Scale Scores	0.78	1.00			0.78	80.40***
3. Economically Disadvantaged	-0.09	-0.11	1.00		-0.01	-0.67
4. RM vs. Non-RM Students	0.02	-0.02	-0.01	1.00	0.04	3.97***
Mean	1582.77	1475.75	0.91	0.50		
Standard Deviation	147.37	155.79	0.29	0.50		
Fifth Grade ($R^2=0.535$; Adjusted $R^2=0.528$)						
1. STAAR 2018 Scale Scores	1.00					
2. Prior Year STAAR Scale Scores	0.73	1.00			0.73	14.98***
3. Economically Disadvantaged	0.10	0.16	1.00		-0.01	-0.26
4. RM vs. Non-RM Students	-0.12	-0.06	-0.03	1.00	-0.07	-1.50
Mean	1647.25	1602.06	0.98	0.50		
Standard Deviation	139.50	150.57	0.15	0.50		

Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018). TerraNova (7/19/2017 and 6/11/2018). STAAR (6/15/2017 and 6/20/2018).

Note: STAAR=State of Texas Assessments of Academic Readiness. $n=7,342$, 7,670, 4,210, 208 for grade two, three, four and five, respectively. * $p<.05$. ** $p<.01$. *** $p<.001$. TerraNova Normal Curve Equivalent (NCE) scores were used for second grade. STAAR scale scores were used for grades three to five. For STAAR, results included first administration data. STAAR Alternate 2 scores were excluded from the analyses. Although all STAAR scores were included, some students' test scores might have been adversely affected by the online testing issues that occurred in April and May 2018.

RM and Non-RM Student Comparisons on TerraNova 2018 and STAAR 2018. As shown in Figure 24, there was no clear pattern of differences in student achievement on TerraNova and STAAR between RM and non-RM comparison students. District information was provided as a reference. For second grade, the percentage of students at or above grade level on TerraNova was the same between RM and non-RM students (62% for both groups). For third and fourth grades, results showed mixed pattern of differences in terms of percentage of students who achieved Approaches +, Meets +, and Masters between RM and non-RM students. For fifth grade, higher percentages of non-RM students than RM students achieved Approaches +, Meets +, and Masters. In general, both RM and non-RM comparison students performed better than the district as a whole.

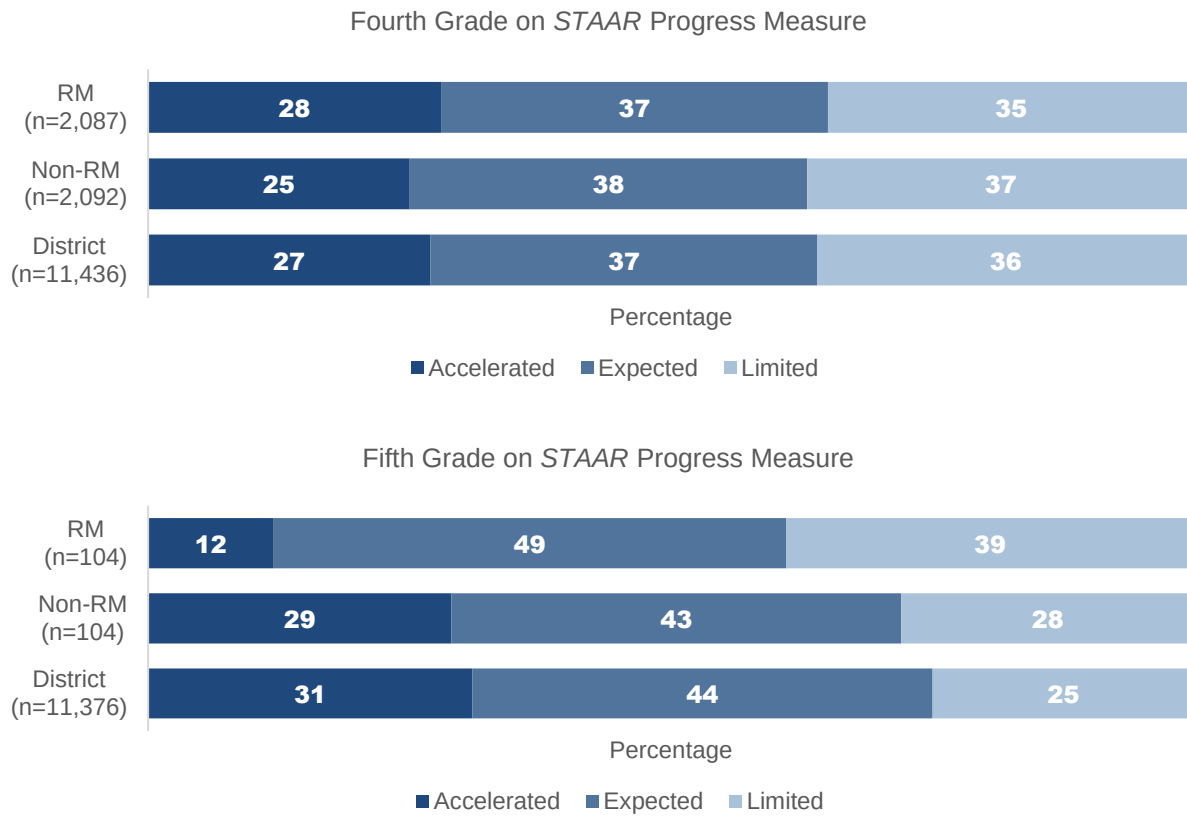
Figure 24: RM and Non-RM Student Comparisons on *TerraNova* 2018 and *STAAR* 2018

Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018). *TerraNova* (6/11/2018). *STAAR* (6/20/2018).

Note: *STAAR*=State of Texas Assessments of Academic Readiness. For *TerraNova*, National Percentile scores were used to determine students that met the grade-level standard (At or Above the 40th percentile). For *STAAR*, results included first administration data. *STAAR Alternate 2* scores were excluded from the analyses. Although all *STAAR* scores were included, some students' test scores might have been adversely affected by the online testing issues that occurred in April and May 2018. Approaches + = Approaches Grade Level or Above. Meets + = Meets Grade Level or Above. District information was provided as a reference.

RM and Non-RM Student Comparisons on *STAAR* Progress Measure. As shown in Figure 25, 65 percent of RM students in grade four made 'Expected' or 'Accelerated' progress in 2017-18, which was two percentage points higher than non-RM comparison students and one percentage point higher than the district as a whole. For fifth grade, 61 percent of RM students made 'Expected' or 'Accelerated' progress in 2017-18, which was 11 percentage points lower than non-RM comparison students and 14 percentage points lower than the district as a whole. However, the results for fifth grade should be reviewed with caution because there was only one centrally funded RM school for fifth grade and the sample sizes for RM and non-RM students were relatively small compared to the district.

Figure 25: RM and Non-RM Student Comparisons on Spring 2018 STAAR Progress Measure



Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018). STAAR (6/20/2018).

Note: STAAR=State of Texas Assessments of Academic Readiness. For STAAR, results included first administration data. STAAR Alternate 2 scores were excluded from the analyses. Although all STAAR scores were included, some students' test scores might have been adversely affected by the online testing issues that occurred in April and May 2018. District information was provided as a reference.

SUMMARY AND RECOMMENDATIONS

Summary

Reasoning Mind (RM) Context

A total of 11,311 students participated in RM during 2017-18. This included 4,346 (37%) of the district's second grade students, 4,464 (36%) of the district's third-grade students, 2,388 (19%) of the district's fourth-grade students, and 113 (1%) of the district's fifth-grade students. The percentage of district students enrolled in RM decreased from 2015-16 to 2016-17 to 2017-18 for all grades. The majority of RM students were economically disadvantaged (85%) and Hispanic (75%). Fifty-six percent were English learners (ELs).

RM was supplemental to the district curriculum in grades two to four, which means teacher time was split between using district and RM curricula. The RM curriculum served as the full core curriculum for grade five. In 2017-18, the district used Title I funds and general operating funds to support a total of 57

elementary schools. These 57 campuses were required to commit and adhere to the implementation requirements (i.e., meeting hour/accuracy targets). The 2017-18 RM budget was \$847,763, which included \$838,962 for the grades two to four supplemental RM program (Title I, Part A funds) and \$8,801 for the grade five core RM program (general operating funds). The total budget amount (\$847,763) was decreased by about half from 2016-17 (\$1,763,660).

RM Implementation

Student Hours and Accuracy. Implementation fidelity continued to be a concern as the majority of campuses were classified as low implementing for second (87%), third (94%), fourth (88%), and fifth (100%) grades. The continued low level of implementation was surprising since most 2017-18 RM campuses had a medium or high level of implementation in 2016-17. Consistent with past years, Dallas ISD fell short of meeting implementation expectations. Most RM students (91% to 100%) did *not* reach the district's 2017-18 target for hours online (target = 62 hours for grades two to four and 174 hours for grade five), and a sizeable proportion (79% to 100%) logged less than 80 percent of expected hours. Whereas many third- (78%), fourth- (66%), and fifth-grade (57%) students met the district accuracy target, many second-grade students (60%) did *not* meet the accuracy target.

