Texas SUCCESS Comprehensive Evaluation Report

for the Texas Education Agency

Texas SUCCESS Comprehensive Evaluation Report

Contributing Authors

Marshall Garland
Joseph Shields
Eric Booth
Shana Shaw, Ph.D.
Karin Samii-Shore

Submitted By:

GIBSON
CONSULTING GROUP
1221 South MOPAC Expressway
Suite 355
Austin, Texas 78746
www.gibsonconsult.com

Contents

Acknowledgements

The authors of this report are grateful to the many organizations, school districts, and individuals who have contributed to this report. Staff from Istation Reading and Think Through Math provided the evaluation team with system usage data and generously donated their time to discuss technical issues related to the students-level usage and performance data generated for each of the online systems. Personnel from selected school districts and campuses across the state campuses directly assisted the evaluation by participating in telephone interviews, which provided valuable data to the research team. This report is sponsored by the Texas Education Agency, and many staff were involved in providing data for analysis, and feedback on results.

List of Tables

List of Figures

List of Acronyms Used in this Report

Accelerated Reading Instruction and Accelerated Mathematics Instruction (ARI/AMI)

Beginning of Year (BOY)

Difference in Difference (DiD)

Education Service Center (ESC)

End-of-course (EOC)

English Lanugage Arts (ELA)

English Language Learner (ELL)

Gibson Consulting Group, Inc. (Gibson)

Istation Reading (Istation)

Istation Indicators of Progress (ISIP)

Istation Indicators of Progress-Early Reading (ISIPER)

Istation Indicators of Progress-Advanced Reading (ISIPAR)

Ordinary Least Squares regression (OLS)

Phase-in 1 Level II Satisfactory standard (STAAR passing standard)

Public Education Information Management System (PEIMS)

Professional Development (PD)

Shore Research, Inc. (Shore)

State of Texas Assessments of Academic Readiness (STAAR)

Student Success Initiative (SSI)

Tejas Lectura en Español (Tejas LEE)

Texas Assessment of Knowledge and Skills (TAKS)

Texas Education Agency (TEA)

Texas Essential Knowledge and Skills (TEKS)

Texas Students Using Curriculum Content to Ensure Sustained Success (Texas SUCCESS)

Think through Math (TTM)

Executive Summary

Background

The Texas Students Using Curriculum Content to Ensure Sustained Success (SUCCESS) program offers state-funded access to computerized interactive mathematics and reading programs provided by two vendors—Istation Reading (Istation) and Think Through Math (TTM)—to all Texas public school students in Grades 3-8. Istation and TTM are adaptive programs designed to support student achievement by adjusting content based on student skill level and incorporating assessments to track student performance changes. When these online programs were selected by the Texas Education Agency (TEA) for the Texas SUCCESS initiative, the correlation between Istation and TTM content and the Texas Essential Knowledge and Skills (TEKS) for reading and mathematics was documented.

In fulfillment of Rider 50 (General Appropriations Act, Article III, 83rd Texas Legislature) Gibson Consulting Group, Inc. (Gibson), in partnership with Shore Research, Inc. (Shore), employed a mixed-methods approach to evaluating the Texas SUCCESS program implemented in school districts across the state. The evaluation plan incorporated in-depth examinations of the Texas SUCCESS program through a number of different sources, including online system usage and student growth data, interviews with district and campus academic intervention staff in local education agencies using the programs, and extensive analysis of student outcomes data related to the 2013-14 State of Texas Assessments of Academic Readiness (STAAR®) performance. Usage and outcomes for the two online learning programs (i.e., Istation and TTM) were examined in great detail by the evaluation team.

Key Findings

Reading

Istation is a supplemental reading program that provides computer-adaptive instruction in an animated environment that designed to improve phonemic awareness, alphabetic knowledge, vocabulary, and reading comprehension. Although Istation includes curricular materials for Grades Pre-Kindergarten through 8, it is offered free-of-charge to Texas public school students in Grades 3-8 as part of the Texas SUCCESS program. Istation includes an integrated assessment tool, administered monthly or upon log-in if more than a month has passed, that tailors the program's curriculum to address students' individual academic needs. The Istation vendor recommends that elementary school students receive a minimum of 250 minutes of exposure to the Istation curriculum and middle school students use the system for a minimum of 200 minutes.

Istation System Usage and Implementation

While 87% of students across the state enrolled in Istation, just over half used the system.

In 2013-14, the vast majority (87%) of students in Grades 3-8 across the state were registered to use the Istation system; however, just over half (55%) actually logged into an Istation curriculum session.

More elementary school students used Istation—and they used the system more frequently and for longer periods of time—than middle school students. Across all grades, lower performing and at-risk students were more likely to be identified to use Istation.

System usage varied widely across grade levels, school characteristics, and student groups. Students in elementary grades were much more inclined to attempt one or more curriculum sessions (67% to 71%, depending upon grade level) than students in middle schools (35% to 46%). Also, elementary school students logged longer and more frequent curriculum sessions compared to students in middle grades. Approximately one third (33%) of the students in Grades 3-5 met the recommended minimum usage threshold of 250 minutes for elementary school students, compared to just 10% of students in Grades 6-8 who met the recommended minimum usage threshold of 200 minutes for middle school students. Differences in Istation usage were also observed across different types of schools, with Title I schools, schools rated as *Improvement Required*, and schools in urban districts using the system at higher rates than other campuses. In addition, geographic differences in usage and attendance at Istation-related professional development were observed.

Variation in system usage was also observed when various student characteristics were taken into account, showing that lower performing students and students classified as English Language Learners (ELL) were more likely to be identified for Istation. Students in the bottom quartile of 2012-13 STAAR-Reading scores were substantially more likely to use Istation, and to use it more frequently, than students in upper quartiles of STAAR performance, a difference that was substantially larger for students in middle school grades. Likewise, students retained in grade in a prior school year and students classified as ELL used the system at higher rates than other students. The differences between ELL and non-ELL students were larger at the middle school level.

Most interviewees felt that they had sufficient instructional staff to implement Istation, but cited that a shortage of computers was a challenge for implementation.

To gain a better understanding of reasons for variation in Istation usage across Texas campuses, the evaluation team conducted telephone interviews with reading interventionists responsible for implementing Istation and other reading programs in their respective districts and schools. At both the campus- and district-levels, most interviewees noted that schools had sufficient instructional staff to implement Istation. What is more, most interviewees did not cite issues with internet connectivity—80% of campus staff were satisfied with connectivity at their schools. However, just 55% of school staff noted that they had an adequate number of computers or laptops to effectively implement Istation at their campus. This shortage of computers was noted as an issue more frequently by staff at campuses using Istation at low levels, as well as staff at middle school campuses.

In addition to inquiries regarding support for Istation implementation, the research team asked interviewees about various professional development offerings on the programs, which were made available to district and campus staff across the state in both the 2012-13 and 2013-14 school years. In total, approximately 469 school district and charter school organizations—representing 46% of districts in the state—took advantage of these trainings. Region 20 education service center (ESC 20, San Antonio) also served a technical support function for districts and campuses that had questions about the programs. Campus staff tended to rate the support, training, and technical assistance from the Istation vendor fairly low, with just 36% reporting that they were "very satisfied." Satisfaction rates regarding ESC 20 telephone support were higher (61% of interviewees stated that they were "very satisfied" with the support).

In terms of actual implementation, campuses made the bulk of decisions regarding how Istation was used, with interviewees commonly noting that they used both the instructional and assessment-based features of Istation and that the system was typically used to support regular classroom instruction, though the settings in which the system was used varied by school level. Almost half of middle schools used Istation exclusively in computer labs, while less than 10% of elementary schools did so. In elementary schools, Istation was typically used in blended classroom settings (i.e., classrooms with dedicated computers). Middle schools were also more likely to report identifying students for Istation based on prior performance on standardized tests. Other variations in implementation were tied to levels of system use. For example, low Istation usage campuses were less likely to report using the system for the entire 2013-14 school year. Regarding other reading interventions and programs, only 8% of school staff reported using only Istation, while 92% indicated that they used at least one other reading program.

Istation Student Outcomes

Changes in Istation Assessment Performance

To examine relationships between Istation use and gains on reading assessments administered via the program in 2013-14 (i.e., the Istation Indicators of Progress (ISIP)), the evaluation team first assessed the frequency and timing of Istation use and assessment administration among elementary and middle school Istation users. Roughly 90% of elementary students took at least two Istation assessments, while the same was true of 70% to 80% of middle school students (depending on grade level). In general, this confirms that Istation use was more widespread in elementary grades and more targeted in middle grades, with middle grade students from lower performance ISIP tiers engaging with Istation more frequently than their peers in higher performance ISIP tiers.[1]

Istation use was more widespread in elementary grades and more prevalent among lower performing students in middle grades.

[1] Students were required to take an ISIP assessment at least once per month, or after more than one month of not logging in to the system.

Descriptive trends showed that, while 70% of students remained in the same performance tier between their first and last Istation assessment, the largest gains were observed among students with more elapsed time between assessments.

Reading gains on the ISIP were assessed among those students who took at least two ISIP assessments. About 70% of students remained in the same performance tier between their first and last assessments.[2] This is not to say that there were no gains on the ISIP—gains were demonstrated across all grades, with the largest gains shown among students with more elapsed time between their first and last assessments. That is, gains were largest among students who took their first ISIP assessment in September and their last assessment in May (nine months between first and last assessments). Students with an eight or nine month gap between their first and last assessment were more likely to use the system more frequently and with more intensity during this period than students with a smaller gap. In addition, the longer the span between students' first and last assessment, the more classroom instruction and content they have been exposed to in school, and within the Istation. Both of these likely contribute to how much a student improves on the ISIP assessments. Irrespective of the number of system usage minutes, students demonstrated considerable growth on the assessment based on the amount of time that elapsed between the first and second assessments.

No clear relationships were observed between students' use of Istation and their performance on Istation assessments, except for small effects seen in Grades 3 and 8.

With the exception of Grades 3 and 8, there were not clear relationships between Istation use and performance on the Istation assessments. In most grades, students' use of the system was not significantly related to how well they performed on the assessments. That said, in Grade 3, increased use was positively correlated with gains in Istation assessments, regardless of the time elapsed between a student's first and last assessment. Patterns in Grade 8 were less straightforward, with only slight gains demonstrated among students, and then only among students who used the system for long periods of time and had nine months between their first and last assessments.

Relationships between Istation Usage and Performance on the STAAR-Reading Assessment

Descriptive results showed that higher levels of Istation usage was associated with poorer reading performance, except in Grades 7 and 8. In those grades, students who used the system for 300 or more minutes had descriptively better gains than their peers.[3] This may be a result of lower performing students using Istation more intensively. Since descriptive analyses did not account for other observable factors

[2] It is not unexpected that the majority of students would remain in the same performance tier because it is likely that students either at the top or the bottom of their starting performance tier would be most susceptible to movement from one tier to another (either up or down).

[3] In the descriptive analyses, this may be a result of lower performing students using Istation more intensively. However, the relationship held when statistical models controlled for prior achievement on the STAAR-Reading assessment.

that can also influence student achievement, statistical models were designed to take into account other factors (student-level and school-level) when considering direct relationships between Istation use and reading gains. Several different methods of measuring achievement outcomes and Istation participation were used in the models to more accurately reflect the nature of changes in student performance and the manner in which Istation was used. In particular, the results below reflect the most precise measurement of student reading outcomes; that is, reading gains that were measured in ways that allowed growth expectations to be different for students based on their 2012-13 STAAR-Reading performance.

Advanced statistical models—which accounted for other observable factors that may influence outcomes, such as students' prior STAAR-Reading performance—showed that, generally irrespective of usage, students using Istation in Grades 4-6 demonstrated less growth on STAAR-Reading than students who did not use Istation. Although not substantial, students in Grades 7 and 8 demonstrated more growth on STAAR-Reading than students who did not use Istation. Descriptively, in Grades 7 and 8, students who used the system 300 or more minutes had higher descriptive reading gains than students who did not use Istation.

Across different model specifications and different categorizations of student performance and program participation, statistical modeling results for students in elementary grades were consistent with descriptive results. Istation use was associated with statistically significant smaller reading gains on STAAR among students in Grades 4 to 6.[4] In models where reading achievement was measured in light of students' prior performance, results in Grades 7 and 8 were small and significantly positive. Istation usage had small, positive associations with reading gains among students in Grades 7 and 8.

With few exceptions, no significant differences emerged among students from different groups in terms of relationships between use of Istation and STAAR-Reading performance.

To address whether relationships between reading gains and different measures of Istation usage varied by student groups, three separate models were analyzed to assess relationships between Istation use and reading achievement, among students by ethnicity, ELL status, and economically disadvantaged status. Results demonstrated largely negligible variability in Istation usage and reading gains among different student groups, with a few exceptions. In Grades 4 and 7, negative relationships between Istation usage and reading gains among ELL versus non-ELL students were statistically significant, meaning that additional time on the system was associated with smaller reading gains among students identified as ELL compared to non-ELL students. In Grade 5, African American students with higher Istation

[4] Across these analyses, effects were typically significant at a minimum of p<0.01, which means that there is less than a 1% chance that these findings were due to chance. It is also important to note, however, that statistical significance is heavily influenced by sample size, meaning that with the large samples used in these analyses, even small relationships between program participation and achievement outcomes may have registered as statistically significant.

usage demonstrated significantly larger, positive gains than Hispanic students with comparable Istation use. It is important to note that the practical significance of these associations was quite minor.

To further explore associations between Istation use and reading gains, five different—more precise—measures of program participation were developed to analyze the effect of "dosage," or incremental increases and decreases in Istation usage.

Considering different intensities of Istation use—or "dosage"—students in Grades 4 and 5 who used Istation more intensely demonstrated less growth on STAAR-Reading than non-users, while greater intensity of Istation use among students in Grades 7 and 8 was associated with small positive gains on STAAR-Reading compared to non-users.

In Grades 4 and 5, each Istation dosage measure was significantly and negatively related to student gains in reading, although the magnitude of the relationships was small. Average negative effects were larger for Grade 4 students than for students in Grade 5. In Grades 7 and 8, dosage effects were positive and statistically significant across each measure of usage intensity. As with previously observed relationships between Istation use and reading outcomes, these effects remained small.

Relationship between Consistency and Timing of Istation Usage and Performance on the STAAR-Reading Assessment

To assess whether consistent and timely Istation use was associated with better reading outcomes, two additional measures of Istation usage were created: usage *proximity* to the STAAR test administration and use *continuity* throughout the school year.[5]

Consistent use of Istation through the school year was associated with smaller reading gains in Grades 4 and 5, but with positive gains on STAAR-Reading among students in Grades 7 and 8.

In Grades 4 and 5, continuous usage was significantly associated with smaller reading gains, although the magnitude of the relationships was small. In Grades 7 and 8, the relationship between usage continuity and reading gains was positive, and gains were notably larger than those observed in Grades 4 and 5.

With regard to findings related to usage proximity to the STAAR test, no statistical association was found between Istation use by students in close proximity to the STAAR test and STAAR-Reading gains. This finding held across all grade levels.

Using Istation in close proximity to STAAR administration was not associated with significant gains on STAAR-Reading performance.

[5] Continuity was defined as the number of months in which a student used Istation for at least 30 minutes per month.

Relationship between Istation Usage and Performance on the STAAR-Reading Assessment – Students at Risk of Being Retained

The evaluation team also explored relationships between Istation usage—and usage intensity—and changes in student performance on STAAR-Reading between students at risk of being retained and those not at risk of being retained. *At-risk* was operationalized as students who, in 2011-12 or 2012-13, failed at least one STAAR-Reading assessment. Analyses were confined to students in Grades 5 and 8 in 2013-14 only—i.e., grades where promotion depends on whether students pass STAAR exams—and who were enrolled in schools where Istation was used.

Students at risk of being retained in Grades 5 and 8 used Istation more intensively in 2013-14 than students not at risk of being retained. Among those categorized as at-risk, Grade 8 students who used Istation at high levels demonstrated better reading results than at-risk Grade 8 students who used Istation at lower levels.

Descriptive analyses demonstrated that, in Grades 5 and 8, just over 30% of students in 2013-14 were at risk of being retained. Also in both grades, students at risk of being retained had lower gains between 2012-13 and 2013-14 than students who were not at risk, with the largest gaps occurring among Grade 8 students. With regard to Istation use, at-risk students in both grades used the system more intensively in 2013-14 than their peers. And, among students who were at risk of being retained in grade, Grade 8 students who used the system at higher dosage levels demonstrated better gains than students who used the system at lower dosage levels.

The descriptive results were supported by multivariate regression results, suggesting that, among Grade 8 students classified as at-risk—who comprise almost a third of the population of Grade 8 students—supplementary instruction provided by Istation may have yielded greater benefits in reading than for students not at risk of being retained. The same could not be said for Grade 5 students classified as at risk.

Relationship between Istation Usage After Failing the First Administration of the STAAR-Reading Assessment and Performance on the Second Administration of the Exam

To assess the relationship between program usage after failing the first administration of the STAAR-Reading assessment and the probability of passing subsequent administrations of the STAAR exam, evaluators examined usage and usage intensity during the period between failure of the first administration and the second administration of STAAR-Reading in Student Success Initiative (SSI) grades (Grades 5 and 8). It is important to note a potential limitation with the findings for this research question. Students who fail the first administration of the STAAR assessment are provided with a wide array of intensive academic interventions which vary by school district, which makes it difficult to tie Istation usage to student outcome results on the second administration of the STAAR assessment.

Grade 5 students who failed the first administration of STAAR-Reading—and who used Istation for 100 or more minutes before the second administration of STAAR—were significantly more likely to pass the retest than students who did not use Istation.

Descriptive usage patterns varied by grade, with approximately 22% of Grade 8 students who failed the first administration using Istation in the period between the first and second administrations, compared to 50% of Grade 5 students. Intensity of usage among Grade 5 students who failed the first administration of Grade 5 STAAR-Reading was nearly three times higher than Grade 8 students (65 minutes compared to 21 minutes). Among students who used the system—relative to those who failed the first administration and did not use the system—the passing rate for Grade 5 Istation users was roughly three percentage points higher than students who did not use the system (41% compared to 38%). Alternatively, Grade 8 Istation users had comparable passing rates to non-users (35.9% compared to 36.2%).

To control for other factors that may have impacted retest passing rates—outside of Istation use—the evaluation team conducted statistical analyses that adjusted for other student attributes.

Among Grade 5 students who failed the first STAAR-Reading assessment, students who used the system for 100 or more minutes in the period between the first and second administrations were significantly more likely to pass the retest compared to students who did not use the system. There was also a positive and statistically significant relationship between intensity of use and reading performance, suggesting that Grade 5 students who used the system more intensively in the interim period were more likely to pass the STAAR-Reading assessment.

Relationships between usage, and usage intensity, were negatively correlated with the probability of passing the second administration among Grade 8 students, although coefficients were not statistically significant.

Mathematics

TTM is a supplemental mathematics program that provides web-based adaptive instruction in an animated environment that is designed to improve students' understanding of critical math concepts and problem-solving skills. TTM includes instructional materials that cover math content for Grades 3-Algebra I. TTM is offered free-of-charge to Texas public school students in Grades 3-8 as part of the Texas SUCCESS program. TTM includes a diagnostic assessment tool that maps out a learning pathway based on students' individual academic needs and students' pathways are adjusted in response performance on quizzes given at the completion of lessons. The TTM vendor suggests that students attempt a minimum of 5 lessons but recommend students attempt 10 or more lessons.

TTM System Usage and Implementation

Usage of Think Through Math decreased by grade level, with the lowest proportion of users in Grade 8.

Over 63% of students in Grades 3-8 across the state had no record of TTM usage in 2013-14.[6] Of those students who logged into TTM, the vast majority completed at least one session, with use of the system decreasing by grade level. Forty-three percent of Grade 3 students and 44% of Grade 4 and 5 students completed at least one TTM lesson as compared to just 21% of Grade 8 students, 23% of Grade 7 students and 29% of Grade 6 students.

System usage also varied widely across school characteristics and student groups. For example, a higher percentage of students in Title I schools attempted the recommended threshold of at least five TTM sessions (78% versus 73%, Title I versus non-Title I respectively), while a higher percentage of non-Title I students passed at least five lessons (52% versus 47%, non-Title I versus Title I respectively). Schools rated as *Improvement Required*, Charter Schools and schools in Rural, Non-Metropolitan Fast Growing, and Non-Metropolitan Stable Growth districts also used the system at higher rates than other campuses.

There was broad use of Think Through Math across student populations. Students did not appear to be systematically selected to use the Think Through Math system based on specific characteristics.

Based on the TTM usage data, schools and teachers did not systematically identify students for the TTM intervention based on students' prior performance. There was little or no descriptive relationship between 2012-13 STAAR-Mathematics assessment scores and the number of TTM lessons attempted during the 2013-14 school year. The same was true considering other student characteristics, such as 2012-13 grade retention status, ELL status, and student demographic characteristics.

To gain a better understanding of reasons for variation in TTM usage across Texas campuses, the evaluation team conducted telephone interviews with mathematics interventionists responsible for implementing TTM and other mathematics programs in their respective districts and schools. Over three quarters (78%) of campus staff interviewed indicated that they experienced some barriers when attempting to implement the TTM program at their schools. The most common barriers included not having enough computers (35%), not having enough time in the daily schedule (26%), and technology issues (23%).

In addition to inquiries regarding supports for TTM implementation, the research team asked interviewees about various professional development offerings related to the program. Trainings for TTM were made available to district and campus staff across the state in both the 2012-13 and 2013-14 school years. In

[6] Approximately 72% of students were accounted for when the evaluation team matched TTM and TEA Public Education Information Management System (PEIMS) data files. This excluded students who registered for TTM but never completed lessons. This is important, because approximately 28% of students from the TTM registration roster could not be linked back to TEA administrative records due to missing or erroneous student identification numbers. This, most likely, led to an underreporting of the number of students who used the system.

total, approximately 438 school district and charter school organizations—representing 43% of districts in the state—took advantage of these trainings. ESC 20 (San Antonio) also served a technical support function for districts and campuses that had questions about the programs. Training did not appear to be a major issue for TTM implementation, as the majority of interviewees (60%) felt that the training they received was sufficient for them to use the system effectively. However, staff at middle schools (69%) were more likely than staff at elementary schools (52%) to state that the TTM-related training was sufficient. The majority of campus mathematics interventionists interviewed in spring 2014 (76%) indicated that staff at their campuses were trained on how to use the TTM program. Not surprisingly, campuses using the TTM system heavily were more likely (87%) to indicate their staff had been trained on the system relative to moderate (66%) or low TTM usage (74%) campuses. District staff also appeared to be generally satisfied with the quality of training received from TTM; however, they felt the training could have targeted specific system features that would have benefitted teachers more.

Considering actual implementation, campuses made the bulk of decisions regarding the nature of TTM use. The TTM program offers both curricular and assessment functionality. Across all campuses, almost half of interviewees noted using just TTM curricular resources. An equal proportion of interviewees noted using both the curricular and assessment-based functions of TTM. Almost three quarters of interviewees noted the system was typically used to support regular classroom instruction. There were also notable differences between elementary and middle schools in the settings in which TTM was used—that is, almost half of middle schools used TTM in computer labs while the same was true of 33% of elementary schools, where TTM was most commonly used in blended classroom settings (i.e., classrooms with dedicated computers). Despite usage results demonstrating that overall TTM use did not seem to be targeted toward specific groups, some middle school staff reported that they identified students for the TTM intervention based on prior STAAR performance on standardized assessments.

Instructional staff needed to implement Think Through Math was sufficient, however a shortage of technological resources to implement an online program may have been an issue at some campuses.

Outside of TTM, most district- and campus-level staff reported using other strategies or programs to support students in meeting SSI grade promotion requirements. With regard to other mathematics programs, only 18% of schools reported using TTM exclusively as supplementary math program, while 82% of schools indicated that they used at least one other mathematics program. Considering whether these other programs were coordinated with TTM, less than half of interviewees indicated that this coordination occurred, citing a lack of time and scheduling difficulties as reasons for the lack of coordination. Despite differences in support staffing levels, at both the campus- and district-levels, most interviewees noted that their schools had sufficient instructional staff to implement mathematics programs. There was less consensus about whether campuses had sufficient computer access and technological resources to implement online mathematics programs, particularly among campuses that registered low and moderate TTM usage in 2013-14.