Professional Development. In 2017-18, first-year RM teachers were not required but were strongly encouraged to take the QC before using RM with students due to delayed processing of the purchase orders for RM in Fall 2017. For this reason, only 36 percent of first-year RM teachers completed the QC by the end of the first six weeks, which was a notable decrease from last year (71% in Fall 2016). Half (50%) of the first-year RM teachers did not complete the QC by the end of the fall semester. About eight percent ($n=28$) of the district's 344 RM teachers participated in optional RM PD during the fall semester.

Staff Perceptions of RM. Per interviews with RM and district staff, the RM project director and the district mathematics director viewed more communication and support by RM ICs as a main success. In 2017-18, RM ICs held two group sessions at each campus based on need and held two individual meetings with each RM teacher. The RM project director also mentioned the single objective curriculum (SOC) as a success. SOC allows teachers to customize the RM curriculum to the district curriculum. Barriers to implementation included technology problems; delayed processing of purchase orders; the district's change in its data management tool to CLEVER; the reduced budget; and fewer ICs.

As in previous years, campus staff survey results revealed that teachers were more positive than campus administrators on the overall survey items with one exception (the benefit of RM on students: 91% for campus administrators and 88% for teachers). Whereas 69 percent to 91 percent of administrators gave positive responses to overall items, 77 percent to 88 percent of teachers did so. Most teachers (80% to 89%) believed students improved on key indicators (e.g., enjoyment of mathematics, confidence in mathematics ability, etc.). The majority of campus administrators and teachers were "satisfied" or "extremely satisfied" with support from the RM implementation coordinator (85% for both groups) or support from district staff members (77% to 88%).

Some campus administrators and teachers reported technology issues, especially with laptops/computers on wheels (COW) carts (38% for campus administrators and 34% for teachers), network (22% for both), and wireless connectivity (21% and 28%, respectively). Between six percent and 30 percent of campus staff members reported other technology issues (6% to 30% for campus administrators and 8% to 18% for teachers). Of the teacher respondents, between 10 percent and 29 percent indicated that they used other mathematics intervention programs (e.g., TTM, HMH Personal Math Trainer, etc.) in addition to RM.

In terms of RM successes, teachers (19% to 74%) were more positive than campus administrators (22% to 58%) with two exceptions. Campus administrators were more positive than teachers on RM's *State of Texas Assessments of Academic Readiness (STAAR)* modules (58% and 43%, respectively) and RM's SOC (22% and 19%, respectively). Twenty-nine percent of campus administrators and teachers indicated the increased RM accuracy as a success. The most cited challenges were technology-related issues (54% for campus administrators and 47% for teachers) and difficulty meeting RM time requirements (49% and 38%, respectively).

Mathematics Achievement of RM Students

Descriptive Statistics for RM Student Spring Mathematics Performance. Overall mean assessment scores for RM students were close to the district mean for grade two, slightly higher for grades three and five, and much higher for grade four (mean differences were 0.8, 7.6, 2.8, and 23.6, respectively). Although the mean difference for grade four was quite large, it should be noted that RM students also had higher prior year (2017) *STAAR* mean scale scores than the district as a whole (mean difference was 26.2). When reviewed by subgroup, mean scale scores for African American students were lower than other ethnic groups, and mean scale scores for White students were higher than other groups.

Comparisons of Assessment Scores and Hour/Accuracy Target Categories. Like past years, RM students that met the accuracy target outperformed those who did not. Students that “met” the hour target tended to perform slightly better than students in the “almost met” and “did not meet” categories with the exception of third grade; however, the differences were not as pronounced as those for accuracy.

Correlation and Multiple Regression Results. Correlation and multiple regression results were consistent with 2016-17 evaluation findings (Kim & Bush, 2017). Students' mathematics scores were more highly correlated with prior achievement ($r=0.71$ to 0.84) and accuracy ($r=0.56$ to 0.73) than with objectives completed ($r=0.13$ to 0.42) or hours online ($r=-0.03$ to 0.17). Multiple regression analyses showed that with the exception of fifth grade, prior achievement, RM accuracy, and RM objectives completed were major predictors of Spring 2018 assessment scores.

Assessment Comparisons between RM and Non-RM Students. When student performance on *TerraNova* 2018 and *STAAR* 2018 was compared between RM and non-RM comparison students, results showed no clear pattern of differences in terms of percentage of students that met the performance standards between the two groups.

For third and fourth grades, multiple regression results showed that when prior achievement and economically disadvantaged status were accounted for, differences in STAAR 2018 mathematics scale scores between RM and non-RM comparison students were statistically significant (favoring RM students) but showed no practical significance for either grade ($f=0.02$ and $f=0.04$, respectively). As for second and fifth grades, there were no statistically or practically significant differences in Spring 2018 mathematics scores between RM and non-RM comparison students.

In regard to RM and non-RM student comparisons on STAAR progress measures, 65 percent of RM students in grade four made 'Expected' or 'Accelerated' progress in 2017-18, which is two percentage points higher than non-RM comparison students. For fifth grade, 61 percent of RM students made 'Expected' or 'Accelerated' progress in 2017-18, which is 11 percentage points lower than non-RM comparison students.

Recommendations

- **Require higher levels of implementation for 2018-19 RM campuses.** Across grade levels, there were more campuses in the “low” implementing category than in the “medium” and “high” implementing categories. Strong implementation of RM is important to effectively impact student achievement and to ensure the district receives an adequate return on its investment of \$847,763. As a result, the district should require higher levels of RM implementation for all RM campuses in 2018-19.
- **Make sure teachers receive adequate guidance on how to meld two different curricula (district and RM) using SOC.** In an effort to alleviate the lack of alignment between district and RM curricula, RM provided SOC PD for teachers. SOC allows teachers to customize the RM curriculum to the district curriculum. According to the teacher survey, 73 percent of the teachers who responded on the SOC-related questions ($n=196$) indicated that they had received the training on RM's SOC. It is important to make certain that every teacher recognizes the availability of SOC and has sufficient guidance on how to customize the RM curriculum to the district curriculum as needed.
- **Make certain all campuses have adequate working technology and are reporting issues to Information Technology as problems occur.** Technology issues continued to be a challenge based on the surveys and interviews. It is important to continue to add and upgrade campus technology while ensuring ample campus support and timely submission of tickets to Information Technology.
- **Strategically plan for next year's RM QC training to make sure more teachers complete the QC prior to or at the onset of the 2018-19 school year.** A notable percentage of first-year RM teachers (64%) did not complete the QC training by the end of the first six weeks of Fall 2017. As recommended in the past, the district needs to make sure more teachers complete the QC prior to or at the very beginning of the school year.
- **Work with RM to include quarter or semester level reports on the RM dashboard that show summary data on metrics (e.g., mean hours, mean accuracy, percentages met hour/accuracy targets, etc.) for all campuses.** Although the RM dashboard provides data on a variety of metrics including student online hours and Level A accuracy, it does not provide the percentage of students

that met the hour target. Also, the dashboard provides data at the campus, feeder pattern, and district levels; however, comprehensive summary data for all campuses are not available from the current dashboard. To help administrators monitor RM implementation overall and by school and intervene in a timely manner, it is important that quarter or semester level summary data for all campuses are included on the dashboard.

- **Decide whether to continue using the fifth-grade RM curriculum when it serves as the full curriculum but does not align with the district curriculum and when the campus for fifth grade met the evaluator criteria for “low” implementation.** Implementation and outcome measures give reasons for concern as all of fifth-grade students (100%) “did not meet” hour expectations (logged less than 80% of expected hours) and there was no significant difference between RM and non-RM comparison students on Spring 2018 STAAR; the same was true for Spring 2017 STAAR results.

REFERENCES

- Bobko, P. (2001). *Correlation and regression: Applications for industrial organizational psychology and management* (2nd ed.). Thousand Oaks, CA: Sage.
- Bush, J. & Kim, M. (2014). Evaluation of Reasoning Mind: Final Report 2013-14. Dallas, TX: Dallas Independent School District.
- Bush, J. & Kim, M. (2015). Evaluation of Reasoning Mind: Final Report 2014-15. Dallas, TX: Dallas Independent School District.
- Bush, J. & Kim, M. (2016). Evaluation of Reasoning Mind: Final Report 2015-16. Dallas, TX: Dallas Independent School District.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Kim, M. & Bush, J. (2017). 2016-17 Evaluation of Reasoning Mind. Dallas, TX: Dallas Independent School District.
- McEnturff, A. (2014). Evaluation of the 2013-2014 Home Instruction for Parents of Preschool Youngsters (HIPPY) Program. Dallas, TX: Dallas Independent School District.
- Reasoning Mind website. www.reasoningmind.org. Accessed March 8, 2018.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25 (1), 1-21.
- Thoemmes, F. J. (2012). Propensity score matching in SPSS. <https://arxiv.org/ftp/arxiv/papers/1201/1201.6385.pdf>. Accessed April 4, 2018.
- Waxman, H. C., & Houston, W. R. (2012). Evaluation of the 2010-2011 Reasoning Mind Program in Beaumont ISD. Reasoning Mind, Inc.