TTM Student Outcomes

TTM Lesson Performance

Overall, 94% of students who used Think Through Math passed at least one lesson in 2013-14, and of all lessons attempted approximately 43% were passed across students in Grades 3-8.

To examine relationships between TTM usage levels and progress within the TTM system, the evaluation team first assessed the TTM assessment passing rates, overall and by different student groups. Overall, 94% of TTM users attempted at least one lesson in 2013-14, though passing rates were higher among elementary students (96% to over 99%) depending upon grade level) than middle school students (85%-90%) depending upon grade level). Of all lessons attempted in 2013-14, roughly 43% were passed, with similar passing rates across elementary and middle grades. Average passing rates across all grades were lowest among students with the lowest number of TTM lessons attempted (1 to 4 lessons) and highest among students attempting 20 or more lessons.

Analyses also considered passing rates by student groups. Across all grades, passing rates were lower among students identified as ELL and economically disadvantaged. There were also gaps in passing rates by race/ethnicity, with Asian students having the highest passing rates and African American and Hispanic students having the lowest. Across all student groups, with the exception of Grade 7, passing rates were highest in August. In subsequent months, passing rates either declined or plateaued.

Across all Think Through Math lessons attempted by students in Grades 3-8, one third were remedial or retaken lessons.

Students' progress in using TTM was defined as a function of whether students remained on prescribed TTM lesson pathways (on path) or had to repeat lessons or take remedial lessons. Across all lessons attempted in 2013-14, almost two thirds were considered on path, with the remaining split between remedial or retaken lessons. Of those students who did have to take at least one remedial lesson—or retake a lesson—the majority of elementary (95%) and middle school (85%) students were able to get back on path.

Last, the relationship between TTM usage levels and associated passing rates and successful progress on TTM lessons were examined. There was not a consistent relationship between the number of attempted lessons (i.e., the primary TTM usage metric) and a students' ability to successfully progress through the TTM system for students attempting the most common types of lessons (i.e., lessons assigned to students based on their performance within the system or lessons designed to provide students with remedial instruction).

Relationship between TTM Usage and Performance on the STAAR-Mathematics Assessment

Because descriptive analyses did not account for other observable factors that can also influence student achievement, statistical models were designed to take into account other factors (student-level and school-level) when considering direct relationships between TTM use and mathematics gains.

Advanced statistical modeling, accounting for other observable factors that may influence student outcomes—such as students' prior STAAR-Mathematics performance—showed that students in Grades 3-8 who attempted 20 or more Think Through Math lessons had statistically significant higher STAAR-Mathematics scores than non-users.

Several different methods of measuring achievement outcomes and TTM use were included in the models to more accurately reflect the nature of changes in student performance and the manner in which TTM was used. In particular, the results below reflect the most precise measurement of student mathematics outcomes, which controlled for differences in students' 2012-13 STAAR-Mathematics performance.

Across different model specifications and different categories of program participation and mathematics outcomes, multivariate analyses demonstrated that TTM users in Grades 3-8 who used the system more frequently—particularly those who attempted 20 or more lessons—had STAAR-Mathematics gains that were significantly and substantively greater than non-users. However, among students who used the TTM program at low levels (i.e., those who attempted between 1 and 4 lessons), smaller gains in 2013-14 were observed.

The evaluation team also explored the extent to which the number of TTM lessons passed—relative to the number of lessons attempted—was associated with STAAR-Mathematics outcomes. Among all TTM users—both students who used the system above and below the five-lesson threshold, STAAR-Mathematics scores improved as TTM lesson passing rates increased. This implies that students' success in passing TTM lessons was positively related to students' performance on STAAR-Mathematics tests.

Passing Think Through Math lessons was positively related to student performance on STAAR-Mathematics tests, meaning that passing Think Through Math lessons was an important indicator of readiness for STAAR-Mathematics content.

Relationship between Consistency and Timing of TTM Usage and Performance on the STAAR-Mathematics Assessment

Consistent usage of Think Through Math at the recommended threshold—five or more lessons completed per month—was associated with statistically significant gains on STAAR-Mathematics, particularly among middle school students.

To assess relationships between the consistency and timing of TTM usage throughout the school year and mathematics outcomes, three additional measures of TTM usage were created: usage *continuity* throughout the school year, usage *proximity* to the STAAR test administration in April, and the *proportion of use* concentrated into the three months before the test was administered.

Of the three measures, usage continuity had the strongest, statistically significant, positive associations with STAAR-Mathematics performance gains, particularly among middle grade students. For example, six months of five or more TTM was associated with larger gains in mathematics performance among Grade 8 students.

Across Grades 3-8, using Think Through Math in close proximity to STAAR administration was associated with small but positive gains on STAAR-Mathematics tests.

With regard to usage proximity to the STAAR-Mathematics test, positive, statistically significant associations emerged across all grades between mathematics gains and each additional lesson attempted in the month before STAAR administration. Statistically significant relationships between usage concentrated in the three months before STAAR and standardized gains were found among middle grade students, though the magnitude of these associations was quite small. The proportion of usage concentrated in the three months before STAAR was not significantly associated with mathematics gains in Grade 4.

Relationship between TTM Usage and Performance on the STAAR-Mathematics Assessment – Students at Risk of Being Retained

The evaluation team also explored relationships between TTM usage and changes in student performance on STAAR-Mathematics between students at risk of being retained and those not at risk of being retained in Grades 5 and 8.[7]

For students at risk of being retained in grade, the relationship between Think Through Math usage and STAAR-Mathematics results were inconsistent. A small positive association with system usage emerged in Grade 5 and a small negative relationship was found in Grade 8.

Results of these analyses were somewhat mixed. Grade 5 students at risk of being retained, who attempted relatively low numbers of TTM lessons (between 5 and 9 lessons) during the 2013-14 school year, performed better on the 2013-14 STAAR-Mathematics assessment than students attempting comparable numbers of TTM lessons who were not at risk of being retained.[8] However, for Grade 8 students at risk of being retained, there was a negative, statistically significant relationship between TTM use and mathematics gains compared to students not at risk of being retained, but only for those attempting 15-19 lessons.

[7] Similar to the Istation analysis, at-risk was operationalized as students who, in 2011-12 or 2012-13, failed at least one STAAR-Mathematics assessment, which resulted in roughly one third of students in both grades being classified as at-risk.

[8] There was no effect for students attempting 10 to 20 lessons.

Relationship between TTM Usage after Failing the First Administration of the STAAR-Mathematics Assessment and Performance on the Second Administration of the Exam

To assess the relationship between program usage after failing the first administration of the STAAR-Mathematics assessment and the probability of passing subsequent administrations of the STAAR exam, evaluators examined usage and usage intensity during the period between failure of the first administration and the second administration of STAAR-Mathematics in SSI grades (Grades 5 and 8).[9]

Grade 5 students who failed the first administration of STAAR-Mathematics and attempted 15 or more Think Through Math lessons before the second administration of STAAR were significantly more likely to pass the STAAR retest than students who did not use Think Through Math.

Descriptive usage patterns varied by grade, with approximately 14% of Grade 8 students who failed the first administration using TTM in the period between the first and second administrations, compared to 29% of Grade 5 students. Grade 5 students who failed the first administration of STAAR-Mathematics completed over two times as many TTM lessons between the first and second administrations of the test as Grade 8 students (2.3 compared to 1 attempt). Among students who used the system—relative to those who failed the first administration and did not use the system—the STAAR-Mathematics retest passing rates for both Grades 5 and 8 TTM users were 3 percentage points higher (45% versus 42% in Grade 5) compared to students who did not use the system after failing the first administration.

Grade 8 students who failed the first administration of STAAR-Mathematics and attempted between 1 and 9 Think Through Math lessons before the second administration of STAAR were significantly more likely to pass the STAAR retest than students who did not use Think Through Math.

To control for other factors that may have impacted retest passing rates—outside of TTM use—the evaluation team analyzed statistical models to determine whether passing the second administration STAAR-Mathematics tests was associated with TTM usage, while adjusting for other student attributes.

Among Grade 5 students who failed the first STAAR-Mathematics assessment, students who attempted 15 or more lessons in the period between the first and second administrations were significantly more likely to pass the retest compared to students who did not use the system. There was no statistically significant increase in the probability of passing STAAR upon retake among students with between 1 and 14 lessons attempted.

Among Grade 8 students who failed the first STAAR-Mathematics assessment, students who attempted between 5 and 9 lessons during the period between the first and second administration were significantly

---

[9] Similar to the limitation expressed for Istation, It is important to recognize that students who fail the first administration of the STAAR assessment are provided with a wide array of intensive mathematics interventions which vary by school district. Therefore, in addition to TTM, a variety of factors and interventions may be contributing to student performance on the second administration of the STAAR-mathematics assessment.

more likely to pass the retest compared to students who did not use the system. There was no statistically significant increase in the probability of passing STAAR-Mathematics upon retake demonstrated among students with more than 15 lessons attempted.

Study Limitations

It is critical to consider the following important caveats related to this study's methodology when considering implications of the results discussed above, and as presented in the rest of this report:

Non-Random assignment of students to the SUCCESS interventions: Through Texas SUCCESS, all public schools in the state had access to Istation and TTM. While this meant that all schools had the benefit of access to these programs designed to support teaching and learning, it also meant that there was not a group of students who *did not* have access to the systems, whose reading and mathematics achievement could be compared to students who *did* have such access. In other words, all schools' access to the systems prevented the evaluation team from comparing reading and mathematics outcomes from a *treatment group*, or students who had access to the programs, to a *control group*, or students who did not. If this condition had been a part of SUCCESS implementation—particularly if students had been randomly assigned to treatment or control groups—it would have been possible to say that the two groups were statistically equal at the start of the program. In this case, any differences in their achievement afterward would be attributable to the one condition that differentiated the groups (i.e., whether they had access to SUCCESS interventions or not).

Because students were not randomly assigned to participate in either Istation or TTM, a key challenge in this evaluation was to use the next best analytic and methodological strategy to estimate the effects of Texas SUCCESS. Since there were many differences between students who used Istation and TTM—besides just whether they used the programs or not—the evaluation team used statistical approaches to control for those differences as much as possible when determining the influence of these programs on reading and mathematics achievement.

In some ways, these efforts allowed evaluators to approximate conditions of random assignment; however, they also relied on the assumption that controls used in the statistical analyses captured the important differences between students who used the programs and those who did not. Unfortunately, because all factors that influence student achievement cannot be measured, it is impossible to test the extent to which this assumption actually holds true. Ultimately, this threatens the "internal validity" of the findings—that is, the confidence that the reported effect of program participation on student achievement represents the *true* effect of the program. What can be said, then, is that the reported estimates of program effects represent the influence of Texas SUCCESS interventions on student achievement, after many other observable factors that also influence student achievement have been taken into account.

Unmeasured teacher quality: The research team did not have access to information about the teachers to whom students were assigned during the period of this evaluation. This is a source of potentially omitted bias, because system usage and usage intensity may be related to teacher quality, or other important attributes of teachers, classroom activities, or contextual features of schools and districts. For instance, if students who were assigned to less effective teachers were also more likely to use either Texas SUCCESS program, lower student test score gains among those students could be attributed to their use of Istation or TTM, rather than to the fact that they may have received poor instruction. This is but one example supporting the notion that—given a lack of information on teacher quality—caution should be used when attributing achievement outcomes to Texas SUCCESS programs.

Missing information about the types of supplemental instruction or interventions students received: Schools and districts implement a plethora of interventions and supplementary services to improve their students' academic outcomes. The research team did not have any systematic information on the other types of supplementary instruction or services participants and non-participants received. This is important, particularly because the assumption underpinning the research design and multivariate analyses is that the difference in outcomes between participant students and non-participants represents the difference between students who use a Texas SUCCESS program compared to students under the "business as usual" condition, or those students who received the typical assortment of program supports and interventions that were available to students who were not Texas SUCCESS participants. This assumption may not hold if, for instance, students who were assigned to use a Texas SUCCESS program were also given a number of other interventions that may have neutralized, or complemented, the effect of either Texas SUCCESS intervention on student performance. Thus, the estimate of the effect of program participation may be impacted by a number of other interventions that are unmeasured in the evaluation.

Unmeasured differences between participating and non-participating students: Despite best efforts, including comparing within-student changes in performance between participating and non-participating students while controlling for other fixed and varying student-level characteristics, supplementing this design with propensity score reweighting based on observable characteristics, and confining the analytic sample to campuses with registered students, no guarantee can be made that participants and non-participants are identical with the exception of their exposure to the SUCCESS program. This is a fundamental, and unavoidable, challenge confronting any attempts to draw inferences about the effect of a social phenomenon (such as an academic intervention) using observational data where students were not randomized to receive, or not receive, treatment. If these unmeasured, or omitted, factors are correlated with program participation or the outcome, the estimates of the effect of program intervention are biased. See Gelman and Hill (2007) and Angrist and Pischke (2009) for accessible discussions of this source of bias.

Error in the measure of student participation in Texas SUCCESS during the 2012-13 school year: Program participation and usage data were obtained from both Istation and TTM for the 2012-13 school year. However, school district staff were not required to use unique student identification numbers for students who were uploaded to each vendor's registration system until the 2013-14 school year. Consequently, the match rate between TEA administrative records and the registration and usage information from each vendor was weaker in 2012-13 compared to 2013-14, and it varied systematically between vendors and across grade levels. Thus, students who participated in 2012-13 but who did not have a unique student identification number in the Istation and TTM systems would not be identified as having participated in

2012-13. This measurement error will produce attenuation bias in the estimates of the effect of 2012-13 Istation participation on the outcome.[10]

Imprecision in Istation dosage measure: Exposure to, and utilization of, TTM was manifested in the number of lessons a student attempted and passed in a defined period of time. This measure directly quantifies students' exposure to the content and assessments that comprise lessons within the system with a great deal of precision. The dosage metric for Istation, however, is less precise because it was not possible to determine *what occurs* and *how a student performs* within or across curriculum sessions. For instance, some students, even after adjusting for prior academic performance and other observable characteristics, may move more slowly through the curriculum. This conflates system *usage* or *dosage* with a number of other student-level characteristics that may also be correlated with student test performance, including their familiarity and comfort with computers and online programs, their general level of engagement or disengagement, classroom distractions, or inattentive or busy teachers who are not able provide assistance quickly to help struggling students. All of these intrinsic and extrinsic factors may contribute to increased time spent in the system and may be confounded with student test performance.

[10] This was a larger issue for TTM than for Istation.

Section 1 – Introduction

Background

Texas Student Success Initiative History

The Student Success Initiative (SSI), enacted in 1999 during the 76th Legislative Session through the passage of Senate Bill 4, provided the legislative framework to ensure that all students in Texas receive the instruction and support that they require to be academically successful in reading and mathematics at grade level. The early SSI legislation created research-based diagnostic assessments such as the Texas Primary Reading Inventory, and its Spanish equivalent, Tejas Lectura en Español (Tejas LEE), to determine students' progress toward K-2 reading standards. It also created high-quality Professional Development (PD) Academies supported by teacher stipends to ensure that K-3 teachers were knowledgeable about scientifically-based reading strategies and scientifically validated instructional practices, and that Grade 5-6 and Grade 7-8 teachers were knowledgeable about best practices in mathematics instruction. SSI provided additional funding for school districts to provide the necessary resources and supports for students struggling in reading and mathematics through the Accelerated Reading Instruction and Accelerated Mathematics Instruction (ARI/AMI) programs.

The initial SSI legislation required that TEA implement requirements that students meet the following standards to qualify for promotion to the next grade. These requirements were phased-in beginning with the first cohort of students entering kindergarten during the 1999-2000 school year. The requirements of the initial legislation required that students:

- Pass Grade 3 Texas Assessment of Knowledge and Skills (TAKS) in reading to be promoted to Grade 4 – first applied to the Grade 3 class of 2002-03.[11]
- Pass Grade 5 TAKS in reading and mathematics to be promoted to Grade 6 – first applied to the Grade 5 class of 2004-05.
- Pass Grade 8 TAKS in reading and mathematics to be promoted to Grade 9 – first applied to the Grade 8 class of 2007-08.

Programs and standards developed under SSI were designed and implemented to support that first cohort of students entering kindergarten in 1999-2000, who were then impacted by changes in grade promotion standards beginning in spring 2003 with the first administration of the TAKS. Thus, the first group of students for which new grade promotion standards applied was the Grade 3 class of 2002-03. District support (i.e., ARI/AMI funding) and teacher PD were designed to follow that first cohort of students and the subsequent cohorts of students. Thus, programs impacted kindergarten students and their teachers

[11] The promotion requirements for Grade 3 students were removed through the passage of House Bill 3 during the 81st Legislative Session in 2009.

in 1999-2000, kindergarten and Grade 1 students and teachers in 2000-01, kindergarten and Grades 1 and 2 students and teachers in 2001-02, and so on.

Because of the timing of the implementation of programs and standards, it was expected that Senate Bill 4 passed by the 76th Legislature was only the beginning of sweeping changes. The SSI provided an umbrella under which additional funding streams and academic programs would seek to meet its goals over time. Over the ensuing years, SSI funding was continued and further expanded through subsequent legislation) that created and funded programs designed to assist students at risk of not meeting state standards in reading and mathematics. The SSI riders included in the biennial appropriation bills (2001 – 2007) also represented a funding stream that has been used since 1999 to accomplish goals laid out that year.[12] The majority of the SSI funding provided to Texas school districts over the 2001 – 2007 period was distributed through formula-funded (i.e., based on the number of students failing the prior year state assessment) ARI/AMI grant programs. The purpose of those grants was to provide districts with additional funding and resources to provide targeted interventions to students struggling to master the grade-level content in reading and/or mathematics.

The Texas SUCCESS program is the latest SSI-related offering designed to help students advance their mathematics and reading skills so that they can meet grade promotion requirements for Grade 5 and Grade 8.

Texas SUCCESS Program

The Texas SUCCESS program offers state-funded access to interactive mathematics and reading programs provided by two vendors; Istation Reading (Istation) and Think Through Math (TTM) to all Texas public school students in Grades 3-8.[13] These free interactive programs are designed to be accessible 24/7 to students both in and out of school. Education service center, Region 20 (ESC 20) served as the SSI Support Center for Texas SUCCESS. The reported enrollment numbers reported by TEA for 2013-14 included 2,161,923 students for Istation and 1,912,062 students for TTM. When these online programs were selected by TEA for the Texas SUCCESS initiative, the correlation between Istation and TTM content and the Texas Essential Knowledge and Skills (TEKS) for reading and mathematics wasdetermined and documented.[14]

Istation and TTM are computer adaptive programs designed to support student achievement by adjusting content based on student skill level and incorporating assessments that track changes in student

---

[12] For further detail on the history of the SSI, See Texas Success Initiative: 2009-10 Biennium Evaluation Report. http://tea.texas.gov/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=2147495699&libID=2147495696

[13] For further detail related to the Texas SUCCESS Initiative, refer to the following TEA webpage: http://texassuccess.org/.

[14] For further detail related to the correlation between Istation and TTM content and the Texas Essential Knowledge and Skills (TEKS) for reading and mathematics, refer to the following Texas SUCCESS Initiative resources: http://texassuccess.org/reading/get_to_know_istation and http://texassuccess.org/wp-content/uploads/2014/05/TTM_TX_2014-15LessonAlignmentTEKS_040414.pdf

performance. Both programs provide reports to educators and parents that demonstrate student progress. Telephone and online support is available for both programs. Students using TTM are provided access to live mathematics tutors. Parts of the TTM curriculum are available in Spanish. Although Spanish versions are not prescribed by Texas SUCCESS, Istation is available in Spanish for some grade levels (K-3) and the benchmark test for this program can also be completed in Spanish if districts choose to purchase it.[15] Texas SUCCESS was marketed to districts, staff, students, and parents through communications provided by TEA, the ESC 20 SSI Support Center for Texas SUCCESS, the vendors themselves, and through the ESCs. ESC 20 provided coordination and technical assistance to all stakeholders. Outreach began in August of 2012 and continued throughout the school year. In addition to providing information about Texas SUCCESS; the ESC 20 SSI Support Center and the vendors provided professional training in the use of the programs directly at regional ESCs across the state and also via webinars, online interactive program guides, and online support. Districts and campuses were encouraged to implement Istation and TTM in ways that they thought would be of most benefit to students, thus the programs were utilized in many different formats. For example, campuses have included the programs in their regular curriculum as after- or before-school offerings, as supplementary or remedial material, and as homework.

Evaluation of the Texas SUCCESS Initiative

In fulfillment of Rider 50 (General Appropriations Act, Article III, 83rd Texas Legislature), TEA entered into a contract with Gibson Consulting Group, Inc. (Gibson) in February 2014 to conduct a comprehensive evaluation of the Texas SUCCESS Initiative. The evaluation focused on the implementation and impact of two online curriculum programs, Istation and TTM, on student academic performance outcomes related to reading and mathematics for students in Grades 3-8. Gibson, in partnership with Shore Research (Shore), employed a mixed-methods approach to evaluating the Texas SUCCESS program implemented in school districts across the state.

Research Questions

The study addressed 10 key research questions related to reading and mathematics broadly, and to the Istation Reading and TTM programs specifically. The following four implementation-related research questions are explored in this report separately for reading and mathematics:

1) What are the program utilization rates across districts and campuses in Texas for the Istation and TTM programs, and to what extent does system utilization vary by school characteristics and geographic regions (i.e., Title 1 school designation, ESC region, urbanicity, state accountability rating)?
2) To what extent is student usage within schools and districts reaching a level of fidelity with the intended implementation model, and how does it vary across various student groups (e.g., past academic performance, race/ethnicity, gender, English Language Learner status)?

[15] However, school districts cannot purchase the benchmark test by grade level.

3) To what extent are program supports available through the online system vendors, and ESC 20 and other service centers, being utilized by participating districts and campuses?
4) How are school districts and campuses using Istation to address the SSI grade promotion requirements for reading and mathematics, and what other academic resources are being used in combination with Texas SUCCESS programs to further support the learning needs of students to meet the SSI grade promotion requirements?

In addition, the following six student outcomes-related research questions are addressed in this report for reading and mathematics:

5) To what extent are students participating in the Texas SUCCESS program showing meaningful progress as evidenced by performance data contained in the Istation and TTM online programs, and to what extent is program implementation fidelity related to student growth on the Istation and TTM online programs?
6) What is the relationship between broader implementation of Texas SUCCESS and other SSI interventions, and student outcomes for reading and mathematics?
7) To what extent do student performance results differ for students participating in the Texas SUCCESS program (i.e., using Istation and TTM) and non-participating students, and how do results differ by student characteristics, prior academic performance, and other key variables?
8) What is the relationship between the timing of system usage (i.e., the proximity to the State of Texas Assessments of Academic Readiness (STAAR®)) and Istation/TTM usage levels, and growth on the first administration of the STAAR exam?
9) To what extent does performance differ between Texas SUCCESS participants (i.e., those using Istation/TTM) and non-participants among students at risk of being retained?
10) What is the relationship between program usage after failing the first administration of the STAAR assessment and the probability of passing subsequent administrations of the STAAR exam for reading and mathematics?

Analytic Methods

The evaluation focused on results for the 2013-14 school year in terms of program implementation and system usage, as well as student outcomes on the spring 2014 STAAR assessments for mathematics and reading.

The evaluation plan incorporated in-depth examinations of the Texas SUCCESS program through a number of different data sources, including:

- Online system usage and student growth data provided by the Istation and TTM vendors;[16]
- Istation and TTM professional development attendance and technical assistance usage data obtained through the Istation and TTM vendors, and ESC 20;
- Interviews conducted by the evaluation team with district and campus academic intervention staff in local education agencies using the programs;[17]
- Archival data on students contained in the Public Education Information Management System (PEIMS) from TEA; and
- Student outcomes data related to 2013-14 STAAR performance (reading and mathematics) from TEA.

To answer questions related to program implementation and system usage, the research team relied on system usage data obtained from the Istation and TTM vendors (matched with PEIMS data from TEA); data collected from interviews with campus- and district-level staff responsible for coordinating reading and mathematics interventions for their respective campuses and districts; and professional development attendance and technical assistance data to assess the extent to which district and campus staff were taking advantage of the supports that were made available to them. Descriptive analyses were conducted to address the implementation-related research questions.

To address the student outcomes questions, the evaluation team utilized system usage and growth/performance data contained in the Istation and TTM systems (matched with PEIMS and STAAR outcomes data) and conducted descriptive analyses and analyses utilizing a series of statistical models to measure the relationship between Istation and TTM system usage and growth on the STAAR assessment scores between 2012-13 and 2013-14. These statistical models controlled for demographic, socioeconomic, and pre-existing academic performance differences between students using the Istation and TTM programs and nonusers included in the comparison group. Please refer to Appendix A for more detail on the research methods related to the reading analyses and to Appendix B for more detail related to the mathematics analyses.

Organization of the Report

Following this introductory section, Section 2 of this report explores each of the ten research questions related to reading (i.e., Istation) and Section 3 examines the same set of research questions for mathematics (i.e., TTM). The reading and mathematics outcomes sections first address the four implementation/usage research questions (outlined above) before turning to the six research questions related to how Istation and TTM usage may be related to changes in student outcomes in their respective

[16] For more detail on Istation and TTM data processing, refer to Appendices A and B, respectively.

[17] See Appendix C for further detail related to the sampling approach utilized for campus- and district-level interviews, as well as fielding results.

subject areas. For each research question, a bulleted list of key findings is provided at the beginning of the section, and a more detailed summary of findings is provided at the end of each research question.

Appendix A contains detailed research methods related to the Istation analyses. Appendix B contains detailed research methods related to the TTM analyses. Appendix C provides additional methodological detail related to the interviews conducted with campus- and district-level staff. Appendix D contains supplementary Istation and TTM system usage tables related to school and student group disaggregations explored in research questions 1 and 2. Appendix E provides a list of study limitations to keep in mind when interpreting the results of this evaluation. Appendices F and G contain the campus staff and district staff interview protocols, respectively. Finally, Appendix H includes references included in this report.

Section 2 – Reading Outcomes

This section of the report is organized into two distinct parts: 1) implementation of Istation and how it is coordinated with other SSI-related reading interventions; and 2) the relationship between Istation usage and student growth on the Istation Indicators of Progress (ISIP) assessment and STAAR-Reading scores. A broader view of SSI as it relates to reading is also examined in this section.

Implementation of Istation and other SSI-related Reading Interventions

The fidelity of Istation implementation can be assessed a variety of ways. The evaluation team has chosen to take a broad view of implementation and explore it in the following ways:

- Students usage patterns (and how usage levels correspond with recommended minimum usage levels established by the vendor);
- District and campus participation in professional development and training required to use the Istation system more effectively;
- Ways in which various districts and campuses reported using the Istation program and other SSI-related reading interventions to enable students to meet the grade promotion requirements for Grades 5 and 8;
- Perspectives of practitioners regarding fidelity of implementation and satisfaction with the Istation program.

The analysis of Istation program implementation serves to address the following four guiding research questions:[18]

Guiding Question 1a: What are the program utilization rates across districts and campuses in Texas for the Istation programs, and to what extent does Istation utilization vary by school characteristics and geographic regions (i.e., Title 1 school designation, ESC region, urbanicity, state accountability rating)?

Guiding Question 2a: To what extent is student usage within schools and districts reaching a level of fidelity with the intended implementation model, and how does it vary across various student groups (e.g., past academic performance, race/ethnicity, gender, English Language Learner status)?

Guiding Question 3a: To what extent are program supports, available through the online system vendors, and ESC 20 and other service centers, being utilized by participating districts and campuses?

Guiding Question 4a: How are school districts and campuses using Istation to address the SSI grade promotion requirements for reading/English Language Arts (ELA), and what other reading resources are being used in combination with Texas SUCCESS programs to further support the learning needs of students to meet the SSI grade promotion requirements?

[18] Please note that parallel research questions for TTM are labeled 1b through 10b.

First, the evaluation team examined Istation system usage patterns at the student level and disaggregated usage results by grade level and the following school-level characteristics:

1) Title I campus
2) State accountability ratings
3) Status as *Priority* or *Focus* school[19]
4) Geographic region of the state (i.e., 20 ESC regions)
5) Urbanicity (e.g., urban, rural, suburban classifications)

The evaluation team also explored Istation system usage patterns by the following student groups:

- Performance on prior year STAAR-Reading assessment
- Race/ethnicity
- Gender
- English Language Learners (ELL) status

These analyses include assessments of the proportion of students using the systems at levels that approach a certain degree of fidelity (i.e., 200 minutes of Istation curriculum session usage per year for middle school students and 250 minutes per year for elementary school students). These recommended usage levels were determined by the Istation vendor.

Second, the evaluation team explored the extent to which staff at school districts and campuses across the state have taken advantage of training and technical assistance provided by the Istation vendor (either onsite or through sessions held at regional ESCs) or technical assistance provided through ESC 20. While data were limited, the results shed light on the extent to which Istation training was made available to district and campus staff across the state and the extent to which practitioners took advantage of this training. Data were provided to Gibson by TEA through information collected from Texas SUCCESS vendors and ESC 20.

Third, the evaluation team conducted interviews in spring 2014 with campus and district staff regarding the manner in which the Istation program was implemented at their schools, including their perspectives on the extent to which the program was implemented with fidelity. During the course of these interviews with district- and campus-level reading interventionists, the evaluation team captured additional information regarding how districts and campuses are using the Texas SUCCESS programs in conjunction with other reading and mathematics interventions to help their students meet the SSI grade promotion requirements for Grades 5 and 8.

[19] *Priority* schools represent Texas' lowest performing schools and account for at least 5% of the state's Title I schools. *Focus* schools account for at least 10%of the Title I schools in Texas and include those with the widest achievement gaps.

Findings for each of the four guiding research questions related to program implementation and system usage are presented below. Key findings are presented first for each guiding research question.

Guiding Question 1a:

What are the program utilization rates across districts and campuses in Texas for the Istation programs, and to what extent does Istation utilization vary by school characteristics and geographic regions (i.e., Title 1 school designation, ESC region, urbanicity, state accountability rating, Focus/Priority school status)?

Key Findings:

- **Over half of students in Grades 3-8 across the state logged into the Istation system and attempted at least one curriculum session; however, utilization at the vendor-recommended number of minutes across the 2013-14 school year was relatively low.** Of the approximately 2.2 million students in Grades 3-8, only 13% never registered to use the Istation system. Thirty-two percent were registered to use the system but never logged a curriculum session, and the remaining 55% recorded time in an Istation curriculum session. However, only 21% of students across the state used the Istation system at the recommended level of minutes (i.e., 250 minutes for elementary school grades and 200 minutes for middle school grades).
- **Students in elementary schools grades (Grades 3-5) used Istation at substantially higher rates than their middle school counterparts (Grades 6-8).** Usage was considerably higher for students in elementary school grades (67% to 71%) than in Grades 6-8 (46%, 40%, and 35%, respectively).
- **Through the Texas SUCCESS Initiative, Istation appears to be serving students in the most need of academic assistance in reading.**
  - ✓ **Campuses with higher proportions of students identified as economically disadvantaged (i.e., Title I campuses) used Istation much more frequently than their counterparts at non-Title I campuses.** Over a quarter of students in Title I schools used the system at or above the recommended threshold levels, compared to 12% of students at non-Title I schools. And, on average, students in Title I schools spent over twice as much time using Istation (202 minutes) than students at non-Title I schools (94 minutes).
  - ✓ **Students in lower performing campuses had higher Istation system utilization rates than students enrolled at campuses designated as *Met Standard*.** On average, students at campuses designated as *Improvement Required* logged 241 minutes of Istation curriculum session time, versus 165 minutes of use among students at campuses designated as *Met Standard*.

- ✓ **Substantive differences in Istation usage patterns were observed in different areas of the state, indicating that additional outreach could be beneficial to some regions of Texas.** Regions 17 and 16 (Lubbock and Amarillo) had the largest proportions of students who used Istation at recommended levels (i.e., 250 minutes for elementary and 200 minutes for middle school students). Regions with the lowest usage were Regions 13 and 6 (Austin and Huntsville).

Of particular interest is the extent to which the more than two million students in Grades 3-8 were engaged in the Istation online program during the 2013-14 school year and used the system for a substantive number of minutes. Usage statistics reported for Istation include Grades 3-8 students who did not register to use the system, who registered but did not use the system, and Istation users. This section addresses the extent to which Istation system usage rates varied across students attending different types of schools (i.e., Title I schools, schools in different geographic regions of the state, rural/urban/ suburban schools, and schools that did or did not meet state accountability standards in 2012-13). That is, are students who are enrolled at schools with high percentages of students identified as economically disadvantaged and students enrolled at persistently low performing schools more or less inclined to use the Istation program than students at other schools in Texas? In addition, exploring regional usage differences is important to ensure that access is uniform across the state, and may identify opportunities for additional outreach to areas of Texas where Istation usage is lower than expected.

Istation provided the evaluation team session-level records for every student who registered for the system and, if they used the system at all, their session history. The session-level records were collected at the day-level, so the research team could determine when, how many times, and how long a student used the system within a given span of time. A high percentage (over 98%) of student records contained in the Istation Reading tables could be reliably joined to TEA administrative and assessment records, which allowed for Istation system usage and assessment performance data to be linked to TEA accountability and assessment records, including student attendance, economic disadvantaged status, and prior-year STAAR assessment performance.[20] In this section, basic descriptive information about the frequency with which the system was used throughout the 2013-14 school year is provided. The focus is exclusively on sessions that were identified as *Curriculum*, as opposed to assessment sessions, meaning the student encountered curriculum and activities directly related to content aligned with Texas Essential Knowledge and Skills (TEKS) standards.[21] The research team did not have any information about how a student performed within a given lesson, or the types or numbers of problems a student attempted or solved. The Istation system does not directly collect information about how students perform on particular lessons.[22]

---

[20] For further detail on the processing of Istation data, refer to Appendix A.

[21] A student was categorized as having recorded a "curriculum activity" if they logged into a curriculum session for any amount of time.

[22] For more detail related to this study limitation, see Appendix E.

Table 2.1 provides the number and percentage of students in the state who: 1) did not register to use the Istation system (and did not attempt a curriculum session); 2) registered in the Istation system but did not attempt a curriculum session;[23] and 3) registered in the Istation system and attempted at least one curriculum session during the 2013-14 school year. [24] This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. Data are further disaggregated by grade level.

Higher grade levels have higher proportions of students having never registered with the Istation system, and lower conversion rates (i.e., registering and then using the Istation system), with the lowest conversion rate being Grade 8 students (35%), and the highest being Grade 3 students (71%). Depending upon grade level, 67% to 71% of students in Grades 3 through 5 who registered in the Istation system attempted at least one curriculum session, compared to just 35% to 46% of students in Grades 6, 7, and 8. There is a clear structural break in conversion rates between elementary and middle grades: the percentage of Grade 5 students who registered and participated in a curriculum session was 67%, 21 percentage points higher than Grade 6 (46%).

[23] System registration was derived from the full roster of unique student IDs in the Istation system during the 2013-14 school year. Students from this file who were not linked to the base system usage file described in Appendix A were classified as having registered, but not used, the system.

[24] Only session records that occurred between August 2013 and June 2014 were retained.

Table 2.1. Frequency and Percentage of Students Who Recorded at Least One Curriculum Session, by System Registration Status, 2013-14

Student Grade Level	Did Not Register		Registered, but Did Not Use Istation		Registered, and Used Istation		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Grade 3	38,306	9.82%	73,562	18.86%	278,109	71.31%	389,977	100.00%
Grade 4	35,483	9.25%	84,399	22.01%	263,657	68.74%	383,539	100.00%
Grade 5	37,165	9.71%	89,828	23.46%	255,899	66.83%	382,892	100.00%
Grade 6	56,829	15.09%	147,339	39.13%	172,390	45.78%	376,558	100.00%
Grade 7	67,421	17.49%	164,367	42.64%	153,677	39.87%	385,465	100.00%
Grade 8	69,214	18.23%	179,048	47.16%	131,418	34.61%	379,680	100.00%
Total	**304,418**	**13.25%**	**738,543**	**32.14%**	**1,255,150**	**54.62%**	**2,298,111**	**100.00%**

Source: Istation session-by-product table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. Students not attempting any curriculum session, and who were not registered in the system, are categorized as "Did Not Register." Students who registered, but did not log any curriculum sessions were included in the "Registered, but did not use" category. Students who attempted a curriculum session were categorized as "Registered, and Used Istation." Registration counts include only students who could be linked to the PEIMS 101 Fall 2013 Snapshot data during the 2013-14 school year.

As noted previously, Istation system usage results were further disaggregated by the school's Title I status, geographic region of the state, urbanicity, state accountability rating, and *Priority/Focus* school status. These groups were purposefully selected to determine whether persistently underperforming schools, and schools with larger percentages of students identified as economically disadvantaged, were using the system more intensively than other schools in Texas. In addition, geographic usage was important to examine whether any areas of the state were being underserved by the program. For each student group analysis, three metrics of system usage are presented: 1) percent of students using the system at recommended threshold levels; 2) percent of students who registered to use the Istation system and attempted a curriculum session; and 3) the mean number of minutes per student spent on Istation curriculum sessions. ***The recommended minimum Istation threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum Istation threshold usage level for Grades 6-8 is 200 minutes.***

The results of these descriptive analyses are presented below. Where available and applicable, qualitative information collected via interviews with campus-level staff is incorporated into descriptions of the summary statistics on Istation usage.

Title I Status

Title I schools enroll a higher percentage of students identified as economically disadvantaged than their non-Title I counterparts, and may be in need of additional resources to serve their student populations. As Table 2.2 shows, students enrolled in Title I campuses used the Istation system much more frequently than their counterparts at non-Title I schools. Over a quarter (25%) of students at Title I schools used the system at or above the recommended threshold levels (compared to 12% of students at non-Title I schools) and 60% of students at Title I schools were registered to use the system and actually attempted at least one Istation curriculum session (compared to just 42% of non-Title I students). Similarly, on average, students at Title I schools logged more than twice as many minutes in Istation curriculum sessions (202 minutes) as students at non-Title I schools (94 minutes).

Table 2.2. Istation System Usage Disaggregated by Campus Title I Status, 2013-14

Campus Title I Status	Number of Students	Percent of Students Using System at or Above Recommended Threshold Levels	Percent of Students Registered to Use Istation Who Attempted a Curriculum Session	Mean Number of Total Minutes per Student Spent on Curriculum Sessions
Not Title 1	661,738	12.45%	41.59%	94
Title 1	1,627,193	25.10%	60.11%	202
Total	2,288,931	21.44%	54.76%	170

Source: Istation session-by-product table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The recommended minimum Istation threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum Istation threshold usage level for Grades 6-8 is 200 minutes.

To gain a better understanding of how usage levels varied by grade levels for students at Title I and non-Title I schools, usage was calculated by the three primary metrics of interest. For all grades, students at Title I schools utilized the Istation system more frequently than students at non-Title I schools, and elementary school students (Grades 3-5) used the system much more frequently than their middle school counterparts (see Table D1 in Appendix D).

Comments from campus-level staff provide some contextual detail regarding Istation usage at both Title 1 and non-Title 1 campuses. Of the 72 schools selected for the campus interview sample, more than 75% were Title 1 campuses. Keeping the composition of the interview sample in mind, there were several differences between Title 1 and non-Title 1 campuses in school staff members' reported usage and perceptions of the system. Campus-level staff were asked about the fidelity of Istation implementation on their campuses, and their level of satisfaction with various aspects of the program, including technical assistance provided by ESC 20. Responses to these questions were disaggregated by schools' Title 1 status to determine whether clear differences emerged that might provide insight into different levels of usage at campuses, by their Title 1 status:

- Title 1 campuses were somewhat more likely to report that the system was being implemented with fidelity (70% versus 56% for non-Title 1 campuses) and that they were satisfied with Istation content (67% versus 40% for non-Title 1 campuses).

- With regard to support provided by ESC 20, no staff members from non-Title 1 campuses reported being "very satisfied" with support from ESC 20, while 76% of Title 1 campuses reported that they were very satisfied with support from the ESC.

Meanwhile, the majority of staff from both types of campuses reported high levels of satisfaction with the overall user-friendliness of the program.

State Accountability Rating

To determine whether Istation system usage was associated with prior year (2012-13) campus academic performance as measured by the state accountability rating, usage statistics were calculated for schools designated as *Improvement Required* or *Met Standard*. As Table 2.3 shows, students at lower performing campuses (i.e. those designated as *Improvement Required*) had higher Istation system utilization rates than students at campuses designated as *Met Standard*. On average, students at *Improvement Required* campuses logged 241 minutes of Istation curriculum session time compared to 165 minutes for students at *Met Standard* campuses.

Table 2.3. Istation System Usage Disaggregated by Campus Accountability Rating, 2013-14

2012-13 Accountability Rating	Number of Students	Percent of Students Using System at Recommended Threshold Levels	Percent of Students Registered to Use Istation Who Attempted a Curriculum Session	Mean Number of Total Minutes per Student Spent on Curriculum Sessions
Improvement Required	163,461	30.01%	64.92%	241
Met Standard	2,103,907	20.79%	54.09%	165
Total	**2,267,368**	**21.46%**	**54.87%**	**171**

Source: Istation session-by-product table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The recommended minimum Istation threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum Istation threshold usage level for Grades 6-8 is 200 minutes.

For all grades, students at campuses with 2012-13 *Improvement Required* state accountability ratings used the Istation system more frequently than their counterparts at schools with *Met Standard* ratings (See Table D2 in Appendix D). Little difference in system usage was observed among students at *Focus* (261 minutes) and *Priority* (284 minutes) schools in terms of average minutes logged into curriculum sessions. This finding persists across all grade levels, with student usage levels slightly higher at *Priority* campuses in elementary school grades, and slightly higher at *Focus* schools in middle school grades. (See Table D3 in Appendix D.)

Geographic Region

As Table 2.4 shows, there were sizable differences in Istation system usage patterns for students in different areas of the state. Regions with the largest proportions of students using the system at recommended threshold levels of 250 minutes for elementary grade students and 200 minutes for middle school students include Region 17 (Lubbock, 35%) and Region 16 (Amarillo, 33%); while the lowest included Region 13 (Austin, 12%) and Region 6 (Huntsville, 13%).

Regions of the state with the highest rates of students registered in the Istation system who attempted a curriculum session include Region 19 (El Paso, 77%), which also had 32% of its students completing the recommended minutes on the Istation system, Region 16 (Midland, 65%) and Region 18 (Amarillo, 64%); while the lowest rates were recorded in the Huntsville (34%) and Austin (34%) regions. The mean number of minutes students spent in Istation curriculum session during the 2013-14 school year was highest in Lubbock (323 minutes) and Region 15 (San Angelo, 274 minutes), and lowest in Austin (89 minutes) and Huntsville (93 minutes).

Table 2.4. Istation System Usage Disaggregated by Geographic Region, 2013-14

ESC Region	Number of Students	Percent of Students Using System at Recommended Threshold Levels	Percent of Students Registered to Use Istation Who Attempted a Curriculum Session	Mean Number of Total Minutes per Student Spent on Curriculum Sessions
1 (Edinburg)	186,708	18.11%	51.26%	136
2 (Corpus Christi)	48,112	32.07%	58.90%	246
3 (Victoria)	23,995	22.78%	48.47%	173
4 (Houston)	509,677	17.93%	49.96%	146
5 (Beaumont)	35,588	20.52%	51.60%	150
6 (Huntsville)	81,687	12.78%	33.57%	93
7 (Kilgore)	76,170	17.46%	46.65%	137
8 (Mount Pleasant)	24,517	20.42%	43.98%	163
9 (Wichita Falls)	16,735	16.07%	52.91%	121
10 (Richardson)	357,496	23.32%	63.51%	195
11 (Fort Worth)	255,983	23.78%	63.36%	176
12 (Waco)	71,415	19.12%	47.24%	137
13 (Austin)	175,414	12.07%	34.45%	89
14 (Abilene)	26,687	26.74%	55.37%	226
15 (San Angelo)	21,682	30.22%	54.10%	272
16 (Amarillo)	37,816	32.52%	64.37%	244
17 (Lubbock)	37,024	35.01%	57.51%	323
18 (Midland)	38,549	32.45%	65.12%	252
19 (El Paso)	78,404	31.83%	76.70%	245
20 (San Antonio)	194,332	26.32%	63.35%	224
Total	**2,297,991**	**21.39%**	**54.62%**	**170**

Source: Istation session-by-product table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The recommended minimum Istation threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum Istation threshold usage level for Grades 6-8 is 200 minutes.

Urbanicity

Minor variation in Istation system usage rates were observed when urbanicity was considered. Data were disaggregated by eight categories captured by TEA: 1) Charter School; 2) Independent Town; 3) Major Suburban; 4) Major Urban; 5) Non-metropolitan Fast Growing; 6) Non-metropolitan Stable Growth; 7) Other Central City; 8) Other Central City Suburban; and 9) Rural.

Students enrolled in schools designated as being from Major Urban (26% at recommended threshold levels) and Independent Towns (27% at recommended threshold levels) tended to use the Istation system at higher rates than other types of jurisdictions. The lowest Istation usage rates were recorded in Charter Schools (8% at recommended threshold levels), Non-metropolitan Fast Growing regions (13% at recommended threshold levels) and Major Suburban regions (19% at recommended threshold levels). (See Table D4 in Appendix D.)[25]

Guiding Research Question 1a: Summary of Findings

The vast majority (87%) of students in Grades 3-8 across the state were registered to use the Istation system; however, just 55% of students actually logged into an Istation curriculum session and only 21% used the system at recommended levels of minutes. Further, system usage patterns varied widely across grade levels, school characteristics, and student groups. Students in elementary school grades (Grades 3-5) used the system at much higher rates (67% to 71%) than their middle school counterparts (35% to 46%).

Disaggregated findings reveal that students at Title I and persistently low performing schools utilized the Istation program more frequently than their counterparts at non-Title I schools and higher performing schools (i.e., *Met Standard*). Key findings include:

- Students attending Title I schools (25%) were approximately twice as likely as those attending non-Title I schools (12%) to use the Istation program at recommended usage levels. Though not representative of the population of Istation users, responses from campus-level staff demonstrated slight differences between Title 1 and non-Title 1 campuses, with non-Title 1 campuses having slightly lower levels of satisfaction with system content and support from ESC 20. Title 1 campuses, conversely, were somewhat more likely to report using the system in the same manner across students and that their campuses were implementing the system with fidelity.
- Students attending schools rated as *Improvement Required* (30%) in 2012-13, had a higher percentage of students using the Istation system at recommended thresholds than students at schools designated as *Met Standard* (21%).
- Geographic differences in usage patterns were evident, with the highest usage rates recorded in Lubbock and Amarillo, and the lowest rates observed in Huntsville and Austin, indicating that targeted outreach efforts may be warranted for some regions of the state.
- Students attending school in Independent Towns (27%) and Major Urban Districts (26%) used the system at recommended usage levels more frequently than students in Charter School (8%), Fast Growing Non-Metropolitan (13%), Major Suburban (19%), and Rural (22%) school districts. Qualitative data revealed mostly similarities in campus-level staff members' responses across regions, though staff from lower usage, rural schools were proportionally more apt to report using the system differently across different types of students than staff from higher usage regions.

[25] For urbanicity definitions, please refer to the following link on the TEA website: http://tea.texas.gov/acctres/analyze/1213/gloss1213.html

Guiding Question 2a:

To what extent is student usage within schools and districts reaching a level of fidelity with the intended implementation model, and how does it vary across various student groups (i.e., past academic performance, race/ethnicity, gender, English Language Learner status)?