APPENDICES

Appendix A

Appendix A: Demographic Characteristics of 2017-18 RM Students and District Students

Characteristic	Second Grade n (%)	Third Grade n (%)	Fourth Grade n (%)	Fifth Grade n (%)	All n (%)
RM Students					
Gender					
Male	2,226 (51)	2,295 (51)	1,225 (51)	65 (58)	5,811 (51)
Female	2,120 (49)	2,169 (49)	1,163 (49)	48 (42)	5,500 (49)
Ethnicity					
African American	784 (18)	699 (16)	391 (16)	16 (14)	1,890 (17)
Hispanic	3,226 (74)	3,449 (77)	1,762 (74)	92 (81)	8,529 (75)
White	191 (5)	196 (4)	135 (6)	3 (3)	525 (5)
Other	145 (3)	120 (3)	100 (4)	2 (2)	367 (3)
Student Group					
Eco. Dis.	3,485 (80)	3,882 (87)	2,140 (90)	106 (94)	9,613 (85)
Gifted	602 (14)	679 (15)	452 (19)	14 (12)	1,747 (15)
Special Education	319 (7)	312 (7)	159 (7)	5 (4)	795 (7)
English Learner	2,454 (57)	2,463 (55)	1,331 (56)	56 (50)	6,304 (56)
Total	4,346 (100)	4,464 (100)	2,388 (100)	113 (100)	11,311 (100)
All District Students					
Gender					
Male	6,104 (51)	6,416 (51)	6,507 (51)	6,482 (52)	25,509 (51)
Female	5,790 (49)	6,149 (49)	6,193 (49)	6,026 (48)	24,158 (49)
Ethnicity					
African American	2,678 (22)	2,763 (22)	2,832 (22)	2,725 (22)	10,998 (22)
Hispanic	8,178 (69)	8,852 (70)	8,940 (70)	8,924 (71)	34,894 (70)
White	717 (6)	638 (5)	637 (5)	582 (5)	2,574 (5)
Other	321 (3)	312 (3)	291 (3)	277 (2)	1,201 (3)
Student Group					
Eco. Dis.	9,285 (78)	10,885 (87)	11,433 (90)	11,331 (91)	42,934 (86)
Gifted	1,724 (15)	1,967 (16)	2,464 (19)	2,522 (20)	8,677 (18)
Special Education	1,059 (9)	1,099 (9)	1,149 (9)	1,192 (10)	4,499 (9)
English Learner	5,985 (50)	6,153 (49)	6,257 (49)	6,212 (50)	24,607 (50)
Total	11,894 (100)	12,565 (100)	12,700 (100)	12,508 (100)	49,667 (100)

Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018).

Note: Eco. Dis.=Economically Disadvantaged. For RM students, analyses were limited to students in the RM database that were in Dallas ISD as of PEIMS snapshot date (10/30/2017) and with RM hours equal to or greater than 0.1 in 2017-18. Also, analyses were limited to students in RM schools that were funded centrally. "Other" ethnicity included Asian, American Indian, Native Hawaiian or Other Pacific Islander, two or more races, and not available.

Appendix B

Appendix B: 2017-18 RM Hour^a and Accuracy^b Statistics by Campus and Grade

High School Feeder	Campus	Grade	n	RM Hours for 2017-18		Accuracy Rates for 2017-18	
				Mean	% Met Target	Mean	% Met Target
All	All	2	4,346	32.44	9	69.79	40
All	All	3	4,464	33.91	7	81.81	78
All	All	4	2,388	25.94	8	77.46	66
All	All	5	113	76.82	0	74.86	57
ACE	Blanton	2	105	3.80	0	82.35	78
ACE	Blanton	3	107	23.48	0	81.52	77
ACE	Blanton	4	97	18.91	0	76.59	64
ACE	Mills	2	48	8.43	0	75.98	65
ACE	Mills	3	53	23.81	8	80.15	69
ACE	Mills	4	53	10.44	0	78.06	67
Adams, B	Bayles	2	77	21.87	0	69.15	38
Adams, B	Bayles	3	85	36.73	13	78.88	74
Adams, B	Bayles	4	86	19.30	0	73.21	55
Adams, B	Casa View	2	101	36.89	2	70.23	40
Adams, B	Casa View	3	106	42.33	18	84.23	82
Adams, B	Hexter	2	74	30.35	0	81.68	80
Adams, B	Hexter	3	92	32.04	1	86.52	90
Adams, B	Hexter	4	83	4.78	0	83.76	78
Adams, B	Kiest	2	94	31.99	0	67.81	34
Adams, B	Kiest	3	118	26.50	0	85.94	92
Adams, B	Kiest	4	82	3.04	0	90.26	89
Adams, B	Reilly	2	84	18.75	0	74.67	50
Adams, B	Reilly	3	85	14.00	0	85.32	88
Adams, B	Reinhardt	2	68	43.45	9	74.01	50
Adams, B	Sanger	2	68	19.24	0	74.22	56
Adams, B	Sanger	3	78	34.45	1	83.62	86
Adams, B	Sanger	4	79	6.22	0	81.27	80
Adams, B	Smith, L	2	116	35.66	4	70.24	39
Adams, B	Smith, L	3	118	45.31	8	81.00	73
Adams, B	Smith, L	4	109	59.93	55	78.18	70
Adams, B	Smith, L	5	113	76.82	0	74.86	58
Adamson	Reagan	2	47	52.22	11	70.93	41
Adamson	Reagan	3	47	33.68	0	83.84	83
Adamson	Reagan	4	52	13.25	0	77.25	67
Carter	Terry	2	63	28.10	2	78.60	62
Carter	Terry	3	54	35.44	0	79.44	76
Carter	Terry	4	64	28.78	0	77.87	67
Conrad	Lowe	2	83	19.77	0	69.31	41
Conrad	Lowe	3	74	39.89	9	76.02	60
Conrad	Lowe	4	87	20.27	0	68.23	41
Conrad	McShan	2	86	25.07	1	71.31	38
Conrad	McShan	3	92	20.71	0	82.49	78
Conrad	McShan	4	103	30.59	1	68.14	43
Hillcrest	Frank	2	183	43.57	4	73.58	48
Hillcrest	Frank	3	167	37.88	1	84.17	86
Hillcrest	Rogers	2	65	39.05	2	72.91	48
Hillcrest	Rogers	3	68	40.62	10	85.52	88
Hillcrest	Rogers	4	67	49.35	31	75.69	64
Intensive Support Network	Burleson	2	91	8.41	0	66.81	38
Intensive Support Network	Burleson	3	106	5.86	0	73.63	54
Intensive Support Network	Rhoads	2	120	32.80	10	59.84	12
Intensive Support Network	Rice	2	80	28.81	4	66.64	30
Intensive Support Network	Rice	3	57	1.42	0	81.31	67
Jefferson	Burnet	2	98	58.77	51	69.59	34
Jefferson	Burnet	3	95	64.67	68	80.74	72
Jefferson	Burnet	4	117	12.07	0	78.44	70
Jefferson	May	2	85	31.96	12	67.88	40
Jefferson	May	3	91	44.09	2	79.66	73
Jefferson	May	4	96	47.12	11	82.15	83

table continues

Table (continued)