Key Findings:

- **Minimal Istation usage was the most common pattern among all students in Grades 3-8 in Texas.** Overall, less than a quarter of students in Grades 3-8 used Istation for 200 or more minutes during the course of the 2013-14 school year. Fifty-five percent of students logged into curriculum sessions for zero to 19 minutes and 22% logged on for 20 to 199 minutes.
- **Elementary school students (Grades 3-5) utilized the Istation system at recommended levels considerably more often than middle school students (Grades 6-8).** While approximately 33% of students in Grades 3-5 met the recommended usage threshold of 250 minutes for elementary school students, approximately 10% of students in Grades 6-8 met the recommended school usage threshold of 200 minutes for middle school students in 2013-14. Grade level usage at recommended levels declined from 39% in Grade 3 to 29% in Grade 5, and to just 7% in Grade 8.
- **It appears that lower performing students and students classified as ELL were systematically identified for the Istation reading intervention more frequently than their higher performing and non-ELL counterparts.**
  - ✓ **Students who performed lower on the prior year STAAR assessment used the Istation system much more frequently than their higher performing peers, a pattern that was consistent across all grade levels.** Students in Grades 4 and 5 who scored in the bottom quartile of STAAR-Reading performance in 2012-13 logged almost twice as much time in Istation as students in the top quartile. While usage, in general, was substantially lower for middle school students, differences between students' Istation use by their prior reading performance was also noteworthy (e.g., in Grade 7, bottom quartile usage of 127 minutes versus top quartile usage of 23 minutes).
  - ✓ **While Istation system use was somewhat higher for students identified as ELL in elementary grades, the differences were much greater among middle school students.** Students identified as ELL used the online reading system almost twice as often as their non-ELL counterparts in Grade 6 (146 versus 87 minutes) and over twice as often as non-ELL students in Grade 7 (122 versus 58 minutes) and Grade 8 (101 versus 41 minutes). What is more, in Grade 8, students identified as ELL were nearly

three times as likely to use Istation at the recommended threshold (200 minutes) as students not classified as ELL (15% versus 6%).

- ✓ **Students previously retained in grade used Istation somewhat more frequently than non-retained students.** On average, in Grade 3, students who were retained in grade in 2012-13 spent over 40 more minutes in the system than students who were previously promoted to the next grade. The difference grew for students in Grade 4 (305 versus 251 minutes for retained versus promoted students, respectively) and Grade 5 (299 versus 224 minutes for retained versus promoted students, respectively). As the usage levels declined substantially in middle school grades, so did the differences in Istation system usage.

- **Not having an adequate number of computers and laptops available was cited as a barrier to Istation implementation.** Staff at most campuses (regardless of their Istation usage level) felt that they had adequate instructional and support staff, and internet connectivity, to implement the system; however, just over half noted that they had an adequate number of computers or laptops to effectively implement Istation. Computer shortages were noted as an issue more frequently by staff at campuses utilizing Istation at low levels and middle schools.

To address Research Question 2a related to fidelity of program usage and variation in usage by key student groups, the evaluation team explored usage patterns at various levels overall and for specific student subpopulations described below. Table 2.5 presents the number and proportion of students using the Istation curriculum activities by the following time intervals: 1) Zero to 19 minutes; 2) 20 to 199 minutes; 3) 200 to 249 minutes (i.e., the minimum recommended usage threshold for middle schools students); 4) 250 to 299 minutes (i.e., the minimum recommended usage threshold for elementary schools students); and 5) 300 or more minutes. To illustrate important differences in usage levels at the elementary and middle school levels, the data are further disaggregated by grade level in this table.

The calculations include students who were not registered for the Istation system (and did not log into a curriculum session), students who registered for the system and did not log into a curriculum session, and students who were registered and did log into at least one curriculum session.[26] In essence, the usage patterns described in this section reflect statewide usage of all students in Grades 3-8, not just system registrants or users. This approach was taken to provide the most accurate view of system usage and impact possible. As Table 2.5 illustrates, less than one quarter of the students used the Istation program for 200 minutes or more during the course of the 2013-14 school year.

[26] These calculations differ from Patarapichayatham (2014). These differences are attributable to the different selection criteria used in that study. The study limited the analytic sample to students in schools classified as "Good Implementation" campuses, which were campuses that met the minimum usage thresholds described above. This reduced the sample considerably, and the reductions in students included in these analyses was disproportionately severe for the middle grades. For instance, in that study, Grade 8 students comprise approximately 2% of the total sample, whereas in this study they comprise approximately 12%.

Stark usage differences were observed between elementary grade students in Grades 3-5 and middle school students in Grades 6-8. Approximately 35% of Grade 3 students, 28% of Grade 4 students, and 25% of Grade 5 students logged into Istation curriculum sessions for 300 minutes or more. This compared to only 10% of Grade 6 students, 6% of Grade 7 students, and 4% of Grade 8 students who used the system for more than 300 minutes. Similarly, between 65% and 79% of students in Grades 6-8 used the system for zero to 19 minutes over the course of the entire 2013-14 school year compared to between 34% and 42% of students in the elementary school grades.

Approximately one third (33%) of the students in Grades 3-5 met the recommended minimum usage threshold of 250 minutes for elementary school students in 2013-14. This includes 39% in Grade 3, 31% in Grade 4, and 29% in Grade 5. This compared to just 10% of students in Grades 6-8 who met the recommended middle school usage threshold of 200 minutes during the 2013-14 school year (14% in Grade 6, 9% in Grade 7, and 7% in Grade 8). This is an important finding because it demonstrates that a relatively small percentage of students were using the system at the threshold level recommended by the Istation vendor (especially at the middle school level), which may help to explain modest impact levels for the system.

Table 2.5. Percent of Students with Various Intervals of Minutes Logged in Istation Curriculum Sessions by Grade Level, 2013-14

Grade Level	0 - 19 Minutes	20 – 199 Minutes	200 – 249 Minutes	250-299 minutes	300+ minutes
Grade 3 (n=389,977)	33.88%	23.43%	4.12%	3.63%	34.95%
Grade 4 (n=383,539)	38.52%	26.41%	4.09%	3.46%	27.51%
Grade 5 (n=382,892)	41.77%	25.95%	3.76%	3.25%	25.26%
Grade 6 (n=376,558)	64.66%	21.22%	2.36%	1.80%	9.96%
Grade 7 (n=385,465)	73.09%	17.87%	1.49%	1.18%	6.37%
Grade 8 (n=379,680)	78.81%	14.60%	1.19%	0.95%	4.44%
Total (n=2,298,111)	**55.01%**	**21.59%**	**2.84%**	**2.39%**	**18.17%**

Source: Istation session-by-product table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. The calculations include students who were not registered for the Istation system (and did not log into a curriculum session), students who registered for the system and did not log into a curriculum session, and students who were registered and did log into at least one curriculum session. Percentages may add up to more than 100 because of rounding. The recommended minimum threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum threshold usage level for Grades 6-8 is 200 minutes.

Istation Usage Patterns Disaggregated by Student Groups

In an effort to better understand the characteristics of students using the Istation system across the State of Texas, Istation usage patterns were further disaggregated by a number of different characteristics, including:

- Students' prior year's performance on the STAAR-Reading exam (i.e., organized by the quartile of their STAAR scale score);
- Students' prior year grade retention status (i.e., retained in grade in 2012-13);
- Students' race/ethnicity;
- Student's gender; and
- Students' ELL status

For each group analysis, similar to the group analyses presented for Research Question #1, the following three metrics of system usage are presented: 1) percent of students using the system at recommended threshold levels; 2) percent of students who registered to use the Istation system and attempted a curriculum session; and 3) the mean number of minutes per student spent on Istation curriculum sessions. As noted previously, the recommended minimum Istation threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum Istation threshold usage level for Grades 6-8 is 200 minutes.

Prior Performance of STAAR-Reading

While data are commonly reported for Grades 3-8 in this report, the disaggregation of Istation usage by prior STAAR performance is reported for Grades 4-8 because prior year assessment data were not available for students in Grade 3 in 2013-14. For this analysis, each student is categorized into quartiles of performance on first administration of the 2012-13 STAAR-Reading exam.

This analysis is very important because it provides key information about whether students are meeting threshold usage levels and whether schools are identifying the students most in need of additional reading support for the online reading intervention. As Table 2.6 illustrates, there are substantive differences in 2013-14 Istation usage levels for students in the top performance quartile of the performance of the 2012-13 STAAR-Reading assessment compared to students in the bottom quartile of 2012-13 STAAR-Reading performance. Further, the difference is more extreme in the middle school grades. Grade 4 students in the bottom quartile of 2012-13 STAAR performance spent an average of 317 minutes working in Istation curriculum sessions (and 38% used the Istation system at recommended usage levels) compared to just 186 minutes (and 24% using the system at recommended usage levels) for students in the top quartile of prior year STAAR-Reading performance.

While usage in general is substantially lower for middle school students, the difference in mean curriculum minutes is vastly different for Grade 7 students (bottom quartile usage of 127 minutes versus top quartile usage of 23 minutes) and Grade 8 students (bottom quartile usage of 100 minutes versus top quartile usage of 13 minutes). A similar pattern holds for the proportion of students using the Istation curriculum system at recommended levels of 250 minutes for elementary school students and 200 minutes for middle school students. This finding strongly suggests that school districts and campuses are focusing their attention on struggling readers (i.e., those most at risk of not meeting state standards on the spring 2014 STAAR-Reading assessment) when deciding which students should be using the Istation system. Campus-level interview data support this finding, though readers should be cautioned that the interview sample represents only a small proportion of the population of schools using Istation in 2013-14: staff from 72 schools were interviewed regarding their schools' use of Istation. That caveat in place, the evaluation team explored responses among interviewees regarding the extent to which they targeted students based on prior STAAR performance, with larger proportions of middle school staff reporting that they targeted students for Istation based on students' prior STAAR performance than did elementary school staff (85% versus 39%, respectively).

Table 2.6. Istation System Usage Disaggregated by Prior Year (2012-13) STAAR-Reading Performance, 2013-14

Grade Level	Prior Year	Number of Students	Percent of Students Using System at Recommended Threshold Levels	Mean Number of Total Minutes per Student Spent on Curriculum Sessions
Grade 4	Bottom quartile	87,622	37.78%	317
Grade 4	Top quartile	84,295	24.43%	186
Grade 5	Bottom quartile	92,002	38.59%	315
Grade 5	Top quartile	83,990	19.74%	146
Grade 6	Bottom quartile	85,584	22.20%	155
Grade 6	Top quartile	84,415	7.88%	50
Grade 7	Bottom quartile	90,527	17.60%	127
Grade 7	Top quartile	84,158	3.10%	23
Grade 8	Bottom quartile	85,955	14.64%	100
Grade 8	Top quartile	71,188	1.62%	13

Source: Istation session-by-product table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 4-8 estimates of usage. The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. Students enrolled in Grade 3 in 2013-14 were not included in this analysis because they did not have prior year STAAR scores. The recommended minimum threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum threshold usage level for Grades 6-8 is 200 minutes.

Prior Year Grade Retention Status

Because SSI has historically been focused on struggling students in danger of not meeting grade-level promotion requirements, an assessment of the extent to which previously retained students are using the system at recommended levels is central to the evaluation. Counter to what may have been expected, whether a student had been retained in grade in the prior school year does not appear to have as large an impact on Istation system usage as prior year performance of the STAAR-Reading assessment. As Table 2.7 shows, the mean number of minutes spent on curriculum sessions during the 2013-14 school year was 378 for Grade 3 students who were retained in grade in 2012-13 versus 334 minutes for students who were promoted to the next grade (Grade 3). The difference grows for students in Grade 4 (305 minutes for retained students versus 251 minutes for students who were promoted) and Grade 5 (299 minutes for retained students versus 224 minutes for students who were promoted. While usage levels declined substantially in the middle school grades, so did the differences in Istation system usage between students promoted or retained in grade.

Table 2.7. Istation System Usage Disaggregated by Prior Year Grade Retention Status, 2013-14

Grade Level in 2013-14	Prior Year Grade Retention Status	Number of Students	Percent of Students Using System at Recommended Threshold Levels	Mean Number of Total Minutes per Student Spent on Curriculum Sessions
Grade 3	Promoted	364,809	38.65%	334
Grade 3	Retained	7,885	43.06%	378
Grade 4	Promoted	362,582	31.05%	251
Grade 4	Retained	4,373	37.02%	305
Grade 5	Promoted	361,852	28.54%	224
Grade 5	Retained	5,299	37.35%	299
Grade 6	Promoted	357,952	14.14%	95
Grade 6	Retained	2,827	15.25%	104
Grade 7	Promoted	365,592	9.00%	65
Grade 7	Retained	3,379	10.42%	76
Grade 8	Promoted	361,165	6.53%	46
Grade 8	Retained	3,616	7.52%	56

Source: Istation session-by-product table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The recommended minimum threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum threshold usage level for Grades 6-8 is 200 minutes.

Race/Ethnicity

For students in Grades 3-8, Hispanic and African-American students tended to be somewhat heavier users of the Istation system than their White counterparts. For example, in Grade 3, 44% of African-American students, 40% of Hispanic students, and 35% of White students spent 250 or more minutes in Istation curriculum sessions (i.e., the minimum threshold for recommended usage) during the 2013-14 school year. Similar patterns were observed in other grades, such as Grade 5 (African American-31%; Hispanic-32%; White-22%), and Grade 8 (African American-7%; Hispanic-8%; White-4%).

Gender

Only modest differences in Istation curriculum session usage were observed between male and female students during the 2013-14 school year, with males using the system slightly more often in each grade level (see Table D5 in Appendix D).

English Language Learner Status

As Table 2.8 illustrates, while Istation system usage is somewhat higher for students classified as ELL in elementary school grades, the differences are much more stark for middle school students where ELL students are using the online reading system almost twice as often as their non-ELL counterparts in Grade 6 (146 minutes versus 87 minutes) and over twice as often as non-ELL students in Grade 7 (122 minutes versus 58 minutes) and Grade 8 (101 minutes versus 41 minutes). Further, in Grade 8, students classified as ELL were nearly three times as likely to use the Istation system (at the recommended level of 200 minutes for the 2013-14 school year) than non-ELL students (15% versus 6%).

Table 2.8. Istation System Usage Disaggregated by ELL Status and Grade Level, 2013-14

Grade Level	ELL Status	Number of Students	Percent of Students Using System at Recommended Threshold Levels	Mean Number of Total Minutes per Student Spent on Curriculum Sessions
Grade 3	Non ELL	287,802	40.33%	344
Grade 3	ELL	102,175	33.65%	303
Grade 4	Non ELL	297,216	30.53%	243
Grade 4	ELL	86,323	32.50%	278
Grade 5	Non ELL	310,996	27.19%	210
Grade 5	ELL	71,896	34.25%	285
Grade 6	Non ELL	323,918	13.04%	87
Grade 6	ELL	52,640	20.79%	146
Grade 7	Non ELL	341,818	8.00%	58
Grade 7	ELL	43,647	17.27%	122
Grade 8	Non ELL	344,988	5.73%	41
Grade 8	ELL	34,692	15.05%	101

Source: Istation session-by-product table Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. The recommended minimum threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum threshold usage level for Grades 6-8 is 200 minutes.

Facilitators and Barriers to Implementation

In an attempt to gain a better understanding of the factors that may be impacting Istation system usage at campuses across the state, staff responsible for overseeing Istation and other reading interventions at their schools were interviewed in spring 2014. Nearly 80% of the interviewees noted some barriers to implementation. This finding was consistent across high, moderate, and low usage campus, as well as elementary and middle schools. Across all campuses, not having enough time in the daily schedule (45%) was the most frequently mentioned barrier to implementing Istation. Not enough computers was mentioned by 31% of the interviewees, and 29% indicated that lack of knowledge about the Istation program was a barrier to implementation.

The most commonly noted facilitators to implementing Istation at their campuses included training (18%), student motivation and enjoyment of using the Istation system (16%), monitoring student use (15%), and time available to use the system (15%). Consistent use of the system (13%), computer access/technology (12%), and buy-in to the use of the system (10%) were also mentioned by interviewees. Training was mentioned more frequently by low usage campuses, and student motivation and enjoyment was not mentioned by any of the low usage campus staff (while it was noted as a facilitator by 23% of high usage campuses and 20% of moderate usage campuses).

Supports for Implementation

Staffing to Support Istation Implementation

About one third (35%) of interviewees indicated that other than the primary staff member for the Istation session, no additional support staff were used to help implement the Istation program at their campuses. As Table 2.9 shows, about a third of campus staff interviewed (32%) indicated that educational aides were used to help support the Istation intervention, and another 32% said that reading interventionists were used to help support Istation use. A smaller proportion of respondents (8%) noted that computer lab staff were used to assist students with Istation. High Istation usage campuses tended to utilize educational aides more frequently than lower utilization campuses to assist students with Istation use.

Table 2.9 Percentage of Campus Staff Indicating Additional Istation Support Personnel by Usage and School Type

	Overall (n=64)	High Usage (n=25)	Moderate Usage (n=23)	Low Usage (n=16)	Elementary School (n=36)	Middle School (n=28)
Educational aides	32%	44%	22%	19%	39%	18%
Interventionists/Coaches	32%	20%	39%	38%	33%	29%
Computer lab staff	8%	8%	13%	0%	0%	18%

Source: Spring 2014 interviews with campus-level staff.

Note: Percentages do not total to 100% because only the most frequent responses to the question were reported in this table.

Adequate Resources to Support Istation Implementation

Interviewees were asked about whether their campuses had adequate resources to effectively implement the Istation program. High percentages of interviewees across all usage groups indicated that they had enough instructional staff (82% to 85%), internet connectivity (76% to 83%), and educational aides (64% to 71%) to implement Istation effectively. It appears that the lack of computers was an issue for campuses that used the program at minimal levels. Low Istation usage (41%) campuses were much less likely than moderate (56%) or high usage (62%) campuses to mention that they had enough computers to effectively implement the online reading program.

Elementary school staff were also much more likely than middle school staff to report that they had adequate numbers of educational aides (75% versus 59%), computers (61% versus 47%), and internet connectivity (85% versus 73%) to effectively implement Istation at their campuses. (See Table 2.10.)

Table 2.10. Percent of Campuses with Enough of the Following Resources to Effectively Implement Istation

	Overall (n=68)	High Usage (n=26)	Moderate Usage (n=25)	Low Usage (n=17)	Elementary School (n=36)	Middle School (n=32)
Instructional staff	84%	85%	84%	82%	86%	81%
Educational aides	68%	64%	71%	67%	75%	59%
Computers/Laptops	55%	62%	56%	41%	61%	47%
Internet connectivity	80%	78%	83%	76%	85%	73%

Source: Spring 2014 interviews with campus-level staff.

Guiding Research Question 2a: Summary of Findings

Overall, minimal usage was the most common Istation usage pattern. Over half (55%) of all students in Grades 3-8 statewide logged less than 20 minutes on Istation curriculum sessions, and just under a quarter of the students (23%) used the system for 200 minutes or more during the 2013-14 school year. Fidelity, as measured by the percentage of students using the system at vendor recommended usage levels (i.e., 250 minutes for elementary school grades and 200 minutes for middle school grades), was also relatively low overall (33%), and declined substantially from Grade 3 (39%) to Grade 8 (7%). This is an important finding because when student outcomes are analyzed later in this report, low usage levels are a key consideration.

As reported in Research Question 1a, Istation usage varied considerably across grade levels with elementary students (i.e., Grades 3-5) using the system much more often than middle school students. Elementary school students were logged into curriculum sessions for more than 300 minutes (25% to 35%, depending upon grade level) during the 2013-14 school year more frequently than students in Grade 6 (10%), Grade 7 (6%), or Grade 8 (4%).

It appears as though students who performed lowest on the prior year STAAR reading assessment and students classified as ELL were identified more frequently for the Istation reading intervention. This is consistent with the SSI goal of providing additional resources to struggling students so they can meet the state standards for grade promotion. Key findings for various student groups include:

- Students in the bottom quartile of 2012-13 STAAR-Reading scores were substantially more likely to use Istation curriculum more frequently than students in upper quartiles of STAAR performance, and the difference between student usage by prior STAAR-Reading scores for lower and upper quartile students was substantially larger for students in middle school grades.
- While the Istation usage differences were not as great as those for students in bottom and top quartiles of 2012-13 STAAR-Reading scores, students previously retained in grade used the system at a somewhat higher rate than non-retained students.
- Only minor differences in Istation system usage were observed between male and female students.

- A much higher percentage of students classified as ELL in middle school used the Istation curriculum at recommended thresholds than their non-ELL counterparts. Differences were more modest between elementary level ELL and non-ELL students.

To gain a better understanding of reasons for variation in Istation usage across Texas campuses, the evaluation team conducted telephone interviews with reading interventionists responsible for implementing Istation and other reading programs in their respective districts and schools. While staff at the majority of campuses (regardless of their overall Istation usage level) felt they had adequate instructional staff (84%), education aides (68%), and internet connectivity (80%), just 55% noted that they had an adequate number of computers or laptops to effectively implement Istation at their campus. A shortage of computers was noted as an issue more frequently by staff at campuses using Istation at low levels, and middle schools.

Guiding Question 3a:

To what extent are program supports available through the Istation vendor, and ESC 20 and other service centers, being utilized by participating districts and campuses?

Key Findings:

- **Various professional development opportunities have been offered on Istation.** Webinars and onsite professional development related to Istation were provided in 2012-13. In 2013-14, in-person training at regional ESCs was offered in fall 2013. Data provided by Istation and the service centers show that 31 different training opportunities were provided by Istation staff at the 20 regional ESCs in 2013-14, with at least one training held at each service center. Almost 1,900 staff members from 469 different school districts (typically campus staff) were documented as having attended Istation-related training at the various ESCs across the state, representing approximately 46% of school districts.
- **ESC 20 staff provided technical assistance and support to district and campus staff during the 2012-13 and 2013-14 school years.** Of the 3,443 technical assistance inquiries received from September 2012 to March 2014, 949 (28%) were exclusively related to Istation and 576 (17%) involved questions about both Istation and TTM. The remainder of calls were related to TTM or unrelated issues. The vast majority (88%) of callers to ESC 20 asked for technical assistance about signing students up for the Istation system or logging in to the online programs. After October 2013, there was a sharp drop off in the volume of technical assistance calls fielded by ESC 20.
- **Overall, campus-level staff seemed satisfied with Istation training, but satisfaction levels were lower at campuses with low usage levels.** Almost two-thirds of school staff interviewed for this report cited that the training they received was sufficient for them to effectively implement Istation, though those satisfaction rates were lower among staff at campuses with low system usage rates (50%).
- **Campus staff appeared to be relatively satisfied with the Istation program and support from ESC 20.** Overall, the majority of interviewees were "very satisfied" with the user friendliness of the Istation program (70%), the appropriateness of the content of Istation lessons for their students (64%), content delivery in Istation (60%), and technical support from ESC 20 (61%). That said, satisfaction levels for all categories noted above were substantially lower for campuses utilizing Istation at low levels.

An important aspect to effective implementation of any online educational program is ensuring that school district and campus staff have received high quality training regarding the use of the program. This includes technical support related to getting students registered to use the system, handling system login issues, and dealing with other problems that may arise when campuses are attempting to use the system with their students. It also includes programmatic support related to lesson planning for students, methods for supporting students in their use of the system, monitoring student usage, and generating system reports.

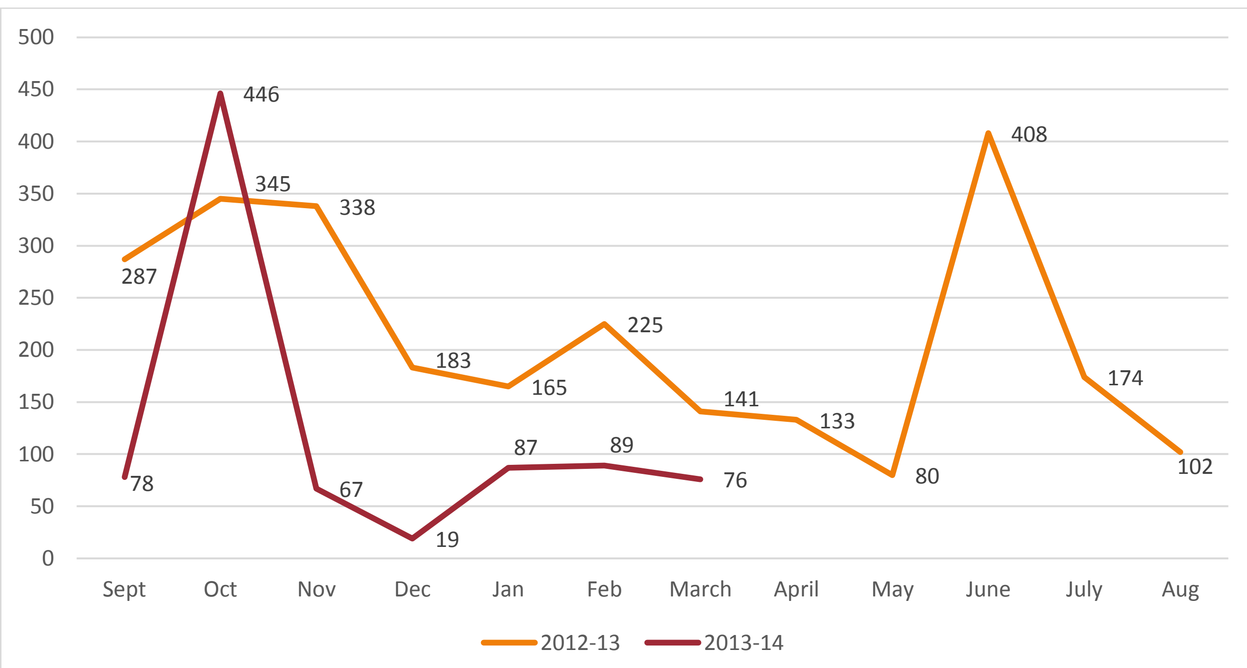
This section explores the extent to which Istation training and technical assistance opportunities were made available to district and campus staff, and the extent to which staff capitalized on these services.

Telephone-Based Technical Assistance for Texas SUCCESS Provided by ESC 20 Staff (Istation and TTM)

During the 2012-13 and 2013-14 school years, staff at ESC 20 provided technical telephone support for both the Istation and TTM programs. Based on data reported by the service center, ESC 20 received a total of 3,443 support calls for both the Istation and TTM programs over the September 2012 to March 2014 period.

As Figure 2.1 shows, after October 2013, there was a sharp decline in the volume of technical assistance calls fielded by ESC 20. A total of 1,153 Istation and TTM technical assistance calls were handled by ESC 20 staff in the first four months of the 2012-13 school year compared to just 610 calls during the same four months of the 2013-14 school year.

Figure 2.1. Number of Istation and TTM Technical Assistance Calls Fielded by ESC 20 Staff, September 2012 – March 2014



Source: Education service center Region 20, 2014.

Note: Data for 2013-14 were only available through March 2014.

Of the 3,443 technical assistance inquiries received over the September 2012 to March 2014 period, 949 were exclusively related to Istation and 576 involved questions about both the Istation and TTM programs. The remainder of the calls were related to TTM or some other unrelated issue.

Most commonly, callers were either teachers (33%) or parents or relatives of the student (33%). A total of 12% of the calls came from campus administrators and 5% originated from district central office staff. The vast majority (88%) of the calls to ESC 20 were related to technical assistance about signing students up for the Istation system or logging in to the online programs. These figures were not disaggregated by whether an inquiry was related to Istation or TTM.

Onsite Vendor Professional Development and Training Provided Through Webinars (Istation only)

During the 2012-13 and 2013-14 school years, Istation utilized two approaches for training district and campus staff on how to use the online reading system: 1) onsite training at the regional ESCs (in 2013-14); and 2) webinars (in 2012-13). Over the July 2012 to June 2013 period, the Istation vendor reported providing professional development directly to over 170 different entities (school districts and campuses), which included 260 participants at onsite training sessions, and 68 webinar attendees. With the rollout of face-to-face trainings at regional ESCs (described below), onsite professional development and sponsored webinars declined substantially in 2013-14. Over the July 2013 to June 2014 period, 82 different districts or campuses were provided training on the system. This included 162 attendees of onsite training and 11 who attended Istation-sponsored webinars.

In-Person Professional Development Provided by Istation Staff at Regional ESCs

Statewide efforts to support widespread usage of the Istation system included technical phone support provided by ESC 20 in San Antonio, webinars sponsored by the Istation vendor, and face-to-face professional development provided by Istation staff at each of the 20 regional ESCs across the state. While data are not available on the number of onsite trainings conducted, it has also been reported that Istation staff served district and campus staff through onsite training and technical support. It appears as though campus and district staff were provided with adequate and varied opportunities to learn how to use the Istation system to support students in reading in their respective jurisdictions.

While webinars and in-person, onsite professional development related to Istation were provided during the 2012-13 school year, in-person training at ESCs did not commence until fall 2013. Data provided by Istation and the ESCs show that 31 different training opportunities were provided by Istation staff at the 20 regional ESCs during the 2013-14 school year. At least one training was held at each of the service centers. In addition, Information regarding Texas SUCCESS was shared information with district and campus staff at dozens of education conferences over the 2012-13 and 2013-14 school years.

As Table 2.11 shows, staff from 469 different school districts (typically campus staff) were documented as having attended Istation-related training held at the various ESCs across the state.[27] This represents approximately 46% of the school districts in the state (excluding charter school organizations). Almost 1,900 school district and campus staff participated in the face-to-face Istation training during the 2013-14 school year.

District participation at training varied substantially across Texas, with the highest percentages of districts (in the 78% to 92% range) at Istation training in Region 19 (El Paso), Region 4 (Houston), and Region 1 (Edinburg). The lowest rates of participation at the district level (in the 26% to 32% range) were in Region 8 (Mount Pleasant), Region 7 (Kilgore), Region 15 (San Angelo), and Region 16 (Amarillo). (See Table 2.11.)

[27] Data related to which or how many campuses were represented in the data was not collected uniformly by all ESCs, so those totals are not available. However, it is clear from the data files that for larger districts, staff from multiple campuses were often represented.

Table 2.11. Istation Professional Development Provided by Vendor at Regional ESCs, 2013-14

ESC	Date(s) of Training	Number of Districts Served in the Region	Number of Districts Represented at Training	Percent of Districts Represented at Training	Number of Attendees at Training
1 (Edinburg)	10/28/13	37	29	78.4%	145
2 (Corpus Christi)	11/21/13 & 2/11/14	42	18	42.9%	127
3 (Victoria)	2/11/14	39	19	48.7%	60
4 (Houston)	11/12/13 & 2/13/14	51	43	84.3%	163
5 (Beaumont)	2/4/14	32	21	65.6%	82
6 (Huntsville)	11/1/13 & 5/15/14	56	25	44.6%	98
7 (Kilgore)	2/10/14	96	30	31.3%	78
8 (Mount Pleasant)	10/23/13	47	12	25.5%	23
9 (Wichita Falls)	1/13/13 & 1/24/14	37	17	45.9%	55
10 (Richardson)	2/13/14 & 4/16/14	80	31	38.8%	148
11 (Fort Worth)	11/11/13	76	27	35.5%	89
12 (Waco)	10/28/13 & 12/10/13	77	27	35.1%	106
13 (Austin)	2/11/2014	56	32	57.1%	100
14 (Abilene)	10/31/13 & 2/28/14	42	24	57.1%	91
15 (San Angelo)	11/18/13 & 2/12/14	42	13	31.0%	32
16 (Amarillo)	10/17/13 & 3/21/14	62	20	32.3%	81
17 (Lubbock)	10/9/13 & 2/11/14	57	20	35.1%	116
18 (Midland)	10/16/13	33	15	45.5%	54
19 (El Paso)	2/4/14	12	11	91.7%	68
20 (San Antonio)	10/15/13 & 2/11/14	52	35	67.3%	171
Total		**1,026**	**469**	**45.7%**	**1,887**

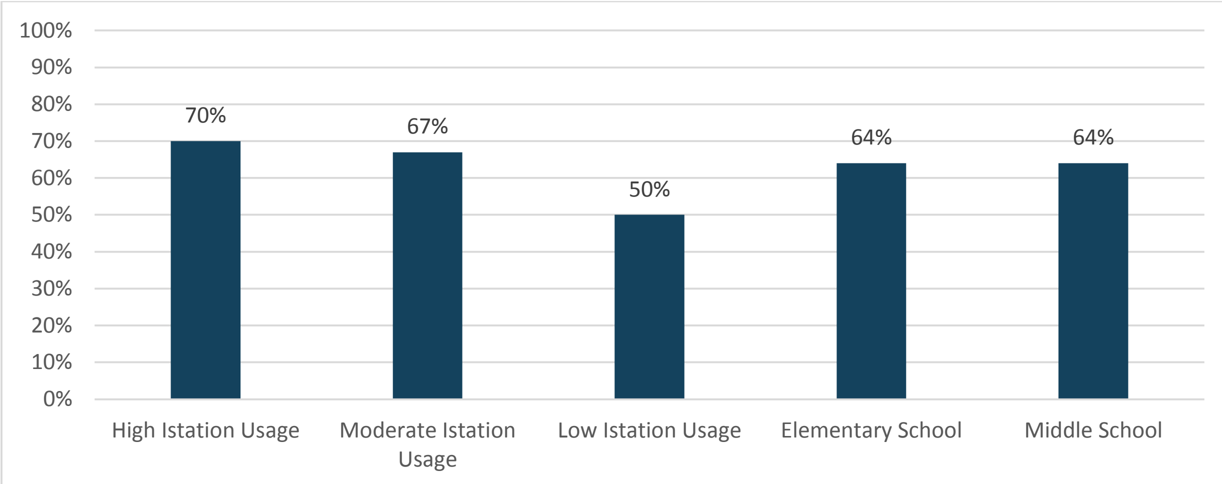
Source: Texas Education Agency (as reported by ESCs), 2014.

Note: District counts do not reflect charter school organizations, but Istation training participation counts do include charter organizations, resulting in inflated participation rates

Campus Staff Perspectives Regarding Istation Training and Technical Support from ESC 20 Staff

The vast majority of interviewees (80%) indicated that staff at their campuses attended training on how to use the Istation program. Overall, staff at 63% of campuses said that the training they received was sufficient for them to effectively use the Istation system. However, a smaller proportion of staff at campuses with low system usage rates (50%) felt the training was sufficient to implement Istation at their campus compared to 70% of high usage and 67% of moderate usage campuses. (See Figure 2.2.)

Figure 2.2. Percentage of Interviewees Noting that Istation Training Received Was Sufficient



Source: Spring 2014 interviews with campus-level staff.

Interviewees were asked to describe any additional training on the Istation system that they would like to receive. Of the 69 interviewees who provided responses to this item, the largest proportions sought information on how to use or read the Istation data (28%), and information about best practices in using Istation (25%). Low Istation usage campuses were much more likely to request additional training on these two items than high and moderate usage campuses.

Overall, the majority of interviewees were "very satisfied" with the user friendliness of the Istation program (70%), the appropriateness of the content of Istation lessons for their students (64%), the engaging nature of how content is delivered through Istation (60%), and technical support from ESC 20 (61%). However, a relatively low proportion (36%) indicated that they were "very satisfied" with the level of support, professional development, or technical assistance from the Istation vendor. Satisfaction levels for all categories noted above were substantially lower for low Istation usage campuses. (See Table D6 in Appendix D.)

Generally, district level staff interviewed in spring 2014 indicated that district-level offices coordinated training and technical assistance for schools, with two-thirds of interviewees (67%) noting that training and technical assistance for school staff on Istation was coordinated at the district level. Six interviewees (20%) noted that individual campuses coordinated training and technical assistance and three others (10%) reported that their central offices collaborated with individual schools to coordinate training and technical assistance. One interviewee noted that no training was provided in 2013-14.

When asked about the sufficiency of professional development opportunities on Istation, most interviewees (71%) replied that they were satisfied. Of those who reported that Istation professional development was not sufficient, their most common criticism was that the training did not provide enough information on specific system features or resources that could have been useful for teachers. Others noted that they would like to have more ongoing training opportunities (i.e., multiple and continued professional development sessions) or opportunities via different modalities (e.g., webinars). While most interviewee reports about vendor-provided professional development were positive, some noted that communications between vendor representatives and their district were more limited than they would have preferred.

Guiding Research Question 3a: Summary of Findings

Based on data provided by the Istation vendor and ESC 20 staff, it is clear that adequate opportunities were provided to school districts and campuses to learn about how to use the online reading system and how to get students registered to use Istation. In addition to various webinars offered during 2012-13, face-to-face Istation training was made available to district and campus staff at each of the 20 ESCs across the state during the 2013-14 school year. In total, approximately 469 (or approximately 46% statewide) school district and charter school organizations (including approximately 1,887 staff members) took advantage of the Istation training delivered between October 2013 and May 2014.

ESC 20 (San Antonio) provided statewide technical support for both Istation and TTM for school districts and campuses with questions about the Istation and TTM programs. The volume of calls was substantially higher in 2012-13, but peaked with 446 calls received in October 2013. Of the technical inquiries, 949 were related exclusively to Istation, and most calls tended to be about how to sign students up in the system or how to log onto the Istation program.

During a telephone survey of campus reading interventionists, respondents were asked a series of questions about their satisfaction with the Istation program, vendor support and training, and ESC 20 support. Overall, the majority of campus staff were very satisfied with the user-friendliness of the program (70%) and the appropriateness of the Istation content (64%), and indicated that the Istation content was delivered in an engaging manner (60%). It is most common for Istation training to be coordinated at the district level, and the majority of district staff (71%) indicated that they were satisfied with professional development opportunities related to the Texas SUCCESS programs.

Guiding Question 4a:

How are school districts and campuses using Istation to address the SSI grade promotion requirements for reading/ELA, and what other reading resources are being used in combination with Texas SUCCESS programs to further support the learning needs of students to meet the SSI grade promotion requirements?

Key Findings:

- **Istation is rarely the only reading intervention program being utilized at campuses to support struggling readers.**
  - ✓ With regard to all reading interventions and programs used on their campuses, only 8% of school staff reported using only Istation, while staff at92% of campuses indicated that they used at least one other reading program.
  - ✓ Approximately 45 different reading programs were in use across 72 campuses included in interview sample.
- **While the decision to use Istation typically occurred at the district level, campus staff made the majority of the decisions regarding the nature of Istation usage at the particular campus.** A number of districts allowed campuses to decide whether to adopt Istation (8 districts, or 28%); though, in general, initial decisions about implementing the program were either completely centralized (59% of districts) or made via discussions between districts and schools (14% of districts).
- **Given the flexibility campuses had in implementing Istation, it is perhaps unsurprising that reports of program use varied across schools; however, some important commonalities in usage were evident.** A large percentage of campus-level interviewees indicated that they used Istation for both assessment and as curriculum (82%). When used for instruction, schools most commonly reported using Istation during regular classroom instruction (84%), as a pullout intervention for struggling students (60%), and/or as part of their after-school program (40%).
- **For the 2013-14 school year, most campuses began using Istation in early fall, but the consistency of usage throughout the school year varied across campuses.** For the 2013-14 school year, the majority of campuses (85%) began using Istation at the beginning of the year (August—October, 2013), while just over half of interviewees from all campuses indicated that they were still using Istation at the time of interviews conducted for this study (May—June, 2014).
- **Monitoring of system usage was uneven across campuses and the extent to which time spent on the system by students was productive was difficult for staff at many campuses**

to determine, suggesting it may not be the strongest measure of usage. Staff at several campuses noted observing patterns of student use showing that students logged in but did not purposefully complete lessons or master lesson content. Interviewees who observed these patterns reported that campus-level staff administering Istation may not be monitoring system use to protect against students becoming distracted while using the programs.

- **Variability in Istation use was tied to levels of system use and school-level.** Low Istation usage campuses and elementary campuses were less likely to report using the system for the entire 2013-14 school year. Low Istation usage campuses were more likely to identify students for Istation use based on grade-level and were less likely to report monitoring system usage. Almost half of middle schools used Istation exclusively in computer labs while less than 10% of elementary schools did so. In elementary schools, Istation was typically used in blended classroom settings (i.e., classrooms with dedicated computers). Middle schools were more likely to report identifying students for Istation based on prior performance on standardized assessments.

The evaluation team had in-depth access to data on Istation usage, as it was one of the primary foci of this evaluation. However, recognizing that Istation was not implemented in a vacuum but, rather, alongside other reading interventions and programs, the evaluation team also sought to explore whether and how schools and districts used other resources in combination with Istation to support student learning in reading. To do so, evaluators posed various questions about the use of other reading resources during telephone interviews with district- and campus-level staff who were involved in overseeing Istation and other reading interventions in their schools and districts. The results reported in this section are based on the responses of campus- and district-level interviewees. See Appendices F and G for the interview protocols. The following sub-questions were explored in this section:

1) How did school districts and campuses use Istation to address SSI requirements for reading?
2) What resources and staff were used by campuses to address SSI grade promotion requirements?
3) How was the use of Istation coordinated with other reading interventions?

How Districts and Campuses Used Istation to Address SSI Requirements

District- and campus-level staff were interviewed to gauge perceptions about various aspects of Istation implementation and use. As indicated in Section 1 of this report, a stratified sampling approach, based on school-level Istation usage (high, medium or low), was used to select the interviewee sample. By using that strategy, the evaluation team was able to differentiate campus-level interview results by whether their school was a high, medium, or low usage school. That is, campus-level usage was used to select the interview sample to ensure that perspectives of school staff, across the spectrum of Istation usage, were considered. Also, campus-level usage was an important condition to consider when disaggregating responses to questions about the system to address, for example, whether high usage campuses would respond more positively about aspects of the system. Campus-level responses were also disaggregated by level (i.e., elementary and middle school). It is important to note that assessing the extent to which interview results varied by Istation usage patterns and school-level was only possible for campus-level interviews. District perspectives represent the perceptions of staff involved with district-wide implementation, regardless of Istation usage patterns or the level of particular schools. As such, district-level interviews provide contextual information regarding district-wide policies and practices related to Istation implementation, but cannot be disaggregated in the same ways as campus-level results. The paragraphs below summarize findings from both district- and campus-level interviews, starting with a description of district-level findings related to Istation implementation. Following that, the remainder of this section focuses on patterns of Istation implementation and use at the campus-level.

District-wide Policies and Practices Related to Istation

A number of districts allowed campuses to decide whether to adopt Istation (8 districts, or 28%); though, in general, initial decisions about implementing the program were either completely centralized (59% of districts) or made via discussions between districts and schools (14% of districts). Outside of the role districts played in initial decisions regarding Istation implementation, centralized mandates about how the system should be used were relatively few. For example, several district-level staff mentioned that they set minimum expectations for Istation use (e.g., minimum number of student log-ins or minimum number of minutes on the system). Others reported that they set requirements about which students should be served by the Istation intervention (e.g., students in specific grades and/or at certain performance levels). Still other districts required use of Istation assessments, but did not mandate use of curricular resources. Besides these district-level expectations, more granular choices about Istation—that is, how, when, where, and with whom the program would be used—were largely left to campuses.

Campus-Level Practices Related to Istation

Given the flexibility campus-level staff had in implementing Istation, it is perhaps unsurprising that reports of program use varied across schools, with differences noted between elementary and middle schools as well as differences by schools' Istation usage patterns. However, there were some commonalities across schools, regardless of school-level and Istation usage patterns. In terms of how Istation was used, a large percentage of campus-level interviewees indicated that they used Istation for both assessment and as curriculum (82%). When used for instruction, schools most commonly reported using Istation during regular classroom instruction (84%), as a pullout intervention for struggling students (60%), and/or as part of their after-school program (40%).

In terms of the timing of Istation use, the majority of campuses (85%) began using Istation in the early fall (August—October, 2013). And, while the timing of use did vary somewhat by schools' overall usage pattern (i.e., high, medium, or low users), over half of interviewees from all campuses indicated that they were still using Istation at the time of interviews conducted for this report (May—June, 2014). Among those who reported less consistent or a shorter duration of use, several noted that inconsistencies were due to the latitude teachers had in deciding how and when they would use the program to support students. While the interview data alone do not allow for conclusions to be drawn about the desirability of consistent usage—i.e., whether consistent usage across a school year leads to better outcomes—several district-level interviewees noted that successful implementation occurred in schools with planned, consistent use of the system.

With regard to students identified for the intervention, most interviewees (80%) reported that the number of hours students used Istation depended on several factors, including student performance level and/or student group. Grade-level was also cited as a factor that informed which students were identified for the Istation intervention, with students in the lower grades in both elementary and middle schools—that is, students in Kindergarten to Grade 2 in elementary and Grade 6 in middle schools—using Istation less frequently than students in upper-elementary and middle school grades. (See Table D7 in Appendix D.)

While this evaluation cannot draw upon interview results for definitive reasons for less frequent identification of students in lower grades, it seems plausible that schools prioritized Istation use among students in tested grades, particularly those with the most substantial accountability implications (i.e., Grades 5 and 8). It is also possible that, in lower elementary grades and Grade 6, schools have less information on students' prior STAAR achievement or performance-level, which many schools cited as considerations when making decisions about which students should be identified for the Istation intervention. Identifying based on prior achievement was particularly prevalent at the middle school-level, with over three-fourths of middle school interviewees indicating that they identified students for Istation participation based on prior STAAR assessment performance, compared to similar statements made by 36% of elementary school interviewees. These and other differences in system use by school-level and Istation schools' usage patterns are outlined below.

Different Patterns of Istation Usage across Campus Types

Elementary versus Middle Schools

In an effort to better understand Istation system usage patterns, campus staff were asked a series of questions related to the settings in which the online reading program was used and how usage was being monitored by campus staff. With regard to school-level, notable differences emerged in terms of the settings in which Istation was used. Over three quarters (78%) of elementary schools used blended classroom/computer lab settings, while blended settings were used in less than half of middle schools (48%). By contrast, 42% of middle schools used Istation exclusively in computer labs, while Istation was used solely in computer labs in 8% of elementary schools. Among those who cited Istation use solely in computer labs, availability of computers and lab space was often mentioned as a challenge to implementation. These interviewees noted that scheduling and logistics played a part in whether students had access to the Istation intervention. This may help to explain some of the variation in usage patterns at the elementary and middle school levels.

Another difference emerged in terms of campus-level monitoring of system use, the responsibility for which typically fell to campus-level administrators. At elementary schools, administrators were more likely to monitor the amount of time students spent on the system (61%), while this was true at 28% of middle schools. Conversely, middle schools were more likely to monitor for student progress (40%) than elementary schools (21%). Outside of school-level monitoring, oversight of Istation usage was also conducted at the district-level. While descriptions of the exact form and nature of district-level monitoring varied, several district-level interviewees noted that monitoring—both at the district- and campus-levels—was important to the effective implementation of Istation and/or TTM. Those who commented on monitoring reported that they most closely tracked system use when expectations about use were mandated, planned and/or incorporated into district-wide curriculum frameworks. As with campus-level staff, district-level interviewees noted that they monitored the number of hours schools spent using the systems each week and/or met periodically with principals or school instructional staff to discuss system implementation. Some interviewees cited apparent issues with monitoring at the campus-level. Specifically, several noted observing patterns of student use showing that students logged in but did not purposefully complete lessons or master lesson content. Interviewees who noted these patterns reported that campus-level staff administering Istation and/or TTM may not be monitoring system use to protect against students becoming distracted while using the programs. This is an important observation to keep in mind when interpreting student outcomes for Istation and the impact of increased minutes on curriculum sessions, in that not all time spent on the system was necessarily productive. As one district-level interviewee noted:

> *You can monitor from the district end that there are areas that haven't done it with fidelity. When you look at the data and see that there are a lot of students who've accessed the program and touched a lot of lessons but there's no mastery of those lessons. You can look at the reports in that way, and it really tells us that it's not implemented with fidelity. You may have folks in the labs that are not monitoring it as they need to.*

Difficulties in monitoring were also noted by campus-level interviewees, with some noting that students used the system in computer labs or other settings that were monitored by school staff other than students' teachers. In those cases—even when student-level usage reports were provided to teachers—the extent to which teachers reviewed and used information in the reports seemed to vary by campus. Several interviewees noted that low incidence of teacher monitoring was due to a lack of teacher awareness about what they should be looking for in student usage reports.

A final school-level difference emerged with regard to the reported consistency and timing of Istation use, with a larger proportion of interviewees at middle schools (87%) noting consistent use of the Istation program across the 2013-14 school year than staff at elementary schools (65%). This difference also emerged by Istation use patterns—that is, staff at high and moderate Istation usage campuses (83% and 79%, respectively) reported that Istation was consistently available in 2013-14, while only half of low usage campuses (50%) reported the same. Other differences in Istation implementation across schools characterized as high, medium, and low users of the system are outlined in the next section.