High School Feeder	Campus	Grade	n	RM Hours for 2017-18		Accuracy Rates for 2017-18	
				Mean	% Met Target	Mean	% Met Target
Jefferson	Polk	2	44	46.76	7	61.33	25
Jefferson	Polk	3	43	34.62	0	75.86	51
Jefferson	Polk	4	66	30.18	0	72.48	56
Jefferson	Walnut Hill	2	47	38.66	9	74.35	47
Jefferson	Walnut Hill	3	46	48.60	15	84.62	89
Jefferson	Walnut Hill	4	38	18.84	0	82.17	84
Kimball	Henderson	2	58	60.64	67	67.94	29
Kimball	Henderson	3	75	63.56	53	85.47	87
Kimball	Henderson	4	67	62.41	64	79.94	75
Kimball	Stemmons	2	104	26.44	14	73.97	57
Kimball	Stemmons	3	129	35.20	2	83.61	85
Kimball	Stemmons	4	110	18.04	1	79.94	70
Molina	Bethune	2	107	60.44	60	66.81	36
Molina	Bethune	3	120	31.99	0	80.59	70
Molina	Bethune	4	18	0.36	0	72.50	60
Molina	Cowart	2	74	40.51	1	68.53	27
Molina	Cowart	3	90	23.28	0	80.75	71
North Dallas	Kennedy	2	42	38.62	10	66.33	26
North Dallas	Kennedy	3	54	40.19	2	82.98	85
North Dallas	Kennedy	4	53	36.82	0	74.29	64
North Dallas	Knight	2	60	42.55	8	73.86	50
North Dallas	Knight	3	85	40.14	21	86.28	93
North Dallas	Maple Lawn	2	68	29.77	1	65.28	26
North Dallas	Maple Lawn	3	74	36.19	11	81.11	73
North Dallas	Medrano, E	2	58	29.59	2	63.69	21
North Dallas	Medrano, E	3	61	33.98	2	82.26	77
North Dallas	Zaragoza	2	38	56.43	61	71.08	35
North Dallas	Zaragoza	3	51	43.29	6	78.67	75
North Dallas	Zaragoza	4	34	10.36	0	76.70	53
Pinkston	Martinez	2	78	6.74	0	65.16	28
Pinkston	Martinez	3	89	17.12	0	80.16	70
Pinkston	Martinez	4	74	11.45	0	79.86	78
Roosevelt	Miller	2	50	38.95	8	59.15	12
Roosevelt	Miller	3	58	43.29	21	78.44	69
Roosevelt	Miller	4	57	27.43	0	77.34	63
Roosevelt	Starks	2	57	25.56	0	62.27	30
Roosevelt	Starks	3	49	46.53	22	80.64	67
Roosevelt	Starks	4	57	51.60	46	80.33	75
Samuell	Pleasant Grove	2	65	26.70	0	66.77	34
Samuell	Pleasant Grove	3	93	25.21	0	77.47	65
Samuell	Pleasant Grove	4	77	26.76	0	71.36	51
Samuell	Runyon	2	86	55.64	28	67.75	38
Samuell	Runyon	3	122	40.41	11	77.99	66
Seagoville HS	Seagoville North	2	125	35.58	2	67.93	34
Seagoville HS	Seagoville North	3	124	37.58	1	80.79	81
Skyline	Rowe	2	58	29.60	0	72.18	47
Skyline	Rowe	3	65	30.40	0	80.76	83
Skyline	Rowe	4	76	28.83	1	71.11	38
Skyline	Urban Park	2	87	3.60	0	73.63	55
Skyline	Urban Park	3	94	8.13	0	81.93	79
Skyline	Urban Park	4	27	1.38	0	81.98	73
South Oak Cliff	Bushman	2	82	6.56	0	60.38	25
South Oak Cliff	Holland	2	60	28.47	18	71.90	43
South Oak Cliff	Holland	3	58	43.45	5	84.14	90
South Oak Cliff	Holland	4	62	24.36	0	77.61	62
South Oak Cliff	Jordan	2	90	37.09	11	62.18	22
South Oak Cliff	Jordan	3	79	28.66	0	75.49	61
South Oak Cliff	Oliver	2	48	18.13	0	63.63	23
South Oak Cliff	Oliver	3	45	37.32	13	74.24	53
South Oak Cliff	Oliver	4	47	44.84	34	71.63	49

table continues

Table (continued)

High School Feeder	Campus	Grade	n	RM Hours for 2017-18		Accuracy Rates for 2017-18	
				Mean	% Met Target	Mean	% Met Target
Spruce	Anderson, W	2	99	40.39	0	63.75	20
Spruce	Anderson, W	3	113	36.12	2	82.74	85
Spruce	Dorsey	2	81	20.20	0	70.05	36
Spruce	Dorsey	3	97	45.09	9	83.36	79
Spruce	Gonzalez	2	87	47.05	13	73.21	46
Spruce	Gonzalez	3	104	40.84	6	87.31	95
Spruce	Gonzalez	4	104	49.77	6	78.76	72
Spruce	Moseley	2	121	42.82	10	64.42	21
Spruce	Moseley	3	100	39.57	8	81.15	78
Sunset	Hooe	2	56	26.34	0	72.22	36
Sunset	Hooe	3	73	43.90	5	80.95	81
Sunset	Peabody	2	84	32.40	0	70.18	37
Sunset	Peabody	3	81	27.22	1	76.80	62
Sunset	Winnetka	2	116	48.95	12	73.13	47
Sunset	Winnetka	3	124	26.41	0	84.87	85
White	Adams, N	2	62	58.84	53	72.16	44
White	Adams, N	3	72	51.90	19	86.32	96
White	Degolyer	2	56	32.07	2	74.04	55
White	Degolyer	3	68	38.61	3	83.13	78
White	Degolyer	4	54	29.04	9	82.67	74
White	Junkins	2	94	13.47	0	74.61	55
White	Junkins	3	98	35.06	3	82.13	78
White	Junkins	4	89	18.57	0	81.25	76
White	Withers	2	68	14.80	0	75.65	57
White	Withers	3	58	22.47	3	85.50	83
White	Withers	4	56	5.11	0	82.29	71
Wilson	Jackson	3	26	1.11	0	90.71	86
Wilson	Lee, R	3	53	17.95	0	84.11	76
Wilson	Lee, R	4	47	8.65	0	87.19	86

Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018).

Note: ACE=Accelerating Campus Excellence. ^a% Met Hour Target=percentage of students who logged 100 percent of expected hours (Hour target=62 for grades two to four and 174 for grade five). ^b% Met Accuracy target=percentage of students who attained Guided Study Level A accuracy rate of 75 percent or higher.

Appendix C

Appendix C: 2017-18 Campuses with a High Level of RM Implementation

High School Feeder	Campus	<i>n</i>	Mean Hours	% Met Hour Target	Mean Accuracy	% Met Accuracy Target
Grade 3						
Jefferson	Burnet	95	64.67	68	80.74	72
Kimball	Henderson	75	63.56	53	85.47	87
Grade 4						
Adams, B	Smith, L	109	59.93	55	78.18	70
Kimball	Henderson	67	62.41	64	79.94	75

Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 06/05/2018 for Spring 2018).

Note: High implementing campuses for each grade were selected based on campus mean statistics, specifically based on the following criteria: (a) campus hour mean was at least 80 percent of the expected hour target, (b) 50 percent or more of students at the campus met the hour target, (c) campus accuracy mean was 60 percent or higher, and (d) 50 percent or more of students at the campus met the accuracy target. Hour target=62 for grades two to four and 174 for grade five. Accuracy (Guided Study Level A) target=75 percent or higher.

Appendix D

Appendix D: 2017-18 Campuses with a Medium Level of RM Implementation

High School Feeder	Campus	n	Mean Hours	% Met Hour Target	Mean Accuracy	% Met Accuracy Target
Grade 2						
Adamson	Reagan	47	52.22	11	70.93	41
Jefferson	Burnet	98	58.77	51	69.59	34
Kimball	Henderson	58	60.64	67	67.94	29
Molina	Bethune	107	60.44	60	66.81	36
North Dallas	Zaragoza	38	56.43	61	71.08	35
Samuell	Runyon	86	55.64	28	67.75	38
White	Adams, N	62	58.84	53	72.16	44
Grade 3						
White	Adams, N	72	51.90	19	86.32	96
Grade 4						
Roosevelt	Starks	57	51.60	46	80.33	75
Spruce	Gonzalez	104	49.77	6	78.76	72

Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018).

Note: The campuses that were not categorized as having either high or low level RM implementation were considered medium level RM implementation schools. High implementing campuses for each grade were selected based on campus mean statistics, specifically based on the following criteria: (a) campus hour mean was at least 80 percent of the expected hour target, (b) 50 percent or more of students at the campus met the hour target, (c) campus accuracy mean was 60 percent or higher, and (d) 50 percent or more of students at the campus met the accuracy target. Low implementing campuses for each grade were also selected based on campus mean statistics, specifically based on the following criteria: (a) campus hour mean was less than 80 percent of the expected hour target, and (b) 49 percent or fewer of students at the campus met the hour target. Hour target=62 for grades two to four and 174 for grade five. Accuracy (Guided Study Level A) target=75 percent or higher.

Appendix E

Appendix E: 2017-18 Campuses with a Low Level of RM Implementation

High School Feeder	Campus	n	Mean Hours	% Met Hour Target	Mean Accuracy	% Met Accuracy Target
Grade 2						
ACE	Blanton	105	3.80	0	82.35	78
ACE	Mills	48	8.43	0	75.98	65
Adams, B	Bayles	77	21.87	0	69.15	38
Adams, B	Casa View	101	36.89	2	70.23	40
Adams, B	Hexter	74	30.35	0	81.68	80
Adams, B	Kiest	94	31.99	0	67.81	34
Adams, B	Reilly	84	18.75	0	74.67	50
Adams, B	Reinhardt	68	43.45	9	74.01	50
Adams, B	Sanger	68	19.24	0	74.22	56
Adams, B	Smith, L	116	35.66	4	70.24	39
Carter	Terry	63	28.10	2	78.60	62
Conrad	Lowe	83	19.77	0	69.31	41
Conrad	McShan	86	25.07	1	71.31	38
Hillcrest	Frank	183	43.57	4	73.58	48
Hillcrest	Rogers	65	39.05	2	72.91	48
ISN	Burleson	91	8.41	0	66.81	38
ISN	Rhoads	120	32.80	10	59.84	12
ISN	Rice	80	28.81	4	66.64	30
Jefferson	May	85	31.96	12	67.88	40
Jefferson	Polk	44	46.76	7	61.33	25
Jefferson	Walnut Hill	47	38.66	9	74.35	47
Kimball	Stemmons	104	26.44	14	73.97	57
Molina	Cowart	74	40.51	1	68.53	27
North Dallas	Kennedy	42	38.62	10	66.33	26
North Dallas	Knight	60	42.55	8	73.86	50
North Dallas	Maple Lawn	68	29.77	1	65.28	26
North Dallas	Medrano, E	58	29.59	2	63.69	21
Pinkston	Martinez	78	6.74	0	65.16	28
Roosevelt	Miller	50	38.95	8	59.15	12
Roosevelt	Starks	57	25.56	0	62.27	30
Samuell	Pleasant Grove	65	26.70	0	66.77	34
Seagoville HS	Seagoville North	125	35.58	2	67.93	34
Skyline	Rowe	58	29.60	0	72.18	47
Skyline	Urban Park	87	3.60	0	73.63	55
South Oak Cliff	Bushman	82	6.56	0	60.38	25
South Oak Cliff	Holland	60	28.47	18	71.90	43
South Oak Cliff	Jordan	90	37.09	11	62.18	22
South Oak Cliff	Oliver	48	18.13	0	63.63	23
Spruce	Anderson, W	99	40.39	0	63.75	20
Spruce	Dorsey	81	20.20	0	70.05	36
Spruce	Gonzalez	87	47.05	13	73.21	46
Spruce	Moseley	121	42.82	10	64.42	21
Sunset	Hoe	56	26.34	0	72.22	36
Sunset	Peabody	84	32.40	0	70.18	37
Sunset	Winnetka	116	48.95	12	73.13	47
White	Degolyer	56	32.07	2	74.04	55
White	Junkins	94	13.47	0	74.61	55
White	Withers	68	14.80	0	75.65	57
Grade 3						
ACE	Blanton	107	23.48	0	81.52	77
ACE	Mills	53	23.81	8	80.15	69
Adams, B	Bayles	85	36.73	13	78.88	74
Adams, B	Casa View	106	42.33	18	84.23	82

table continues

Table (continued)

High School Feeder	Campus	n	Mean Hours	% Met Hour Target	Mean Accuracy	% Met Accuracy Target
Grade 3						
Adams, B	Hexter	92	32.04	1	86.52	90
Adams, B	Kiest	118	26.50	0	85.94	92
Adams, B	Reilly	85	14.00	0	85.32	88
Adams, B	Sanger	78	34.45	1	83.62	86
Adams, B	Smith, L	118	45.31	8	81.00	73
Adamson	Reagan	47	33.68	0	83.84	83
Carter	Terry	54	35.44	0	79.44	76
Conrad	Lowe	74	39.89	9	76.02	60
Conrad	McShan	92	20.71	0	82.49	78
Hillcrest	Frank	167	37.88	1	84.17	86
Hillcrest	Rogers	68	40.62	10	85.52	88
ISN	Burleson	106	5.86	0	73.63	54
ISN	Rice	57	1.42	0	81.31	67
Jefferson	May	91	44.09	2	79.66	73
Jefferson	Polk	43	34.62	0	75.86	51
Jefferson	Walnut Hill	46	48.60	15	84.62	89
Kimball	Stemmons	129	35.20	2	83.61	85
Molina	Bethune	120	31.99	0	80.59	70
Molina	Cowart	90	23.28	0	80.75	71
North Dallas	Kennedy	54	40.19	2	82.98	85
North Dallas	Knight	85	40.14	21	86.28	93
North Dallas	Maple Lawn	74	36.19	11	81.11	73
North Dallas	Medrano, E	61	33.98	2	82.26	77
North Dallas	Zaragoza	51	43.29	6	78.67	75
Pinkston	Martinez	89	17.12	0	80.16	70
Roosevelt	Miller	58	43.29	21	78.44	69
Roosevelt	Starks	49	46.53	22	80.64	67
Samuell	Pleasant Grove	93	25.21	0	77.47	65
Samuell	Runyon	122	40.41	11	77.99	66
Seagoville HS	Seagoville North	124	37.58	1	80.79	81
Skyline	Rowe	65	30.40	0	80.76	83
Skyline	Urban Park	94	8.13	0	81.93	79
South Oak Cliff	Holland	58	43.45	5	84.14	90
South Oak Cliff	Jordan	79	28.66	0	75.49	61
South Oak Cliff	Oliver	45	37.32	13	74.24	53
Spruce	Anderson, W	113	36.12	2	82.74	85
Spruce	Dorsey	97	45.09	9	83.36	79
Spruce	Gonzalez	104	40.84	6	87.31	95
Spruce	Moseley	100	39.57	8	81.15	78
Sunset	Hoe	73	43.90	5	80.95	81
Sunset	Peabody	81	27.22	1	76.80	62
Sunset	Winnetka	124	26.41	0	84.87	85
White	Degolyer	68	38.61	3	83.13	78
White	Junkins	98	35.06	3	82.13	78
White	Withers	58	22.47	3	85.50	83
Wilson	Jackson	26	1.11	0	90.71	86
Wilson	Lee, R	53	17.95	0	84.11	76
Grade 4						
ACE	Blanton	97	18.91	0	76.59	64
ACE	Mills	53	10.44	0	78.06	67
Adams, B	Bayles	86	19.30	0	73.21	55
Adams, B	Hexter	83	4.78	0	83.76	78
Adams, B	Kiest	82	3.04	0	90.26	89
Adams, B	Sanger	79	6.22	0	81.27	80

table continues

Table (continued)

High School Feeder	Campus	n	Mean Hours	% Met Hour Target	Mean Accuracy	% Met Accuracy Target
Grade 4						
Adamson	Reagan	52	13.25	0	77.25	67
Carter	Terry	64	28.78	0	77.87	67
Conrad	Lowe	87	20.27	0	68.23	41
Conrad	McShan	103	30.59	1	68.14	43
Hillcrest	Rogers	67	49.35	31	75.69	64
Jefferson	Burnet	117	12.07	0	78.44	70
Jefferson	May	96	47.12	11	82.15	83
Jefferson	Polk	66	30.18	0	72.48	56
Jefferson	Walnut Hill	38	18.84	0	82.17	84
Kimball	Stemmons	110	18.04	1	79.94	70
Molina	Bethune	18	0.36	0	72.50	60
North Dallas	Kennedy	53	36.82	0	74.29	64
North Dallas	Zaragoza	34	10.36	0	76.70	53
Pinkston	Martinez	74	11.45	0	79.86	78
Roosevelt	Miller	57	27.43	0	77.34	63
Samuell	Pleasant Grove	77	26.76	0	71.36	51
Skyline	Rowe	76	28.83	1	71.11	38
Skyline	Urban Park	27	1.38	0	81.98	73
South Oak Cliff	Holland	62	24.36	0	77.61	62
South Oak Cliff	Oliver	47	44.84	34	71.63	49
White	Degolyer	54	29.04	9	82.67	74
White	Junkins	89	18.57	0	81.25	76
White	Withers	56	5.11	0	82.29	71
Wilson	Lee, R	47	8.65	0	87.19	86
Grade 5						
Adams, B	Smith, L	113	76.82	0	74.86	58

Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018).

Note: ACE=Accelerating Campus Excellence. ISN = Intensive Support Network. Low implementing campuses for each grade were selected based on campus mean statistics, specifically based on the following criteria: (a) campus hour mean was less than 80 percent of the expected hour target, and (b) 49 percent or fewer of students at the campus met the hour target. Hour target=62 for grades two to four and 174 for grade five. Accuracy (Guided Study Level A) target=75 percent or higher.

Appendix F

Appendix F: RM Student Hour and Accuracy Results by Grade Level from 2015-16 to 2017-18

Grade	School Year	n	Mean	Target	Met Target %	Almost Met Target %	Did Not Meet Target %
Hour							
2	2015-16	13,078	50.16	60	35	20	45
2	2016-17	12,076	42.64	64	18	19	63
2	2017-18	4,346	32.40	62	9	12	79
3	2015-16	12,848	47.69	60	29	22	49
3	2016-17	12,368	39.39	64	16	14	70
3	2017-18	4,464	33.90	62	7	14	79
4	2015-16	6,152	33.93	60	17	14	69
4	2016-17	4,702	25.76	64	9	9	82
4	2017-18	2,388	25.90	62	8	8	84
5	2015-16	811	128.98	205	6	16	78
5	2016-17	568	101.90	177	8	11	81
5	2017-18	113	76.80	174	0	0	100
Accuracy							
2	2015-16	13,076	68.72	75%	36	39	25
2	2016-17	12,052	68.27	75%	35	38	27
2	2017-18	4,317	69.80	75%	40	38	22
3	2015-16	12,845	82.19	75%	80	16	4
3	2016-17	12,308	81.94	75%	78	17	5
3	2017-18	4,423	81.80	75%	78	17	5
4	2015-16	6,089	75.70	75%	60	27	13
4	2016-17	4,403	76.73	75%	64	23	13
4	2017-18	2,184	77.50	75%	66	22	12
5	2015-16	810	71.24	75%	42	40	18
5	2016-17	567	73.17	75%	49	36	15
5	2017-18	113	74.90	75%	57	33	10

Source: PEIMS district database (10/30/2015, 10/31/2016, and 10/30/2017). RM implementation data (1/15/2016, 1/12/2017, and 1/04/2018 for fall and 6/16/2016, 6/19/2017, and 6/05/2018 for spring).