High, Medium, and Low Usage Campuses

Differences that emerged by Istation usage level include changes in how the system was used over time, monitoring of system use, and as described more fully in the next section, other resources used in tandem with Istation to support students. In terms of changes over time, for interviewees indicating that Istation had been used at their schools prior to the 2013-14 school year, more high usage campuses stated that the way in which they had used Istation had changed over time, relative to reports of changes among low usage campuses (74% versus 60%, respectively).

Generally, it was more common for schools to report using Istation more in 2013-14 (38%), with a smaller proportion of schools reporting less use (10%; see Table 2.12). Some schools also noted that they had begun using the program in a different way. For example, one interviewee said:

> *When we first began utilizing the Istation program, it was in a pullout setting. Now we have a combination, of where kids are put into a reading lab class where they go every single day, and they still have also a pullout. So we have a combination setting.*

Another campus staff member indicated that "*we are using it as an intervention this year, but last year we only used it for assessment*." When considering differences by Istation usage patterns, both high and low using campuses reported similar increases in Istation usage (43% and 44%, respectively). However, high and moderate use campuses were more likely to report differences in the way they used the system—that is, accessing different functions or using the system in different settings—than low usage campuses (39% and 28% versus 11%, respectively). High and moderate use campuses were also less likely than low using campuses to report reductions in use in 2013-14 (4% and 11% versus 22%, respectively). In other words, among campuses that reported using Istation before the 2013-14 school year, those characterized as low users in 2013-14 were more likely than other campuses to report the change as a reduction in use. (See Table 2.12.)

Table 2.12. Major Themes Related to Changes in Istation Usage, Overall and By Usage Patterns

	Overall (n= 50)	High Usage (n= 23)	Moderate Usage (n= 18)	Low Usage (n= 9)
Used Istation more	38%	43%	28%	44%
Used Istation less	10%	4%	11%	22%
Used Istation differently[a]	30%	39%	28%	11%

Source: Spring 2014 interviews with campus-level staff.

Note: [a] Different use refers to reductions in use, different functions used and/or use of the system in different settings.

Overall, staff at over three quarters of the campuses reported that they monitor Istation usage in some manner. The majority of high and medium-level usage schools (89% and 84% of schools, respectively) reported monitoring use while monitoring of Istation use was reported at less than half of low usage campuses (44%). Those interviewees who reported monitoring Istation use at their campuses noted tracking both student usage patterns and student progress.

Interviewees were also asked to indicate whether TTM was implemented in the way they had planned on their campuses. Overall, 54% of the interviewees indicated that they felt the program was implemented with fidelity. However, a great deal of variation was reported by system usage level. The majority of high usage campuses (80%) noted the program was implemented with fidelity compared to 39% of moderate and 33% of low usage campuses. (See Table D18 in Appendix D.)

Resources and Staff Used by Campuses to Address SSI Requirements

To better understand the larger context of SSI requirements to school districts related to serving students at risk of not meeting the requirements for grade promotion, the evaluation team queried campus staff about what reading resources and programs were being used, in addition to their core reading program delivered through regular class instruction and the Istation online program, to support students at their campuses. Information that follows helps to shed light on the fact that Istation is not the only reading intervention in place at campuses using Istation and that campuses not using Istation are most likely using other online and face-to-face reading intervention strategies.

On average, most campuses reported that they used a variety of reading-related services and programs, along with Istation, to help students meet SSI requirements in 2013-14. In fact, staff at less than 10% of campuses indicated that Istation was the only reading program used at their school. Staff at 42% of schools indicated that they used one additional reading program, 28% said that they used Istation and two additional reading programs, and over 20% reported that they utilized three or more other programs in addition to Istation at their campuses (see Table 2.13). Across the 72 campus included in the interview sample, 45 different reading programs were identified as being in use.

Table 2.13. Number and Percent of Other Resources Used to Address SSI Requirements

	Number of Other Resources	Percent
No supplemental reading programs other than Istation	6	8%
One other program	30	42%
Two other programs	20	28%
Three other programs	7	10%
Four other programs	6	8%
Five or more other programs	3	4%
Total	**72**	**100%**

Source: Spring 2014 interviews with campus-level staff.

In addition to other reading interventions that were used, interviewees mentioned different instructional strategies and settings used to help struggling readers address SSI requirements, including:

- In-class strategies (71%)
- Out-of-school strategies (41%)
- Other online programs (35%)
- Assessments (14%)

For further detail, see Table D8 in Appendix D.

With regard to in-class strategies, 73% of campuses used small group and/or direct teaching sessions to supplement Istation. Out-of-class strategies included using Istation and other programs during pull-out interventions or remediation, and during afterschool and tutoring sessions. Students were identified for additional reading support or remediation via their performance on their state test scores from 2012-13 and/or via benchmark or interim assessment results during the 2013-14 school year. The most commonly noted formative indicators of students' reading skill development were the Istation assessment, ISIP (24%), teacher knowledge of individual student needs (19%), and benchmark tests (19%).

Considering differences in other resources used by school-level and Istation usage, middle schools tended to use in-class strategies more frequently than elementary schools (88% versus 56%); while staff at elementary schools more frequently used the program in before- or after-school programs more frequently (47% versus 33%). Elementary and middle schools reported using assessments and other online programs with similar frequency. With regard to other resources used, low usage schools were less likely than high or moderate usage schools to use other online programs (22% versus 38% and 36%, respectively).

Aside from campus-level provision of resources for struggling readers, most district-level interviewees noted specific recommendations or strategies that they promoted for working with struggling students at schools in their districts. Only one interviewee noted that decisions about serving struggling students were left completely up to individual schools. All others noted some district-level involvement, which most commonly took the form of providing schools with a list, guide, or menu of instructional approaches to take with struggling students. District-level coordination of support for schools was often dependent on the size of the district. Support for schools in smaller districts often came from one central office department, while support in larger districts was coordinated through content-specific departments.

Coordination of Istation with Other Reading Interventions

Despite the multitude of resources, strategies, and programs used to support students in meeting SSI requirements, almost three quarters of campus-level staff (48 schools, or 71%) noted that they did not coordinate services and programs funded through other sources with the Istation interventions. In other words, most campus-level interviewees did not report systematic coordination of resources designed to support students in meeting grade-level requirements. Those that did report coordinating services were asked to report on challenges they faced in doing so. Commonly cited issues were related to logistics, e.g., planning and scheduling among different teachers and students who needed access to the resources and having sufficient time to address students' reading needs as well as other areas where they may also be having difficulty. Another challenge cited was having sufficient staff who were trained and available to support students in using various resources and programs.

Guiding Research Question 4a: Summary of Findings

At the district-level, staff were mostly involved in initial decisions about whether to implement Istation. Some districts also set expectations for schools' use of the program and/or monitored program use. Campuses made the bulk of decisions regarding the nature of Istation use, with interviewees most commonly noting that they used both the instructional and assessment-based features of Istation and that the system was typically used to support regular classroom instruction.

Variability in Istation use was tied to levels of system use and school-level:

- Low Istation usage and elementary campuses were less likely to report utilizing the system for the entire 2013-14 school year.
- Almost half of middle schools used Istation exclusively in computer labs, while less than 10% of elementary schools did so. In elementary schools, Istation was typically used in blended classroom settings (i.e., classrooms with dedicated computers).
- Middle schools were more likely to report identifying students for the Istation intervention based on prior performance on standardized assessments.
- Low Istation usage campuses were more likely to identify students for Istation use based on grade-level.
- Low Istation use campuses were also less likely to monitor system usage.

With regard to all reading interventions and programs used on their campuses, Just 8% of school staff reported using only Istation, while the vast majority indicated that they used at least one other reading program. A wide array of different supplemental reading programs were in place at campuses. Among the 72 campuses included in the interview sample, 45 different reading programs were identified in addition to Istation.

At both the campus- and district-levels, most interviewees noted that schools had sufficient instructional staff to implement reading interventions. There was less agreement that campuses had sufficient computer access and technological resources to implement online reading programs, particularly from campuses registering low Istation and other online reading program usage in 2013-14.

Student Outcomes Related to Istation Reading and Other SSI-related Reading Interventions

Relationships between Istation implementation and reading outcomes were explored in a variety of ways. First, to explore the extent to which students evidenced progress within the system, the evaluation team considered whether students who used the system showed meaningful progress as evidenced by performance data contained in the online program. The extent to which this relationship varied by the fidelity of Istation implementation in different schools and districts was also considered (Guiding Question 5a). Second, the evaluation team explored relationships between school-level reading achievement and schools' implementation of a broader complement of SSI interventions, including Istation (Guiding Question 6a). Next, to explore relationships between Istation use and reading achievement on STAAR exams, the evaluation team used several analytic strategies to provide a multi-faceted view of those associations. Those included:

- For Guiding Questions 7a.1, 7a.2, and 7a.3, the evaluation team explored whether relationships between Istation usage and reading outcomes were influenced by the timing and intensity of system usage, or *dosage*. The evaluation team explored whether different concentrations or saturations of system use resulted in stronger associations between Istation participation and reading achievement overall, and among specific student groups.
- For Guiding Question 8a, the evaluation team explored the relationship between the timing of system usage (i.e., the proximity to the STAAR test) and consistency of use (i.e., dosage as measured by number of minutes spent on Istation curriculum sessions) and performance on the first administration of the STAAR-Reading exam. Of particular interest in exploring these associations were possible relationships between Istation use in close proximity to the first administration of the STAAR-Reading test and test outcomes. These indicators of student usage were again included as independent variables in the statistical models predicting reading gains. In sum, by exploring both the timing *and* intensity of system use, these analyses were designed to build on those conducted for the previous question by examining whether stronger associations between students' STAAR performance and Istation use were stronger when use increased during periods where there was less time between that use and STAAR administration.
- To explore whether relationships between Istation use and reading achievement would vary among students whose academic performance classified them as at risk for academic difficulties in 2013-14, the evaluation team explored:
  - For Guiding Question 9a, the extent to which relationships between Istation use and reading performance would differ among students at risk of being retained versus those not at risk was explored.
  - For Guiding Question 10a, the evaluation team explored the relationship between program usage after failing the first administration of the STAAR-Reading test and the probability of passing a second administration of the exam.

Guiding Question 5a:

To what extent are students participating in the Texas SUCCESS program showing meaningful progress as evidenced by performance data contained in the Istation online program, and to what extent is program implementation fidelity related to student growth on the Istation online program?

Key Findings:

- **Middle grade students were less likely to use the system, and use it consistently throughout the school year, and, consequently, less likely to take the assessment multiple times, than elementary grade students.** In Grades 3-5, approximately 20% of the students took at least nine ISIP assessments. In middle school, this pattern was reversed, particularly in Grades 7 and 8, where between 26% and 30% of students took only one assessment.
- **Overall, nearly 70% of students remained in the same performance tier between their first and last ISIP assessment.** Minimal differences are observed across grade levels, with the exception of Grade 4 students, where a slightly higher proportion (15%) of students' tier improved between their first and last assessment, than other grade levels. This is not to say that there were no gains in Istation scores—gains were just typically not large enough to move from one performance tier to the next. In addition, gains were largest among students who used the system for a longer period of time and took their first Istation assessment in September and their last assessment in May (nine months between first and last assessments).
- **The relationship between dosage and growth is inconsistent, with a few exceptions (particularly in Grade 3 and Grade 8).** For Grade 3 students there is a clear linear, positive relationship between system usage categories and gains on the ISIP assessment. It is important to keep in mind that Grade 3 students had higher levels of system usage throughout the school year. The trend for Grade 8 is non-linear and inconsistent when considering usage levels; however, students who used the system at least 250 minutes demonstrated larger scaled score gains (46 scaled score points for students with 250 to 300 minutes of usage, and 50 scaled score points for students with more than 300 minutes of usage).

This section of the report examines student usage of Istation's assessment tools—Istation Indicators of Progress – Advanced Reading (ISIPAR) for Grade 4 and up and Istation Indicators of Progress – Early Reading (ISIPER) for Grade 3 and below—and changes in student performance measured by these assessments (generically referred to as the ISIP assessment). The evaluation also links these performance changes to students' system usage, and the timing of the assessments, to explore the association between usage and growth. The primary focus is on gain scores, which capture the scaled score difference between students' first and last ISIP assessment. This approach is analogous to that used by Patarapichayatham (2014).[28]

Assessment Frequency and Timing

Before examining student growth in reading on the Istation ISIP assessment tools (i.e., ISIPER and ISIPAR), it is important to have a clear understanding of the frequency and timing of ISIP assessments taken by Istation users. For each month of the 2013-14 school year, the last ISIP assessment score a student earned during the month was extracted from the system.[29] Upon their first system login each month, students were automatically directed to the assessment tool, which they had to complete before moving on to Istation curriculum. Thus, theoretically, students who logged in every month of the school year were given the opportunity to complete an assessment.[30]

Roughly 90% of elementary students took at least two Istation assessments, while the same was true of 70% to 80% of middle school students (depending on grade level). In Grades 3-5, approximately one in five students took at least nine ISIP assessments. In Grades 6-8, this pattern was reversed, particularly in Grades 7 and 8, where between 26% and 30% of students took only one assessment. This is consistent with the session login history data previously reported in Research Question 1a: middle grade students were less likely to use the system, and use it consistently throughout the school year, and, consequently, less likely to take the assessment multiple times, than elementary grade students.

Student Growth on Istation System Metrics: Progression between ISIP Performance Tiers

While student growth on the STAAR exam is the primary focus of this study, it is also important to understand the extent to which students progressed on outcome measures (i.e., ISIP assessments) within the Istation system. Based on beginning of the month ISIP assessments, Istation categorizes students into three performance tiers (low, medium, and high). Overall, nearly 70% of students remained in the same performance tier between their first and last ISIP assessment (see Table 2.14). However, it is important to note that students in the bottom to middle of a particular performance tier would be less likely to advance to the next tier up and students in the middle to upper end of a performance tier would be less likely to drop down to a lower tier. Minimal differences were observed across grade levels, with the exception of

---

[28] One critical difference is the rule that Patarapichayatham (2014) used to define the first and last assessment. In that study, only students whose first and last assessments occurred in September and March, respectively, were included in the growth calculation. In this report, we include all students with more than one valid assessment score in more than one month. However, calculations adjust for the amount of time that elapsed between the first and last assessments to ensure that comparisons in changes in ISIP performance across different time intervals are not being made.

[29] The selection rule was slightly more nuanced. For students with more than one assessment *type* (i.e., an ISIPER and an ISIPAR) in a month, the research team first selected the assessment that matched the student's grade level—ISIPER for students in Grade 3, and ISIPAR for students in Grades 4-8.

[30] A small number of students (approximately 3%, or around 41,000 unique students) had no valid assessment scores. Because of this, they could not be linked to their grade level. These students are not represented in any analyses conducted in this section.

Grade 4 students, where a slightly higher proportion (15%) of students' tier improved between their first and last assessment, than other grade levels.

Importantly, middle school students (Grades 6-8) used the ISIP system less frequently and were nearly twice as likely to not have more than one ISIP assessment than students in Grades 3-5 (24% compared to 11%). Thus, they were not included in the growth analyses presented in this section. The consequence of this is that middle grade students were systematically *less* likely to have a measure of growth within the system due to these missing data than were elementary grade students. The severity of this missing data worsens as grade level increases. Approximately 30% of Grade 8 students (approximately 43,000 students) had only one assessment, compared to 10% of Grade 3 students (approximately 30,000 students). This is important because it underrepresents the population of middle school students who started using the Istation system, and it could reflect students either scoring high on the first ISIP assessment and being pulled from the intervention by their school or other systemic reasons for not continuing to use the online reading system.

Table 2.14. Difference between First and Last ISIP Monthly Performance Tier, by Grade Level, 2013-14

Grade Level	Unchanged		Regressed		Improved		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Grade 3	187,190	70.33%	42,880	16.11%	36,089	13.56%	266,159	100.00%
Grade 4	175,452	69.58%	38,263	15.17%	38,451	15.25%	252,166	100.00%
Grade 5	172,739	70.74%	42,712	17.49%	28,748	11.77%	244,199	100.00%
Grade 6	108,782	70.32%	27,620	17.85%	18,304	11.83%	154,706	100.00%
Grade 7	90,518	70.73%	22,892	17.89%	14,564	11.38%	127,974	100.00%
Grade 8	73,467	70.56%	17,736	17.04%	12,911	12.40%	104,114	100.00%
Total	**808,148**	**70.32%**	**192,103**	**16.71%**	**149,067**	**12.97%**	**1,149,318**	**100.00%**

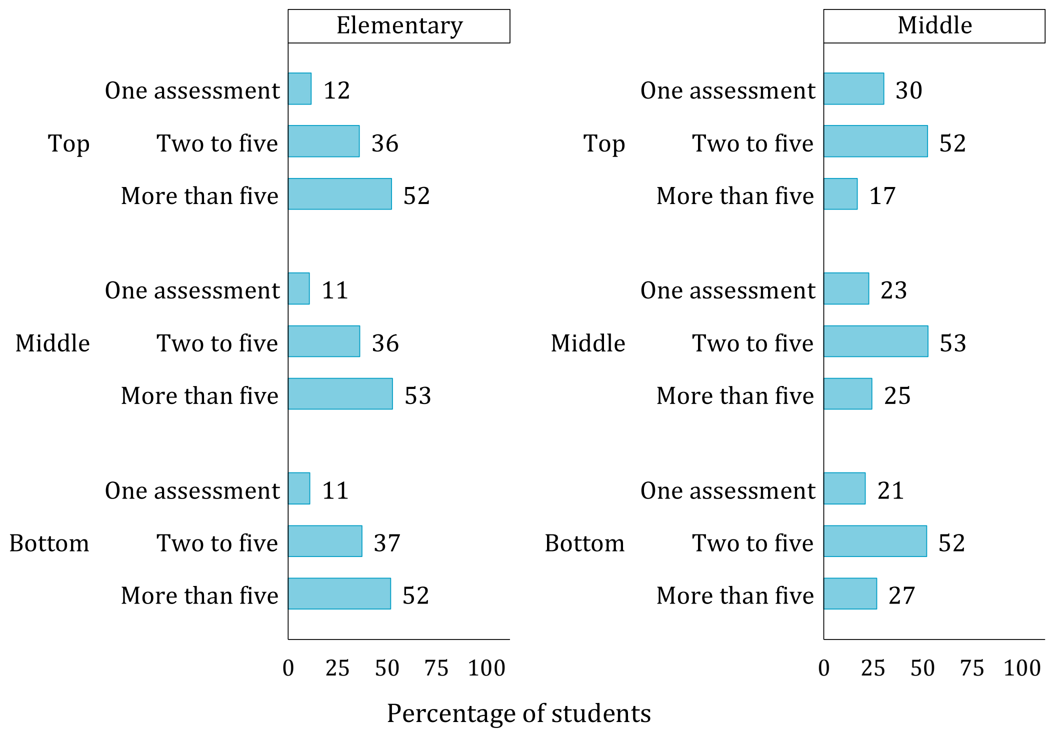
Source: Istation assessment table, 2014.

Note: Counts only include students with at least two ISIP assessments across at least two months. Only assessments with non-missing scores were included in the calculations.

This is an important finding to keep in mind throughout this section: the population of students who had at least two ISIP assessments is different than the population of all other system users. To further explore the differences between students with more than one assessment to those with one or fewer, the research team calculated the percentage of students, within each tier (top, middle, bottom) determined by their first ISIP assessment, with one assessment. Because of more focused targeting of interventions revealed from the campus and district interviews, it is anticipated that in middle grades, students in the top tier based on their first ISIP assessment are *more likely* to have only one assessment than students in the bottom tier. Smaller differences among students in elementary grades are anticipated.

Figure 2.3 provides information to investigate this assumption. The percentage of students with one assessment, between two and five assessments, or more than five assessments is shown by the students' beginning ISIP performance tier and school level. First, irrespective of beginning tier, middle school students were more likely to have only one assessment than elementary students. Again, independent of beginning tier, middle school students were approximately half as likely to have more than five assessments. However, the difference between tiers *within* a school level is revealing. Middle school students in the highest performing ISIP tier were roughly 10 percentage points more likely to have only one assessment than students in the lowest performing ISIP tier: among elementary students, there were miniscule differences. This lends additional empirical support to the behaviors uncovered in the campus and district interview data suggesting greater targeting of the tool in middle schools compared to elementary schools, particularly among lower performing students. The data on the number of ISIP assessments and movement within performance tiers sets the stage for the ISIP growth analysis that follows.

Figure 2.3. Percentage of Students with One or Multiple ISIP Assessments, by School Level and Beginning Tier Level



Source: Istation assessment table, 2014.

Student Growth on Istation System Metrics: Difference in Gain Scores between First and Last ISIP Assessment

Next, patterns of student growth measured by the difference between students' first and last ISIP assessment were explored. The review team classified students according to the first and last month in which they took an ISIP assessment and, in Figures 2.4 and 2.5, restrict the comparison to only students who took their first assessment in September 2013.[31] Students' gain score is calculated by simply subtracting students' first ISIP assessment scaled score from their last ISIP scaled score. Only gain scores for the same assessments are used because the Grade 3 ISIPER assessment is measured on a different scale than the ISIPAR assessment used in Grades 4 through 8. In addition, when presenting the results of the calculations, ISIPAR assessments are presented separately from ISIPER assessments. Only students with assessments in two different months were assigned a gain score and were subsequently included in the calculations.

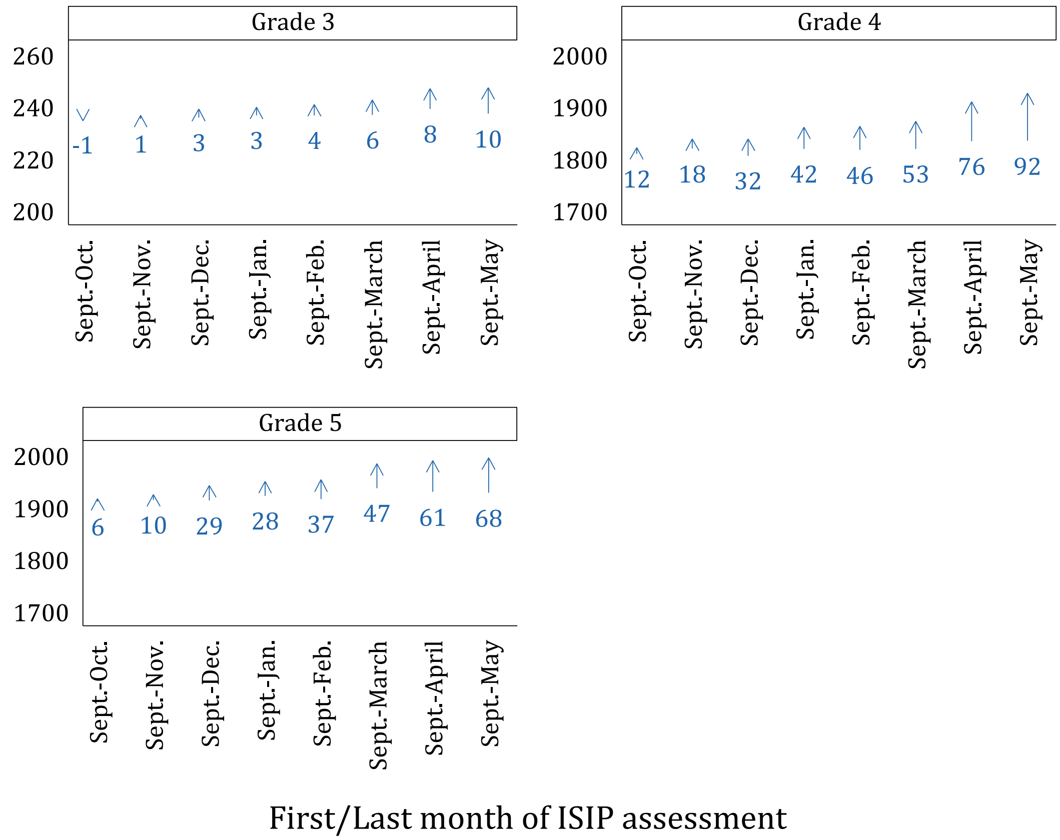
Figure 2.4 includes Grade 3-5 students, and Figure 2.5 includes Grade 6-8 students. The evaluation team calculated the mean gain score by the first and last month in which students had a valid ISIP assessment, by grade level. Some orientation to the design of the visualizations is warranted. The data are ordered by the amount of time that elapsed between assessments (e.g., "Sept.-Oct." is one month, and "Sept.-March" is six months). Each arrow, by first and last assessment month, conveys three pieces of information: the average score on the September assessment (the base of the arrow), the average score on the last assessment (the top point of the arrow), and the average change in scale score points between the first and last assessment (the length of the arrow, which is denoted by the value label at the base of the arrow).

Figures 2.4 and 2.5 demonstrate the importance of incorporating elapsed time into measures of within-system growth. Across every first and last month of assessment group, average gains were larger among students in Grades 4 and 5 compared to students in Grades 6 through 8. Students in Grade 4 experienced the largest average gains: between September and March, the average growth on ISIPER was 53 scaled score points, while the average growth between September and May was 92 scaled score points. The differences were larger when the time between the first and last assessment increased. Although students' ISIP scores in Grades 6 through 8 rose on average, and the magnitude of this growth as the amount of elapsed time also grew, gains in scale scores were larger among elementary grade students than among middle grade students. Confirming earlier findings related to system usage, elementary grade students used the system more consistently, frequently, and intensely, than middle grade students. That is, students who used the system more frequently and for more months of the 2013-14 school year

[31] September was selected because this was the modal month in which students took their first assessment, and this represents the first full month of the school year. This classification according to first and last month of assessment differs from Patarapichayatham (2014) in that students' first and last month assessment are mutually exclusive: thus, students whose first assessment occurred in September, but took subsequent assessments in January, February, and March, would only be placed in the "September-March" group. This classification imposes some comparability across grades and students because it standardizes the amount of time that elapsed between assessments.

experienced larger gains in ISIP scale scores. This finding may be a function of system usage, elapsed time, or a combination of both.

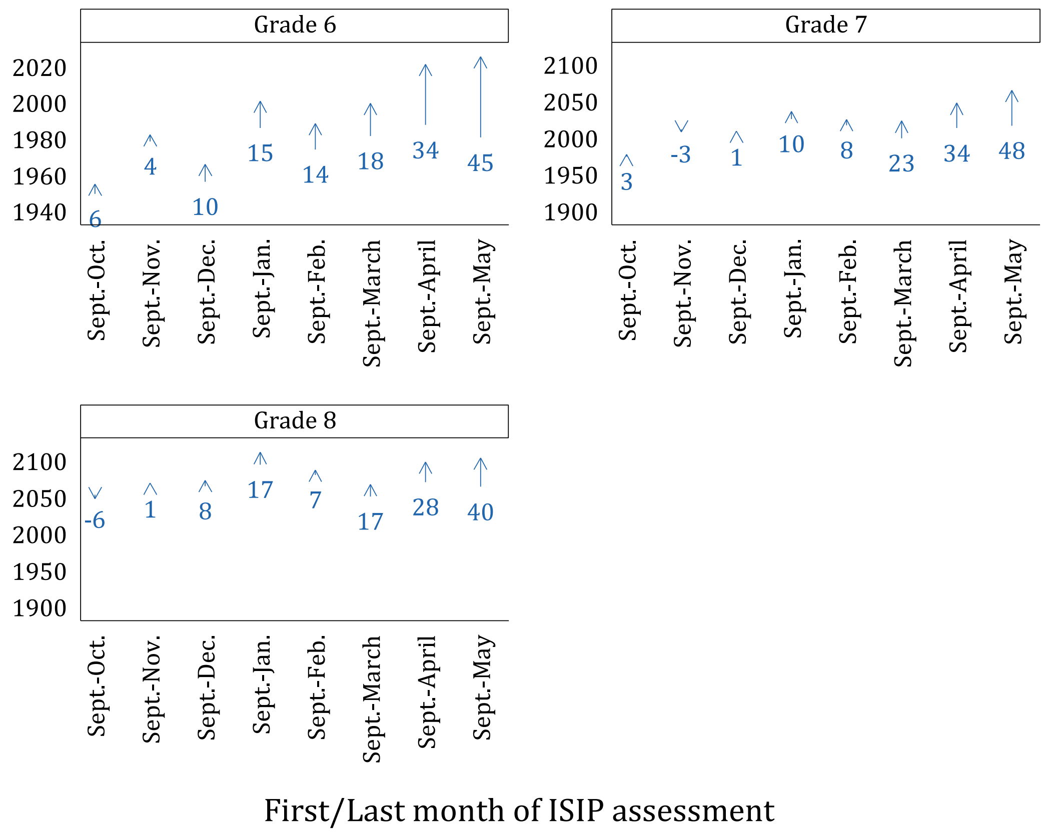
Figure 2.4. Mean Change in ISIP Scaled Score between First and Last Assessment, by Grade Level and Time between First and Last Assessment, Grades 3-5, 2013-14



Source: Istation assessment table, 2014.

Note: Only students with an assessment in September were included in the calculations. First/last month categories were mutually exclusive. Each arrow, by first and last assessment month, conveys three pieces of information: the average score on the September assessment (the base of the arrow), the average score on the last assessment (the top point of the arrow), and the average change in scale score points between the first and last assessment (the length of the arrow, which is denoted by the value label at the base of the arrow).

Figure 2.5. Mean Change in ISIP Scaled Score between First and Last Assessment, by Grade Level and Time between First and Last Assessment, Grades 6-8, 2013-14



Source: Istation assessment table, 2014.

Note: Only students with an assessment in September were included in the calculations. First/last month categories were mutually exclusive. Each arrow, by first and last assessment month, conveys three pieces of information: the average score on the September assessment (the base of the arrow), the average score on the last assessment (the top point of the arrow), and the average change in scale score points between the first and last assessment (the length of the arrow, which is denoted by the value label at the base of the arrow).

Student Growth on Istation System Metrics: Impact of Elapsed Time between First and Last ISIP Assessment

Elapsed time between the first and last assessment is strongly correlated with how heavily students use the system. That is, students with an eight or nine month gap between their first and last assessment are more likely to use the system more frequently and with more intensity during this period than students with a smaller gap. What is more, the longer the span between students' first and last assessment, the more classroom instruction and content they have been exposed to in school, and within the Istation. Both of these likely contribute to how much a student improves on the ISIP assessments. Because the research team knew when students took the assessments, and how much students used the system between their first and last assessment, the research team attempted to disentangle the influence of time and system usage on students' gain scores.

To investigate this, the amount of curriculum system usage minutes a student recorded between their first and last ISIP assessment was calculated. This allowed comparisons to be made between students with the same amount of time between their first and last assessments with comparable curriculum minutes during this period. Put another way, students who took assessments in September and March, and who used the system for 200 minutes during this period, can be compared to students who also took assessments in September and March but who used the system less than 20 minutes. Contrasts between these two groups will shed light on whether system dosage, independent of time, contributes to student growth within the system.

Figure 2.6 provides information on the link between growth on the ISIP assessment, system usage, and time. System usage is categorized into five groups: Fewer than 11 minutes, between 11 and 199 minutes, between 200 and 250 minutes, between 250 and 300 minutes, and more than 300 minutes.[32] These categories were chosen for two purposes: to investigate the contrast in ISIP gain scores between students who did not use the system or used it minimally (and, are most likely different from those who use the system at even low levels)—and those who used the system at different levels of intensity that correspond with the levels recommended by Istation. The calculations were further disaggregated by the elapsed time between students' first and last assessment for students whose first assessment occurred in September, allowing comparisons to be made among students, in the same grade level with a similar dosage level in Istation.[33]

Several patterns emerge. First, irrespective of the number of system usage minutes, students demonstrated considerable growth on the assessment based on the amount of time that elapsed between the first and second assessment. For instance, Grade 4 students, who demonstrated larger scale score gains on the ISIP assessment than students in other grades, who used the system less than 11 minutes, gained 55 scaled score points if the amount of time elapsed was four months (September to January), and 90 scaled score points if the amount of time that elapsed was eight months (September to May). A similar finding was observed for Grade 4 students who used Istation for over 300 minutes. These students gained 47 scaled score points if the amount of time elapsed was four months (September to January), and 94 scaled score points if the amount of time that elapsed was eight months (September to May). However, recall that 70% of the students did not move from one performance tier to another between their first and last ISIP assessment. Overall, this finding suggests that regardless of usage levels, students who were using the system for a longer period of time showed larger gains on their ISIP assessments than students

[32] The range of minutes in curriculum sessions for those using the system for 300 or more minutes was 300 to 4,000 minutes.

[33] For concision and clarity, only students whose first ISIP assessment occurred in September, and whose last assessment occurred in January, March, or May are shown in the figure.

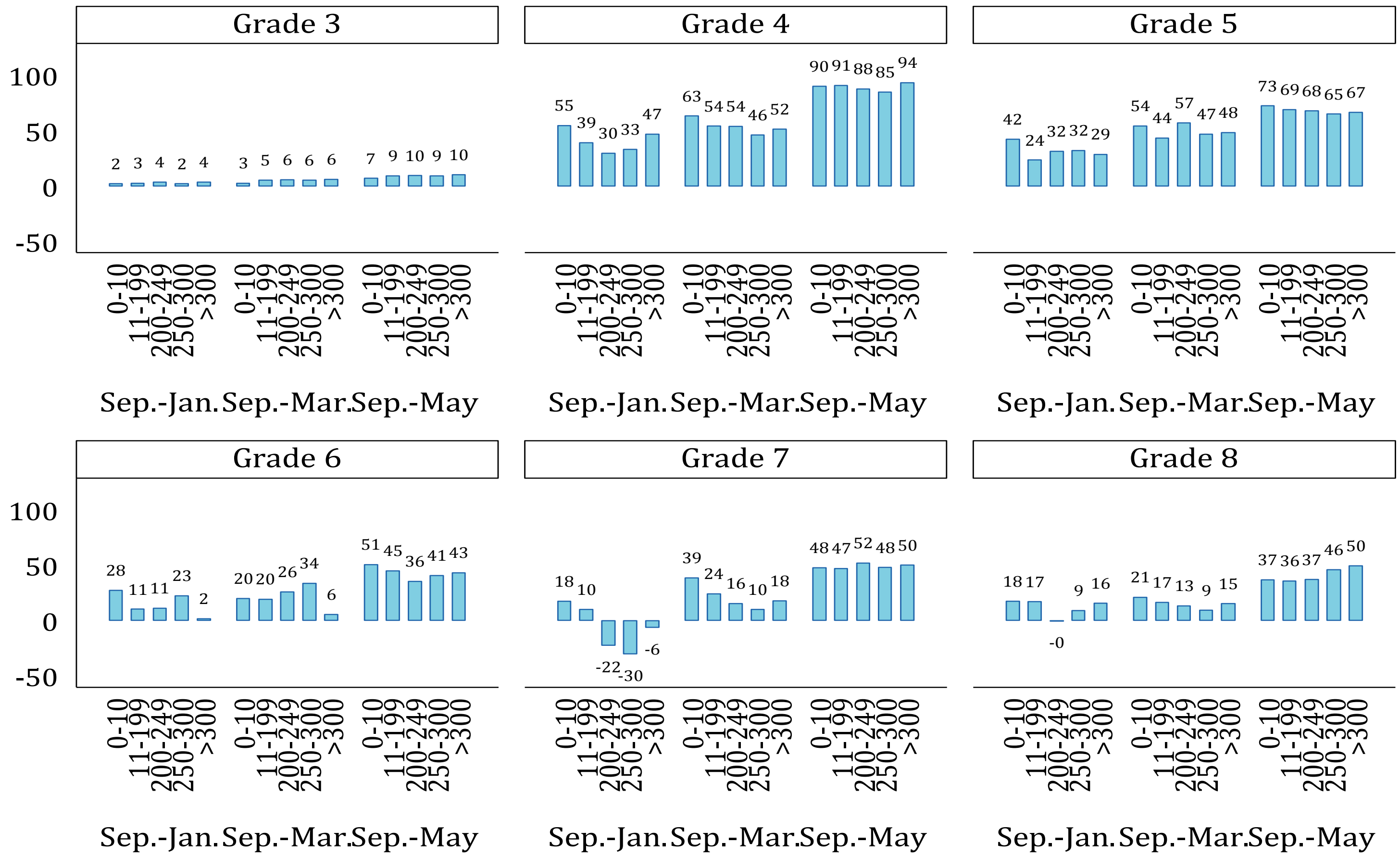
who used the system for fewer months. It is important to note that the gains experienced over a longer period of time may be a function of additional classroom instruction provided to students.

Second, the relationship between dosage (i.e., the number of minutes a student spent on curriculum sessions in 2013-14) and growth (i.e., the difference between the first and last ISIP assessment scale score in 2013-14) is inconsistent and non-linear (i.e., growth does not increase with the number of minutes on the system) with the exception of Grade 3 and Grade 8. That is, for most grades, increasing the number of minutes on the Istation system does not result in larger gains in ISIP scores between the first and last assessment.

For example, in Grades 5 and 6, regardless of how long they used the Istation system, low users (fewer than 11 minutes) demonstrated larger scaled score gains than high users (i.e., students who used the system over 250 minutes). However, for Grade 3 students there is a clear linear, positive relationship between system usage categories and gains on the ISIP assessment. It is important to keep in mind that Grade 3 students had higher levels of system usage throughout the school year. Grade 8 students who used the system throughout the year (September to May) and logged at least 250 minutes demonstrated larger scaled score gains (46 scaled score points for students with 250 to 300 minutes of usage, and 50 scaled score points for students with greater than 300 minutes of usage) than students using the system throughout the year but for fewer minutes.[34]

[34] The research team expected, and found, growth to be greater among students' whose performance on the first ISIP assessment placed them in the bottom tier compared to those in the top tier. However, the impact of system dosage on growth was not strongly associated with students' starting performance level.

Figure 2.6. Mean ISIP Gain Score, by Elapsed Months between First and Last Assessment, Usage Minutes, and Grade Level, 2013-14



Source: Istation assessment and session-by-product tables, and author's calculations.

Note: The number of students represented in each of the subcategories displayed in Figure 2.6 varied substantially—ranging from 56 to 24,191. The lowest numbers of students were found in the 0-10 minute usage category from September to March, where student counts ranges from 56 to 273.

Guiding Research Question 5a: Summary of Findings

Analyses related to this research question first address the frequency and timing of system usage and ISIP assessments during the 2013-14 school year, then turn to an examination of student growth on the Istation assessment. Roughly 90% of elementary students took at least two Istation assessments, while the same was true of 70% to 80% of middle school students (depending on grade level). This supports previously discussed findings that Istation use was more widespread in elementary grades and more targeted in middle grades, with middle grade students from lower performance tiers engaging with Istation more frequently than their peers in higher performance tiers. To be included in the ISIP growth analyses, a student had to complete at least two ISIP assessments during the 2013-14 school year.

Next, reading gains were assessed among those students who took at least two Istation assessments. Almost 70% of students remained in the same performance tier between their first and last assessments. This is not to say that there were no gains in Istation scores—gains were demonstrated across all grades; however, the gains were not generally large enough to move across ISIP performance tiers. The largest gains between first and last ISIP assessments were demonstrated among students with more elapsed time between their first and last assessments. That is, gains were largest among students who took their first Istation assessment in September and their last assessment in May (nine months between first and last assessments). It is important to note that gains may be attributable to other external unmeasured factors.

Along with the timing of assessments, relationships between the frequency of Istation use and gains were examined, with results showing mostly inconsistent (i.e., the relationship did not hold for students using the system for a partial year versus a full year) and non-linear relationships (i.e., increased dosage did not translate to increased gains in ISIP scores), except in Grades 3 and 8. In Grade 3, increased use was positively correlated with gains in Istation assessments, regardless of whether the first and last assessments were administered five, seven, or nine months apart. Patterns in Grade 8 were less straightforward, with non-linearity in the relationships between Istation use frequency and reading gains when assessments were administered five or seven months apart. However, slight incremental gains were demonstrated among students with nine months of elapsed time between their first and last assessments *and* who used the system for at least 250 minutes. No clear relationships were found between gains on Istation assessments and Istation use among students in other grades.

Guiding Question 6a:

What is the relationship between broader implementation of Texas SUCCESS and other SSI interventions and student outcomes for reading?

Key Findings:

- **Descriptive analyses of overall reading interventions revealed schools with higher Istation use utilized fewer supplemental reading programs than low Istation users.** Campuses demonstrating high levels of Istation use were more inclined to not use any other supplemental reading programs or just use one other than campuses that did not use Istation very intensively (68% versus 27%). Conversely, a higher percentage of campuses demonstrating low Istation usage reported using at least two other supplemental reading programs to promote SSI grade promotion requirements than high Istation users (73% versus 32%).
- **No consistent patterns emerged in terms of differences in reading achievement by Istation and/or non-Istation usage patterns in either 2012-13 or 2013-14.** It may be instructive that schools using Istation with greater frequency tended to rely less on other reading programs than low Istation usage campuses. Further exploration would be needed to uncover possible reasons behind this trend as well as possible implications for the use of different reading interventions on reading outcomes.

While a substantial portion of this report focuses on relationships between Istation use and reading outcomes, activities conducted for the evaluation also explored other contextual factors—outside of Istation—that may have influenced whether students' met SSI grade promotion requirements in reading in 2013-14. Consideration of other factors was operationalized in the evaluation design and methodology in several ways. As described in the next section, statistical models analyzed for the impact analyses disaggregated the examination of relationships between Istation use and reading outcomes by grade-level as well as student demographic factors. This approach provided a more robust view of relationships between Istation use and STAAR-Reading performance than simply exploring those trends in aggregate. In other words, impact analyses explored the extent to which Istation use was related to student performance in reading, while also considering the extent to which other factors moderated the relationship between reading outcomes and system use.

To assess the broader picture of how schools sought to meet grade level promotion and accelerated instruction (SSI) requirements, this section explores relationships between reading achievement, and contextual and instructional factors in schools, through a different lens: by considering Istation use along with other reading interventions and programs used in schools. As described previously, Istation was not the only reading program used on most campuses in 2013-14; in fact, less than 10% of school-level interviewees reported that Istation was the only reading program used at their schools. School staff most commonly reported using either one or two other programs (42% and 28%, respectively), with two programs reported as the average and median number of other reading programs used. Study Island was cited most often (17%), followed by Leveled Literacy Intervention kits (14%), and Read 180 (13%).

Given that over 90% of school-level interviewees reported using Istation and at least one other reading program, analyses described in this section sought to provide perspective on relationships between the use of different resources and reading achievement. Data elements used in the analyses included information on Istation usage, data on other programs used to support reading instruction in 2013-14—collected via interviews with school staff—and school-level STAAR-Reading outcomes. Before describing the findings, it is important to note several limitations to these data. For one, while fine-grained data on Istation usage patterns were available, data on schools' use of other programs were collected via interviewee responses regarding questions about their schools' use of "other reading interventions or other reading online learning programs." In other words, programs other than Istation were cited as either used or not used, while information on the frequency, timing, and duration of use, or use with particular grades or types of students, was not reported systematically. As such, relationships between use of programs other than Istation and reading outcomes could not be disaggregated by the nature of how those programs were used (e.g., frequency, intensity, etc.), nor could relationships be explored between non-Istation program use and reading achievement among students in specific grades or groups. Also, information on non-Istation program use was only available from schools selected for interviews (n=75), unlike Istation data, which were collected from all users of that system in public schools in the state. Given these features of the data, findings reported in this section provide a purely descriptive and exploratory view of school-level interventions geared toward improving student reading outcomes. These descriptive analyses were guided by the following two primary questions:

- What was the nature of cumulative, school-level efforts to help students meet SSI grade promotion requirements, operationalized as the combination of Istation usage and the use of other reading programs?
- Were there relationships between these cumulative efforts and school-level reading outcomes?

Descriptive Analysis of Programs Implemented for SSI Requirements

To explore the nature of combined use of Istation and other reading programs, the distribution of school-level Istation usage patterns were compared to the number of other programs reported as used. Analyses focused on whether relationships between Istation and non-Istation program use were positive or negative—that is, whether high Istation using campuses used more or fewer other programs than low Istation usage campuses. Table 2.15 provides results of this comparison, contrasting Istation use levels (i.e., low, moderate, or high usage) with the percentage of schools that were either above or below the mean number of programs used (i.e., fewer or more than two other programs). As the table shows, descriptive usage patterns imply an inverse relationship between Istation and non-Istation program use, with a larger percentage of high Istation usage campuses using zero or one other reading programs than low Istation usage campuses (68% versus 27%). Conversely, a higher percentage of low Istation users reported using at least two other programs to promote SSI grade promotion requirements than high

Istation users (73% versus 32%). Thus, it appears that when campuses made the decision to use Istation at a high level, they were much less inclined to also use other reading interventions, while low Istation use campuses had students using a variety of different reading programs (or perhaps different groups of students using different programs). The percentage of schools categorized as moderate Istation users fell in the middle, reporting using two or more programs (65%) with greater frequency than high Istation users (32%) but with less frequency than low Istation using schools (73%).

Table 2.15. Inverse Usage Pattern between Istation Use and Use of Other Programs (N=73 Schools[a])

	0 to 1 Other Reading Programs (n=35)	2 or More Other Reading Programs (n=38)
Low Istation Usage (n=22)	27%	73%
Moderate Istation Usage (n=17)	35%	65%
High Istation Usage (n=34)	68%	32%

Source: Spring 2014 interviews with campus-level staff.

Note: [a] The research team was unable to match Istation usage data for two of the 75 schools included in the interview sample, resulting in a matched usage-interview sample of 73 schools.

In practical terms, these findings imply that schools with higher Istation use utilized fewer other programs than low Istation users. While these data do not allow causal inferences to be drawn about this trend, several reasonable possibilities can be posed. For one, schools allocating significant time to Istation may have had less time for students to use other programs. Also, as cited in interviews with school- and district-level staff, limited technological resources and scheduling challenges made it difficult for schools to coordinate multiple interventions for students—in fact, less than a third of school-level staff reported coordination among reading interventions (29%). Another possible explanation for the inverse relationship between Istation and non-Istation program use comes from previously reported positive relationships between Istation use and interviewees' satisfaction with Istation content—high Istation users reported more satisfaction with the user-friendliness and content of Istation than did low system users. Given that low Istation users reported less satisfaction with the system than higher Istation users, it is possible that a lack of satisfaction with Istation led low using schools to employ other methods to build students' capacity in reading. The next section explores this possibility further by examining relationships between Istation and non-Istation usage, and school-level reading achievement.

Relationships between Programs Implemented to Meet SSI Requirements and School-Level Reading Outcomes

Analyses conducted for this section explored descriptive patterns in reading achievement—from both 2012-13 and 2013-14—by schools' Istation and non-Istation program usage (see Table 2.16).[35] Across all program usage categories, most schools demonstrated higher school-level reading achievement in 2013-14 than in the previous year. The only exception to this was observed in the 10 schools classified as high

[35] While this report focuses on relationships between Istation use and reading outcomes in 2013-14, analyses throughout have considered prior achievement when examining that relationship. As such, 2012-13 achievement is included here to show, descriptively, average prior reading achievement in schools included in the analysis.

Istation users and users of two or more other programs, where 2012-13 reading scores were one point higher than scores in 2013-14. In general, users of zero or one other reading program had higher reading achievement both years than users of two or more other reading programs. Otherwise, no consistent pattern emerged in terms of noticeable differences in reading achievement by Istation and/or non-Istation usage patterns in either year. That said, it is important to emphasize that these relationships are not causal and that they involve different students each year. That is, average school-level reading achievement was calculated for different cohorts of elementary and middle school students in 2012-13 and 2013-14.

These types of cross-sectional analyses do not allow for conclusions to be drawn about whether school-level achievement differences from one year to the next may have been expected given the performance of student cohorts in the prior year. Findings in this section present a purely contextual and descriptive view of reading interventions and achievement recently demonstrated in Texas public schools.