Note: The number of students with hour and accuracy data differ because some students did not have Level A Guided Study accuracy data. For hours, Met=Logged 100 percent of expected hours; Almost Met=Logged 80 to 99 percent of expected hours; Did Not Meet=Logged less than 80 percent of expected hours. For Accuracy, Met=Attained Guided Study Level A accuracy rate of 75 percent or higher; Almost Met=Accuracy rate between 60 and 74 percent; Did Not Meet=Accuracy rate of less than 60 percent.

Appendix G

Appendix G: Number and Percentage of Teachers in 2017-18 that Completed the RM Qualification Course by Feeder Pattern and Time Period

High School Feeder	During First Six Weeks <i>n</i> (%)	After First Six Weeks <i>n</i> (%)	Not Trained <i>n</i> (%)	Total <i>n</i>
ACE	0 (0)	0 (0)	3 (100)	3
Adams, B	3 (20)	1 (7)	11 (73)	15
Carter	0 (0)	0 (0)	1 (100)	1
Hillcrest	3 (100)	0 (0)	0 (0)	3
Intensive Support Network	3 (38)	0 (0)	5 (63)	8
Jefferson	5 (56)	1 (11)	3 (33)	9
Kimball	2 (67)	1 (33)	0 (0)	3
North Dallas	3 (43)	1 (14)	3 (43)	7
Pinkston	1 (100)	0 (0)	0 (0)	1
Samuell	4 (80)	1 (20)	0 (0)	5
Seagoville HS	0 (0)	1 (20)	4 (80)	5
South Oak Cliff	0 (0)	2 (50)	2 (50)	4
Spruce	1 (25)	2 (50)	1 (25)	4
Sunset	2 (100)	0 (0)	0 (0)	2
White	1 (33)	1 (33)	1 (33)	3
Wilson	0 (0)	0 (0)	5 (100)	5
All	28 (36)	11 (14)	39 (50)	78

Source: RM PD data (01/10/2018).

Note: ACE=Accelerating Campus Excellence. These numbers were current as of the end of the fall semester. Some percentages may not add up to 100 due to rounding.

Appendix H

Appendix H: Spring 2018 RM Campus Administrator Survey Results

Please mark your level of agreement with the following.	Strongly Agree n (%)	Agree n (%)	Neither Agree Nor Disagree n (%)	Disagree n (%)	Strongly Disagree n (%)
I would like to have Reasoning Mind (RM) at my campus next year.	32 (39)	34 (42)	11 (14)	3 (4)	1 (1)
RM helps teachers on my campus to be more effective in the classroom.	25 (31)	30 (38)	20 (25)	5 (6)	0 (0)
Students at my school benefit from RM.	28 (35)	45 (56)	6 (7)	2 (2)	0 (0)
I would recommend RM to other administrators.	26 (33)	40 (50)	12 (15)	2 (2)	0 (0)
Teachers on my campus are positive toward RM.	20 (25)	35 (44)	20 (25)	4 (5)	1 (1)
How satisfied are you with the following related to RM implementation at your campus this year?	Extremely Satisfied n (%)	Satisfied n (%)	Neither Satisfied Nor Dissatisfied n (%)	Dissatisfied n (%)	Extremely Dissatisfied n (%)
Support from RM Implementation Coordinator	35 (43)	34 (42)	8 (10)	4 (5)	0 (0)
Support from campus administrators	35 (44)	35 (44)	9 (12)	0 (0)	0 (0)
Support from math campus instructional coach	33 (48)	27 (40)	8 (12)	0 (0)	0 (0)
How often have the following issues kept teachers on your campus from implementing RM with students as planned?	Very Frequently n (%)	Frequently n (%)	Occasionally n (%)	Rarely n (%)	Never n (%)
Problems with laptop(s) from COW cart	15 (20)	14 (18)	27 (35)	15 (20)	6 (7)
Problems with computer(s) in wired computer lab	5 (7)	11 (15)	25 (34)	18 (24)	15 (20)
Computer lab unavailable/delayed due to repairs/reimaging	5 (7)	4 (5)	17 (24)	18 (25)	28 (39)
Computer lab unavailable due to use by other students during RM scheduled time (e.g., testing, etc.)	6 (8)	5 (7)	17 (23)	22 (30)	24 (32)
School schedule conflicts (e.g., field trips, assemblies, testing, etc.)	1 (1)	4 (5)	21 (28)	37 (49)	13 (17)
Student log-in issues	3 (4)	2 (3)	20 (27)	35 (46)	15 (20)
Computers not charged/not holding their charge	8 (11)	11 (15)	17 (23)	25 (34)	12 (17)
Wireless connectivity issues	7 (9)	9 (12)	29 (38)	25 (33)	6 (8)
Network issues	7 (9)	10 (13)	21 (27)	31 (40)	8 (11)
Limited access to computers/scheduling issues (e.g., not enough working computers, difficulty scheduling laptop carts, etc.)	10 (13)	13 (17)	18 (24)	24 (32)	11 (14)

Source: RM campus administrator survey administered in Spring 2018.

Note: Total number of campus administrator respondents=81. Of 81, some administrators skipped a couple of survey items. In addition, administrators that marked "not applicable" were not included. For these reasons, adding up the number of respondents for some items may not equal to 81.

Appendix I

Appendix I: Spring 2018 RM Teacher Survey Results

	Every Day n (%)	2-3 Times a Week n (%)	Once a Week n (%)	2-3 Time a Month n (%)	Once a Month n (%)	1-2 Time a semester n (%)
How often did you use the following?						
Grade 2 RM	10 (10)	67 (68)	20 (20)	1 (1)	0 (0)	1 (1)
Grade 2 HMH Personal Math Trainer	0 (0)	8 (29)	9 (32)	1 (3)	0 (0)	10 (36)
Grade 2 Other	11 (39)	14 (50)	2 (7)	1 (4)	0 (0)	0 (0)
Grade 3 RM	10 (12)	58 (71)	12 (15)	0 (0)	0 (0)	2 (2)
Grade 3 Think Through Math (TTM)	1 (7)	5 (33)	5 (33)	1 (7)	0 (0)	3 (20)
Grade 3 HMH Personal Math Trainer	0 (0)	7 (39)	4 (22)	3 (17)	0 (0)	4 (22)
Grade 3 Other	3 (13)	9 (39)	8 (35)	2 (9)	0 (0)	1 (4)
Grade 4 RM	6 (15)	25 (63)	7 (17)	2 (5)	0 (0)	0 (0)
Grade 4 Think Through Math (TTM)	0 (0)	2 (50)	1 (25)	1 (25)	0 (0)	0 (0)
Grade 4 HMH Personal Math Trainer	1 (13)	0 (0)	1 (13)	4 (50)	0 (0)	2 (24)
Grade 4 Other	2 (17)	6 (50)	1 (8)	3 (25)	0 (0)	0 (0)
Please mark your level of agreement with the following.						
	Strongly Agree n (%)	Agree n (%)	Neither Agree Nor Disagree n (%)	Disagree n (%)	Strongly Disagree n (%)	
I would like to continue teaching with Reasoning Mind (RM) next year.	100 (47)	86 (41)	15 (7)	4 (2)	7 (3)	
RM helps me to be more effective in the classroom.	72 (34)	92 (44)	27 (13)	13 (6)	7 (3)	
Students in my classroom benefit from RM.	85 (41)	99 (47)	18 (9)	2 (1)	5 (2)	
I would recommend RM to other teachers.	87 (42)	88 (42)	23 (11)	7 (3)	5 (2)	
Teachers on my campus are positive toward RM.	62 (30)	99 (47)	35 (17)	8 (4)	5 (2)	
How satisfied are you with the following as you implement RM this year?						
	Extremely Satisfied n (%)	Satisfied n (%)	Neither Satisfied Nor Dissatisfied n (%)	Dissatisfied n (%)	Extremely Dissatisfied n (%)	
Support from RM Implementation Coordinator	69 (34)	106 (51)	26 (13)	5 (2)	0 (0)	
Support from principal or another campus administrator	60 (29)	97 (48)	42 (21)	4 (2)	0 (0)	
Support from math campus instructional coach	66 (34)	97 (49)	30 (15)	4 (2)	0 (0)	
To what extent did students change in the following areas as a result of using RM?						
	Significantly Improved n (%)	Improved n (%)	No Change n (%)	Regressed n (%)	Significantly Regressed n (%)	
Reasoning skills	35 (18)	130 (67)	30 (15)	0 (0)	0 (0)	
Non-standard thinking abilities	34 (18)	120 (62)	38 (20)	0 (0)	0 (0)	
Independence in learning	56 (28)	122 (61)	21 (11)	0 (0)	0 (0)	
Confidence in mathematical ability	48 (24)	123 (62)	27 (14)	0 (0)	0 (0)	
Enjoyment of mathematics	61 (31)	114 (57)	21 (11)	3 (1)	0 (0)	
How often have the following issues kept you from implementing RM with students as planned?						
	Very Frequently n (%)	Frequently n (%)	Occa- sionally n (%)	Rarely n (%)	Never n (%)	
Problems with laptop(s) from COW cart	31 (16)	33 (18)	42 (23)	42 (23)	37 (20)	
Problems with computer(s) in wired computer lab	11 (7)	18 (11)	37 (23)	34 (21)	63 (38)	
Computer lab unavailable/delayed due to repairs/reimaging	5 (3)	11 (7)	17 (11)	28 (19)	90 (60)	
Computer lab unavailable due to use by other students during RM scheduled time (e.g., testing, etc.)	8 (5)	10 (6)	21 (14)	30 (20)	84 (55)	
School schedule conflicts (e.g., field trips, assemblies, testing, etc.)	6 (3)	13 (8)	47 (28)	47 (28)	56 (33)	
Student log-in issues	4 (2)	12 (6)	33 (18)	77 (42)	59 (32)	
Computers not charged/not holding their charge	13 (7)	20 (11)	45 (25)	55 (31)	46 (26)	
Wireless connectivity issues	20 (11)	31 (17)	62 (33)	53 (28)	20 (11)	
Network issues	18 (10)	23 (12)	63 (34)	59 (32)	22 (12)	
Limited access to computers/scheduling issues (not enough working computers, difficulty scheduling laptop carts, etc.)	13 (7)	15 (8)	41 (23)	48 (27)	64 (35)	

Source: RM teacher survey administered in Spring 2018.

Note: Total number of teacher respondents=212. Of 212, some teachers skipped a couple of survey items. In addition, teachers that marked "did not use", "not applicable", or "do not know" were not included. For these reasons, adding up the number of respondents for some items may not equal to 212. HMH=Houghton Mifflin Harcourt. Results for fifth grade was not reported due to the small number of fifth-grade teacher respondents ($n < 5$).

Appendix J

Appendix J: RM Campus Staff Survey Results for Successes and Challenges

	Campus Administrator	Teacher	All
	<i>n</i> responded=72	<i>n</i> responded=194	<i>n</i> responded=266
Successes	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Increased independent learning	39 (54)	144 (74)	183 (69)
Increased mathematics and problem-solving skills	36 (50)	124 (64)	160 (60)
Increased student engagement	33 (46)	116 (60)	149 (56)
Additional ways to teach/reinforce concepts	38 (53)	111 (57)	149 (56)
RM's STAAR modules	42 (58)	84 (43)	126 (47)
Data monitoring/dashboard	29 (40)	93 (48)	122 (46)
Improved vocabulary, reading, and note-taking skills	17 (24)	85 (44)	102 (38)
More one-on-one interactions with struggling students	21 (29)	80 (41)	101 (38)
Increased use of RM	22 (31)	71 (37)	93 (35)
Increased RM accuracy	21 (29)	56 (29)	77 (29)
RM's single objective curriculum (SOC)	16 (22)	36 (19)	52 (20)
RM's online live tutoring	7 (10)	13 (7)	20 (8)
None	3 (4)	7 (4)	10 (4)
Other	1 (1)	2 (1)	3 (1)
Challenges	<i>n</i> responded=79	<i>n</i> responded=198	<i>n</i> responded=277
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Technology-related issues (e.g., inadequate working technology/access/support)	43 (54)	92 (47)	135 (49)
Difficulty meeting RM time requirements	39 (49)	75 (38)	114 (41)
Managing RM and Dallas ISD curriculum expectations	37 (47)	58 (29)	95 (34)
Preparation for district's six-week common assessments and/or ACP	23 (29)	61 (31)	84 (30)
Lack of alignment between district and RM curricula (e.g., the order in which the standards are taught throughout the year vary)	13 (17)	44 (22)	57 (21)
Scheduling issues	15 (19)	40 (20)	55 (20)
None	10 (13)	43 (22)	53 (19)
Other	3 (4)	9 (5)	12 (4)

Source: RM campus administrator survey and RM teacher survey administered in Spring 2018.

Note: Respondents were allowed to choose more than one category from each list that was made based on prior years' thematic results of open-ended campus staff survey items.

Appendix K³⁵

Appendix K: Propensity Score Matching Detailed Methods

Experimental designs with randomly assigned treatment and control groups are the “gold standard” for research and program evaluation. Random assignment means that each member of a population has an equal chance of receiving the treatment. When everyone has an equal chance of receiving the treatment, it is assumed there are no systematic differences between the treatment and control groups. Any post-treatment differences between the groups are then assumed to result from the treatment, not from pre-existing differences between the groups. An experimental design with random assignment was not possible in this evaluation. Rather, students were assigned to RM classes based upon a number of factors such as principal choice, technology availability, etc. Propensity score matching (PSM) is one alternative when random assignment is not possible (Stuart, 2010; Thoemmes, 2012).

A propensity score is a measure of the likelihood to receive treatment based on individual characteristics (Stuart, 2010). The characteristics used to determine an individual's propensity for treatment are called “covariates.” PSM resulted in individual propensity scores for RM students and non-RM students. For example, imagine a Hispanic male whose primary language at home is Spanish, does not have a disability, and is a RM student. In PSM, he will receive a propensity score that indicates the probability that someone with his characteristics will become a RM student. This male will be matched with a member of the control group (non-RM students of the same grade from the district) with a similar propensity score. Following this process, each RM student is matched with a control group member who has the most similar propensity score, thus forming a treatment and comparison group of equal sample sizes. This process is known as 1:1 nearest neighbor matching because a single treatment group member is matched with a single control group member (Thoemmes, 2012). The result is not a comparison group that matches the exact demographics of the treatment group. For example, the Hispanic male may not be matched with another Hispanic male; rather, he will be matched with a control group member with the same probability of receiving the treatment. This mimics the effect of random assignment.

In this report, PSM was performed independently for each grade from second grade to fifth grade to make a control group for RM students. The treatment group included RM students, and the control group included all other district students that did not participate in RM in 2017-18.³⁶ The following covariates from the Dallas ISD data files formed the basis for matching:

- Race
 - Asian*
 - African American
 - Hispanic/Latino
 - White
 - American Indian or Alaska Native*
 - Two or more races*
- Sex
 - Male
- Non-LEP: Student Opt-in Two-Way Dual Language Program*
- Non-LEP: Student Opt-in One-Way Dual Language Program
- English Learner (EL) Status
- Served in the Two-Way Dual Language Program*
- Served in the One-Way Dual Language Program
- ESL Program*
- Unserved in Bilingual Program*
- Parent Denial of the Dual Language, Bilingual, and ELS Programs
- Served in the Newcomer Program*

³⁵ This appendix was adapted from Appendix H in McEnturff (2014).

³⁶ Students who did not use RM in RM schools were not included in the control group.

- Female
- Talented and Gifted Program Eligibility
- Economically Disadvantaged Status
- Lunch
 - Free
 - Reduced
 - Application Denied
 - Direct Certification
- HIPPY Program*
- Dyslexia
- At Risk
- Special Education
- Homeless*
- Refugee*
- Birth Country Mexico
- Birth Country Other
- Child Primary Language Spanish
- Child Primary Language English/Spanish
- Child Primary Language All Other Languages*
- Spanish Spoken at Home
- English and Spanish Spoken at Home
- Home Language All Other Languages*
- Primary Disability
 - Emotional Disturbance
 - Learning Disability
 - Speech Impairment
 - Autism*
 - Mental Retardation
 - Auditory Impairment*
 - Orthopedic Impairment*
 - Other Health Impairment
- Prior Year Math Achievement (*TerraNova* or *STAAR* Mathematics Scores in 2016-17)

Note: *STAAR*=State of Texas Assessments of Academic Readiness. *TerraNova* Normal Curve Equivalent (NCE) scores were used for students in grades two and three, while *STAAR* scale scores were used for students in grades three to five. *STAAR* included first administration data. *STAAR Alternate 2* scores were excluded from the analyses. Covariates with an asterisk (*) were excluded for grade five because there were no fifth-grade students from either RM or non-RM group in the respective covariate category.

It is recommended to select as many covariates as possible. Furthermore, covariates should include more than demographic variables whenever possible, and covariates should not directly affect the outcome (Stuart, 2010). In the analyses, an academic covariate, prior mathematics achievement, was used in addition to demographic factors: *TerraNova* NCE scores in 2016-17 for grades two and three and *STAAR* mathematics scale scores in 2016-17 for grades four and five.

After selecting the covariates, a dichotomous treatment/control variable was created to separate RM students from the control group. Then, the covariates were used in a logistic regression with RM participation as the treatment indicator. The logistic regression resulted in a “propensity score” for each treatment and control participant, which was used as the basis for matching.

After PSM was performed for each grade, the balance of covariates was assessed using univariate and multivariate balance measures. The aim of PSM is to improve balance, which refers to the similarity of covariate distributions between the treatment and control groups. When covariates exhibited large univariate imbalance, the propensity score model was respecified to remove these covariates. Typically, large univariate imbalance was a result of having covariates with zero variance (e.g., no RM fifth-grade students were homeless). Removing covariates with zero variance resulted in better balance in all models where balance was an issue. The multivariate imbalance measures represent balance for the group of covariates as a whole, both before and after matching (see Appendix K Table). For “relative multivariate imbalance,” there is no set criterion for acceptable balance, so these are evaluated by comparing imbalance values before and after PSM. PSM successfully improved the balance of covariates when the “after” value was less than the “before” value. This was the case in the current analyses, indicating improved balancing of the covariate distributions. A successful “overall balance test” occurs when $p > .05$, indicating no significant imbalance among the covariates. The PSM samples in the current analyses met this criterion.

Appendix K Table: Sample Size and Balance Statistics from Propensity Score Matching

	Original Sample Size		Relative Multivariate Imbalance		Overall Balance Test		
	Treatment	Control ^a	Before	After	Chi-squared	df	p
Grade 2	3,671	5,765	0.638	0.547	58.465	138	1.000
Grade 3	3,835	5,981	0.644	0.553	52.326	127	1.000
Grade 4	2,105	6,483	0.585	0.388	39.875	43	0.608
Grade 5	104	10,819	0.867	0.288	14.101	27	0.980

Source: PEIMS district database (10/30/2017). TerraNova 2017 (7/19/2017). STAAR 2017 (6/15/2017).

Note: ^aAfter matching, the control group (non-RM students) had the same sample size as the treatment group (RM students).

Appendix L

Appendix L: Spring 2018 TerraNova/STAAR Mathematics Results by RM Hour^a and Accuracy^b Categories

Grade		Hour			Accuracy		
		Met	Almost Met	Did Not Meet	Met	Almost Met	Did Not Meet
TerraNova							
Grade 2	<i>n</i>	382	538	3,195	1,644	1,553	893
	Mean	56.06	51.75	47.79	62.62	46.22	29.38
	<i>SD</i>	18.67	18.29	20.54	15.82	15.24	16.53
	% at Grade Level +	74	68	58	88	56	20
STAAR							
Grade 3	<i>n</i>	305	607	3,357	3,330	712	189
	Mean	1494.61	1497.80	1459.88	1504.11	1344.41	1283.87
	<i>SD</i>	151.27	138.67	144.15	131.89	98.74	96.67
	% at Approaches +	80	86	75	88	41	20
Grade 4	% at Meets +	53	54	42	54	10	5
	<i>n</i>	190	183	1,911	1,390	451	241
	Mean	1671.15	1612.11	1571.98	1640.65	1497.60	1423.23
	<i>SD</i>	158.35	144.09	148.14	133.05	109.55	121.05
Grade 5	% at Approaches +	92	88	77	93	63	32
	% at Meets +	75	60	48	67	22	13
	<i>n</i>	-	-	111	63	37	11
	Mean	-	-	1627.95	1698.57	1550.84	1482.82
	<i>SD</i>	-	-	130.00	114.93	74.76	85.36
	% at Approaches +	-	-	84	97	76	36
	% at Meets +	-	-	57	83	24	18

Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018). TerraNova (6/11/2018). STAAR (6/20/2018).

Note: STAAR=State of Texas Assessments of Academic Readiness. *SD*=standard deviation. For TerraNova, scores were considered at or above grade level if they were at or above the 40th percentile. % at Grade Level+ = % At or Above Grade Level. % at Approaches+ = % at Approaches Grade Level or Above. % at Meets+ = % at Meets Grade Level or Above. ^aMet=Logged 100% of expected hours. Almost Met=Logged 80% to 99% of expected hours. Did Not Meet=Logged less than 80% of expected hours. ^bMet=Attained Guided Study Level A accuracy rate of 75% or higher. Almost Met=Accuracy rate between 60% and 74%. Did Not Meet=Accuracy rate of less than 60%. For descriptive statistics, mean Normal Curve Equivalent (NCE) scores were used for TerraNova, and mean scale scores were used for STAAR. For STAAR, results included STAAR first administration data. STAAR Alternate 2 scores were excluded from the analyses. Although all STAAR scores were included, some students' test scores might have been adversely affected by the online testing issues that occurred in April and May 2018. Cells marked with a dash (-) indicate no students met or almost met the hour target.

Appendix M

Appendix M: Multiple Regression Results for Spring 2018 Mathematics Scores and Various Predictors (RM Students in Grades 2 to 5)

	Correlations among Variables									β	t-value
	1	2	3	4	5	6	7	8	9		
Second Grade ($R^2=0.651$; Adj. $R^2=0.651$)											
1. TerraNova 2018 Scores	1.00										
2. GS Level A Accuracy	0.69	1.00								0.40	30.17***
3. N of Objectives Completed	0.42	0.19	1.00							0.19	17.08***
4. Prior Year TerraNova Scores	0.73	0.64	0.37	1.00						0.41	30.10***
5. Economically Disadvantaged	-0.11	-0.14	-0.02	-0.13	1.00					0.01	0.78
6. Gender	0.00	-0.10	0.08	-0.02	-0.03	1.00				0.04	3.87***
7. Race (African American)	-0.13	-0.13	-0.17	-0.10	0.07	0.02	1.00			-0.01	-0.61
8. Race (Hispanic)	0.03	0.01	0.13	0.00	0.05	-0.01	-0.78	1.00		-0.01	-0.47
9. Race (Other)	0.05	0.07	0.02	0.03	-0.04	-0.02	-0.08	-0.34	1.00	0.01	0.75
Mean	49.94	70.36	13.80	47.68	0.80	0.51	0.16	0.77	0.03		
Standard Deviation	19.93	13.91	8.09	19.34	0.40	0.50	0.36	0.42	0.18		
Third Grade ($R^2=0.589$; Adj. $R^2=0.588$)											
1. STAAR 2018 Scores	1.00										
2. GS Level A Accuracy	0.65	1.00								0.31	21.79***
3. N of Objectives Completed	0.29	0.24	1.00							0.09	8.26***
4. Prior Year TerraNova Scores	0.71	0.64	0.26	1.00						0.48	34.45***
5. Economically Disadvantaged	-0.10	-0.10	-0.02	-0.11	1.00					-0.02	-1.68
6. Gender	0.00	-0.08	0.10	0.05	0.00	1.00				-0.01	-0.97
7. Race (African American)	-0.18	-0.13	-0.08	-0.10	-0.04	-0.01	1.00			-0.10	-4.59***
8. Race (Hispanic)	0.10	0.08	0.08	0.04	0.20	0.02	-0.80	1.00		-0.02	-0.87
9. Race (Other)	0.05	0.03	0.00	0.01	-0.06	-0.01	-0.06	-0.31	1.00	0.03	1.88
Mean	1471.71	82.80	7.18	49.03	0.88	0.52	0.13	0.80	0.02		
Standard Deviation	145.05	10.92	5.56	19.07	0.33	0.50	0.34	0.40	0.15		
Fourth Grade ($R^2=0.640$; Adj. $R^2=0.638$)											
1. STAAR 2018 Scores	1.00										
2. GS Level A Accuracy	0.57	1.00								0.18	10.15***
3. N of Objectives Completed	0.25	0.16	1.00							0.11	7.33***
4. Prior Year STAAR Scores	0.78	0.58	0.20	1.00						0.64	35.62***
5. Economically Disadvantaged	-0.09	-0.06	0.09	-0.09	1.00					-0.03	-2.06*
6. Gender	0.03	-0.02	0.06	0.08	-0.02	1.00				-0.02	-1.30
7. Race (African American)	-0.20	-0.10	0.05	-0.16	-0.02	0.01	1.00			-0.09	-3.11**
8. Race (Hispanic)	0.15	0.09	-0.01	0.13	0.21	0.00	-0.76	1.00		-0.01	-0.40
9. Race (Other)	-0.01	-0.06	0.05	-0.03	-0.06	-0.04	-0.09	-0.38	1.00	0.00	-0.04
Mean	1589.34	78.30	5.64	1472.19	0.92	0.52	0.15	0.76	0.04		
Standard Deviation	150.60	15.28	6.70	151.64	0.28	0.50	0.36	0.43	0.20		
Fifth Grade ($R^2=0.731$; Adj. $R^2=0.717$)											
1. STAAR 2018 Scores	1.00										
2. GS Level A Accuracy	0.74	1.00								0.27	3.39***
3. N of Objectives Completed	0.13	0.15	1.00							-0.02	-0.30
4. Prior Year STAAR Scores	0.84	0.75	0.15	1.00						0.64	8.05***
5. Economically Disadvantaged	0.09	0.09	0.01	0.13	1.00					-0.01	-0.27
6. Gender	0.02	0.03	0.11	-0.04	-0.03	1.00				0.04	0.68
Mean	1631.26	75.54	38.23	1593.04	0.97	0.58					
Standard Deviation	131.04	10.22	24.05	138.40	0.17	0.50					

Source: PEIMS district database (10/30/2017). RM implementation data (1/04/2018 for Fall 2017 and 6/05/2018 for Spring 2018). TerraNova (7/19/2017 and 6/11/2018). STAAR (6/15/2017 and 6/20/2018).

Note: STAAR=State of Texas Assessments of Academic Readiness. $n=3,589, 3,737, 1,817$ and 104 for grades two, three, four, and five, respectively. GS=Guided Study. Adj. R^2 =Adjusted R^2 . * $p<.05$. ** $p<.01$. *** $p<.001$. Results included STAAR first administration data. STAAR Alternate 2 scores were excluded from the analyses. Although all STAAR scores were included, some students' test scores might have been adversely affected by the online testing issues that occurred in April and May 2018. For fifth grade, ethnic groups were not included due to small number of students in White and other ethnic groups.