Table 2.16. Average School-Level STAAR Reading Scale Scores, by Istation and Other Reading Program Use

Zero or One Other Reading Program (n=35)[a]			Two or More Other Reading Programs (n=37)		
Istation System Usage Level	**2012-13**	**2013-14**	**Istation System Usage Level**	**2012-13**	**2013-14**
Low Istation Usage (n=6)	1572	1579	Low Istation Usage (n=16)	1511	1538
Moderate Istation Usage (n=6)	1474	1493	Moderate Istation Usage (n=11)	1498	1519
High Istation Usage (n=23)	1595	1603	High Istation Usage (n=10)	1568	1567

Source: Spring 2014 interviews with campus-level staff.

Note: Ns represent the number of schools in each category.

Guiding Research Question 6a: Summary of Findings

Schools with higher Istation use tended to implement fewer supplemental reading programs than schools that did not use Istation intensively. Campuses demonstrating high levels of Istation use were more inclined not to use any other supplemental reading programs or just use one other than campuses which did not use Istation very intensively (68% versus 27%). Conversely, a higher percentage of campuses demonstrating low Istation use reported using at least two other supplemental reading programs to promote SSI grade promotion requirements than high Istation users (73% versus 32%).

With regard to descriptive relationships between reading program usage and reading achievement, no noteworthy, consistent patterns emerged in terms of differences in reading achievement by Istation and/or non-Istation usage patterns in either 2012-13 or 2013-14.

Guiding Question 7a.1:

To what extent do student performance results differ for students participating in the Texas SUCCESS program (i.e., using Istation) and non-participating students and how do results differ by student characteristics, prior academic performance, and other key variables?

Key Descriptive Findings:

- **Istation assessments were strongly related to performance on the STAAR test, suggesting that the instrument was an important diagnostic tool for assessing students' preparation for STAAR.** The diagnostic performance of the ISIP assessments did differ across student performance levels. On average, about half of students in the lowest ISIP performance tier met the phase-in 1 Level II (Satisfactory) standard in 2013-14, compared to more than 95% of students in the top performance tier.
- **Istation use was associated with lower STAAR passing rates among students in lower performance tiers.** Among students in the lowest two Istation performance tiers, students who used Istation more frequently met the passing standard at lower rates than students who used the system less, or who did not use it at all. These differences were largest in Grades 5, 7 and 8.
- **In most grades, students who did not use Istation demonstrated larger performance gains in reading than students who used Istation.** These differences remained even after students' performance in the previous year was considered. Further, the performance differences persisted, and in some instances grew, as Istation system dosage increased.

Key Multivariate Findings:

- **In Grades 4, 5, and 6, students who used Istation did not perform as well on STAAR-Reading as students who did not use the system**. These differences persisted after accounting for student-level characteristics and different schooling environments. Students who used Istation had reading gain scores that were consistently lower than students who did not use the system, although these differences were substantively small, ranging, on average, from 5 and 8 scale score points. The scale score differences were largest in Grade 6, and smallest in Grade 5.
- **After controlling for prior reading performance, students who used Istation in Grades 7 and 8 had larger performance improvements in reading between 2012-13 and 2013-14 than students who did not use Istation.** The differences were largest in Grade 8, and smallest in Grade 7, although both differences were substantively small.

To address Research Question 7a, the evaluation team linked student-level longitudinal state administrative and STAAR data stretching from 2011-12 to 2013-14 to student-level Istation system usage in 2012-13 and 2013-14. Students were not randomly assigned to SUCCESS interventions. Academically low-performing students may have elected to participate, or may have been instructed to participate, by their parents, teachers, or school principals. What is more, students at risk of being retained may have been urged to participate in the program to improve their preparedness for the state assessment. This type of intentional assignment based on student characteristics that are linked to academic performance may distort the relationship between SUCCESS program participation and academic performance. The evaluation team observed outcomes for students who elected to participate, and those who did not. Because of this non-random assignment, quasi-experimental design was used to compare longitudinal within-student changes in outcomes between participating and non-participating students (Angrist & Pischke, 2009; Wooldridge, 2002).

In this approach, commonly referred to as a difference-in-difference (DiD) design, with longitudinal panel data in which individual students are observed at multiple time points, each student serves as his or her own comparison. In doing so, differences in student performance across time (i.e., differences in their pre- and post-test scores, or in the probability of repeating a grade) among students who participated in one of the SUCCESS interventions are compared to those who did not participate in either program. This approach is flexible, and incorporates other student characteristics that are stable (such as ethnicity or gender) or change (such as behavioral referrals or attendance) to investigate whether association between program participation and changes in STAAR scores differs across student groups, such as student ethnicity or economically disadvantaged status.

The evaluation team also fit a number of additional statistical models with different assumptions about the data generating process and how students are selected into participating in the program. In addition, for the DiD models, the evaluation team estimated a propensity score that calculated the probability that each student, based on a plethora of student, school, and district covariates, participated in the program at different dosages (e.g., at or above the recommended usage threshold; Rosenbaum & Rubin, 1983). The evaluation team fit these supplementary models to assess the stability and robustness of the primary statistical model.

In subsequent paragraphs, associations between different measures of program participation and student achievement using simple comparative descriptive statistics are described for key student groups, such as prior student performance and measures of the intensity of students' system usage. Next, the results from the multivariate analyses are presented for several outcomes.

As discussed in the section addressing Research Question 2a, the population of students who used the system differs from the population of non-users, although this varied considerably by grade level, as usage in middle grades became less widespread and less intense compared to usage in Grades 3 through 5. It is important, then, to take into account these factors in the analyses of differences in student outcomes between Istation participants and non-participations. In this section, differences in both levels of, and changes in, student performance on grade-level STAAR-Reading exams in 2013-14 are described. These analyses take important student-level attributes, such as prior performance, that are jointly related to current STAAR performance and the likelihood of using Istation, into account.

Relationship between Istation Assessment Performance (ISIPAR and ISIPER), Istation System Usage, and STAAR-Reading Performance

The evaluation team linked a number of metrics from the Istation system usage data to TEA administrative and assessment files. These included monthly ISIP assessment scores, which are required each time a student logs into Istation for the first time in a given month. In addition, the evaluation team also extracted the amount of time, each month, a student spent in the system working on a curriculum session. This allowed the research team to assess the relationship between students' performance on the beginning of year (BOY) ISIP assessment (September 2013) and their subsequent performance on STAAR in the spring of 2014. Figure 2.7 depicts the percentage of students, at each grade level, who met the phase-in 1 Level II (Satisfactory) standard on the first administration of STAAR in the spring of the 2013-14 school year.[36] These calculations are further disaggregated by two additional student attributes: the number of minutes a student used the system in 2013-14 prior to the first administration of STAAR, and the performance tier associated with his or her BOY ISIP assessment (Tier 1 are scores in the top 20% of scores in the respective grade and month, and Tier 3 are scores in the bottom 20%).[37] Only students who took an ISIP assessment in September 2014 are included in the calculations.[38]

Student performance on the BOY ISIP assessment is strongly associated with STAAR-Reading performance during the same school year, and the relationship is stable across all grade levels (see Figure 2.7). This suggests that the material tested on the assessment is closely linked to content tested on the corresponding grade-level STAAR-Reading test, and that this instrument can provide valuable diagnostic information for campus staff. This finding corresponds to Patarapichayatham, Fahle, and Roden (2014), although their sample included only a single district in Texas in 2012-13.[39] Nearly 100% of students in the top tier of each grade level ISIP assessment met the STAAR-Reading passing standard, and nearly 80% of

---

[36] Henceforth this will be referred to as the "STAAR passing standard" for brevity.

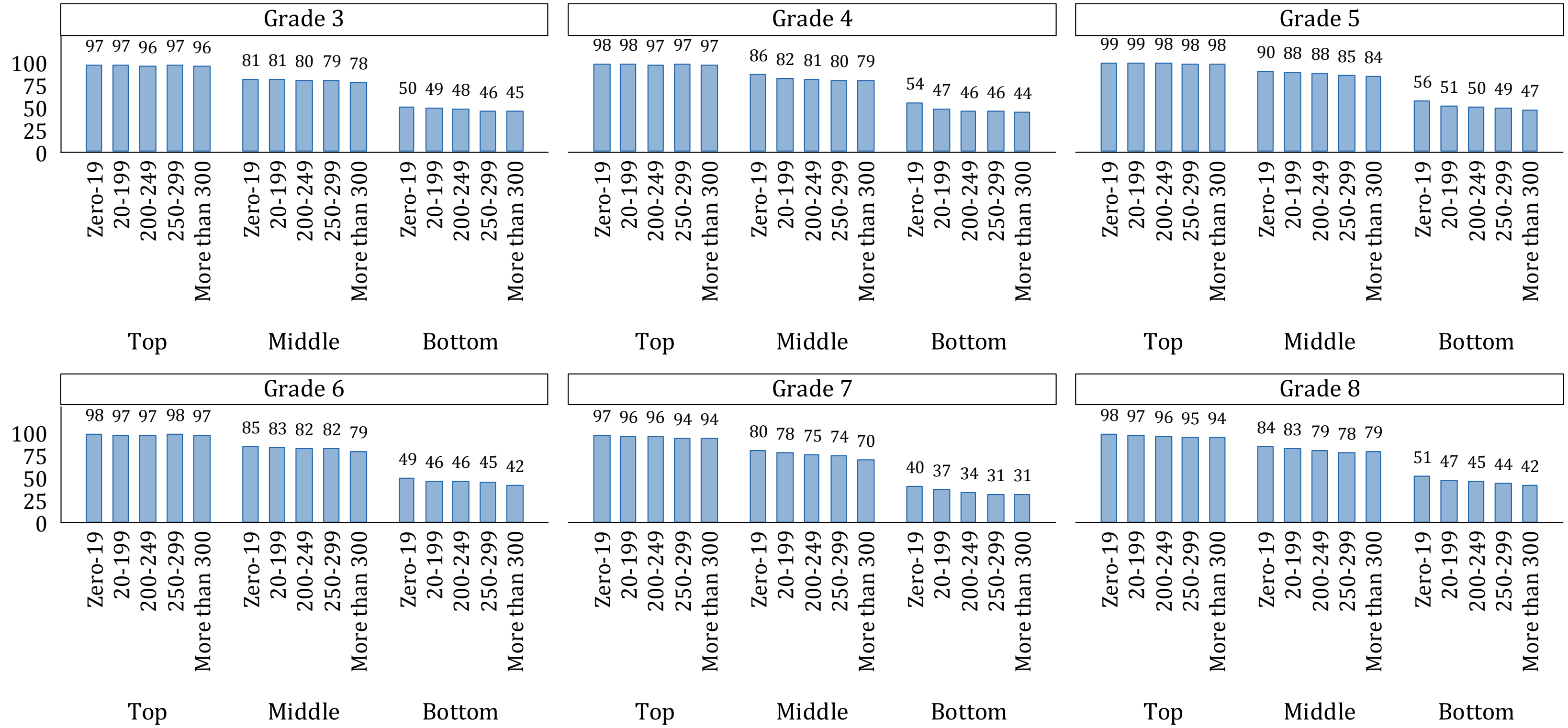
[37] Istation staff recommend 250 minutes for students in Grades 3-5, and 200 minutes for students in Grades 6-8. Minutes were summed prior to the first test administration, which was April 2nd in Grades 5 and 8, and April 23rd in Grades 3, 4, 6, and 7.

[38] Only ISIP assessments that were taken in September are included for two reasons. First, September was the first full month of the school year and, thus, a student's score represents his or her baseline measure of performance on the assessment. Second, and related to the first reason, it standardizes the baseline comparison to ensure that students taking the same assessment, within the same grade level, are taking the assessment at the same time during the school year. This is important, since ISIP performance is strongly correlated with the amount of time that has elapsed in the school year.

[39] To examine the reliability of estimates of the relationship between ISIP scores and performance on STAAR, the evaluation team attempted to replicate the correlation coefficients reported in Patarapichayatham et al. (2014) using 2013-14 data for the same school district. The coefficients for the correlation between BOY ISIP scores and STAAR scaled scores ranged between .68 and .71 for Grades 3-8, and were within one to two hundredths of a decimal point of the coefficients reported in Patarapichayatham et al. (2014).

students in the middle tier met the passing standard. Students in the lowest tier (that is, in the bottom 20% of scores in the respective grade level ISIP assessment in September), however, were considerably less likely to earn a passing score on STAAR-Reading. In addition, students in the lowest performance tier who used Istation more intensively were less likely to meet the STAAR-Reading passing standard. For instance, in Grade 5, 56% of students who logged fewer than 20 minutes in Istation met the STAAR-Reading passing standard, compared to 47% of students who used the system 300 minutes or more. The largest difference in the percentage of bottom tier students who met the passing standard between low and high dosage students was in Grades 4 (10 percentage points), 7 and 8 (9 percentage points).

Figure 2.7. Percentage of Students who Met the Phase-In 1 Level II (Satisfactory) Standard on STAAR-Reading, by Grade Level, Number of Istation Curriculum Minutes, and September 2013 ISIP Assessment Performance Tier, 2013-14 School Year



Source: Istation assessment and session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish versions only (i.e., not modified or alternate versions). Calculations included only students who took an ISIPER or ISIPAR assessment in September 2013. Labels “Top”, “Middle”, and “Bottom” in each graph refer to performance tiers on the beginning of year (BOY) ISIP assessment. Curriculum usage minutes are summed prior to the first STAAR test administration; the month of STAAR first administration is contingent upon grade level.

Relationship between Istation Usage and Changes in Student Performance on STAAR-Reading

The evaluation team calculated decile-standardized STAAR-Reading test-score gains between the first test administration in 2012-13 and the test in 2013-14.[40] The decile-standardized scores have a mean of zero and a standard deviation of 1 for each decile based on students' prior-year (2012-13) STAAR-Reading score, in each grade and school year, so that a positive score indicates that a student's performance *gain* was higher than the average *gain* of other students in that grade, year, and prior STAAR-Reading score decile, while a negative score indicates a student's *gain* was lower than the mean *gain* of other students in the same grade, year, and *prior STAAR-Reading score decile*. The quantity of interest for the evaluation is the *difference* in the gain score between Istation users and non-users. This standardization facilitates comparisons across grade levels for different grade-level assessments because the scale is standardized to have equal means and variance for each assessment, in each year and grade level.

Table 2.17 provides the number of students, by grade level, for which both grade and year-standardized and decile standardized gain scores could be calculated. Gain scores could not be calculated for Grade 3 students, since STAAR-Reading tests are only administered in Grades 3-8 and no comparable prior-year measure of student performance was available. Consequently the association between Istation program participation and STAAR-Reading performance in 2013-14 could not be calculated for Grade 3 students.

Table 2.17. Counts of Students with Valid STAAR-Reading Gain Scores between 2012-13 and 2013-14, by Grade Level

Student Grade Level	Count of Students without Valid Gain Scores	Count of Students with Valid Gain Scores	Total
Grade 4	48,197	335,189	383,386
Grade 5	50,330	332,324	382,654
Grade 6	48,120	327,846	375,966
Grade 7	48,667	336,023	384,690
Grade 8	48,981	330,669	379,650
Total	**244,295**	**1,662,051**	**1,906,346**

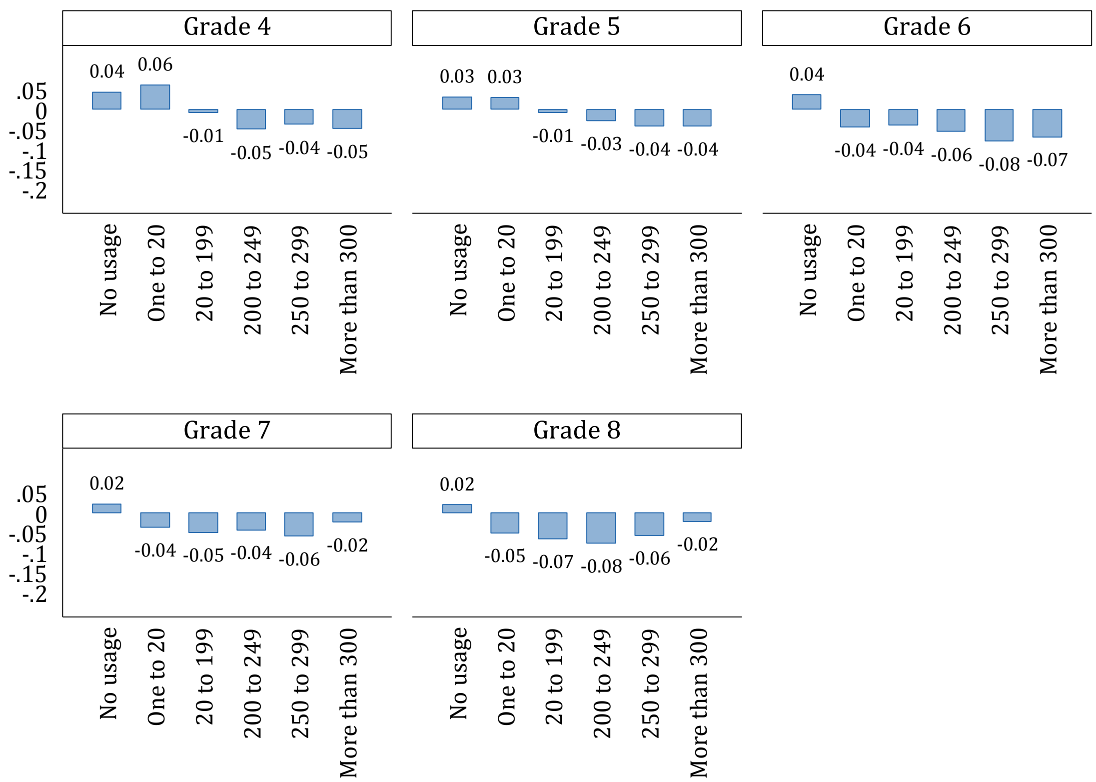
Source: State of Texas Assessments of Academic Readiness data, 2012-13 and 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level included the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions). Only students with consecutive grade level assessments (e.g., Grade 3 in 2012-13 and Grade 4 in 2013-14) were included in the calculations. Counts included students who were Istation users and non-users.

[40] More technical detail about the rationale for decile-standardization of the gain score, and the construction of the measure, can be found in Appendix A.

Figure 2.8 shows the decile-standardized gain scores, respectively, between 2012-13 and 2013-14. The evaluation team calculated the average gain score by grade level and by the amount of time a student spent using Istation curriculum lessons in 2013-14 prior to the first STAAR administration. Across each grade level, non-users, or students who used the system 20 minutes or less, showed larger decile-standardized performance gains than students who used the system at higher dosage levels. In Grade 5, students who used the system more than 250 minutes had lower decile-standardized gain scores than students at every other dosage level.

Figure 2.8. Mean Prior Achievement Decile, Grade and Year Standardized Gain Scores between 2012-13 and 2013-14, by Grade Level and Number of Istation Curriculum Minutes, 2013-14



Source: Istation session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level included the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions). Calculations only included students who took an ISIPER or ISIPAR assessment in September 2013 and who had at least two years of state assessment data available.

Although these descriptive results are an important first step in understanding the relationship between Istation usage and student STAAR-Reading performance, they do not take into account other student, school, or district-level characteristics that are measurable with available administrative data and that may contribute to students' test performance. Thus, this descriptive evidence showing that increased

Istation system usage was associated with lower decile-standardized gain scores compared to non-users should not be interpreted as implying a causal link between the two.[41] In the next section, results from multivariate analyses that adjust for observable student, school, and district characteristics that may distort the descriptive mean difference are presented.

Multivariate Analysis of the Relationship between Istation Participation and Student Performance

Outcomes analyses for this research question focused on decile-standardized gains between 2012-13 and 2013-14 on the STAAR-Reading assessment. The evaluation team fit a number of regression models, all of which are not presented in the body of the report, but are placed in Appendix A.[42] As discussed in more detail in the introductory paragraph to this section, a quasi-experimental DiD design with multivariate regression was used to minimize the observable and unobservable differences between students who participated and those who did not.[43] This approach was used in this evaluation. In addition, in section 7a.2, the extent to which the effect of participation varied across three types of student groups is presented: student ethnicity, economic disadvantaged status, and ELL status. Last, in section 7a.3, the evaluation team defined Istation participation in several different ways to investigate whether STAAR-Reading participation was sensitive to participation dosage –that is, the amount of time a student spent using curriculum sessions in Istation. More detail about the supplementary analyses conducted by the evaluation team, as well as additional technical material describing the multivariate model presented and discussed below, can be found in Appendix A, Research Question 7a.

Model Results

Basic information about the controls and outcome measure used in each model is placed at the bottom of each table, with more technical detail placed in the corresponding appendix (Appendix A, 7a.1). Furthermore, for Guiding Research Question 7a, the measure of the relationship between program participation and student STAAR-Reading performance was confined to two versions of program participation: whether the student used a curriculum session in Istation for at least one minute and a continuous count of the total number of curriculum sessions logged by students in 2013-14. To assess the association between dosage and students' STAAR-Reading performance, additional measures were investigated for Guiding Research Question 7a.3.

---

[41] Although this is particularly true in the simple comparative mean difference analysis conducted in this section, this also remains true for the multivariate results presented in the next section. The limitations of those analyses, and the threats to the internal validity of research design, will be discussed in the Limitations section in Appendix E.

[42] All of the models were fit using ordinary least squares (OLS) regression, because the outcome measure across research questions was continuous. Additional detail can be found in Appendix A.

[43] Even these methods are no panacea. For instance, a crucial element missing from estimates of participation in Istation and STAAR-Reading performance is teacher quality. This will be discussed in more detail in the *Limitations* discussion that bookends this evaluation report.

Table 2.18 presents the multivariate results estimating the relationship between Istation usage and decile-standardized gain scores between 2012-13 and 2013-14 (i.e., results that took into account students' prior STAAR-Reading scores). To review, the measure standardized students' gain scores by their prior performance level. That is, the interpretation of the results in Table 2.18 incorporated an additional element of standardization: prior-achievement decile. For instance, comparison of gain score differences between participating and non-participating students were relative to students in the same grade and from the same prior-achievement decile in 2012-13. Furthermore, to account for differences between participating and non-participating students, the evaluation team adjusted for an array of student and school-level characteristics in the multivariate model.[44]

The results presented in Table 2.18 indicate a statistically significant negative relationship between Istation usage and decile-standardized gain scores in elementary grades, where the largest negative difference between students who used Istation at least one minute relative to students who did not use the system at all was .052 standard deviations in Grade 4. Interestingly, the magnitude of the difference between Istation users and non-users was smaller in Grade 5 (.035) and Grade 6 (.028) relative to the difference in Grade 4. However, in both Grades 7 and 8, students who used Istation at least one minute experienced larger achievement gains than non-users. The largest, positive difference between at least one minute of usage and student gain scores is in Grade 8 (.048 standard deviations), which is nearly twice as large as the effect for Grade 7 students (.027). Last, the measure of Istation usage based on the number of curriculum hours a student logged during the 2013-14 school year generally corresponds with the finding derived from the binary measure of program participation. That is, in Grades 7 and 8, as the number of Istation curriculum hours increased, the magnitude of the decile-standardized reading gain score relative to lower level of system usage increased, while the relationship was reversed in elementary grades. In sum, the multivariate results suggest that Istation users in Grades 7 and 8 experienced statistically significant, albeit substantively small, larger achievement gains compared to non-users, while the gains among elementary grade students who used the system were smaller relative to non-users.[45]

[44] A full list of the covariates included in the statistical model can be found in Appendix A.

[45] Throughout this report, the research team assessed the statistical reliability of the estimates from the multivariate models as being either statistically significant or not statistically significant. This refers to the confidence in concluding that a relationship between our outcome (here, student test scores) and our variable of substantive interest (here, SUCCESS program participation) exists, or whether it was due purely to chance or sample aberrations. The research team adopted the conventional minimum threshold of *p<.05*, meaning that the likelihood of obtaining the association when no association exists, was less than five percent.

Table 2.18. Estimated Effects of Istation Usage on STAAR-Reading Test Score Gains between 2012-13 and 2013-14, by Student Grade Level and Measure of Program Participation

		B	SE	N
Grade 4	Curriculum hours	***-0.003****	0.001	335,089
	Used at least one minute	***-0.052****	0.007	
Grade 5	Curriculum hours	***-0.002**	0.001	332,264
	Used at least one minute	***-0.035****	0.007	
Grade 6	Curriculum hours	0.001	0.001	327,800
	Used at least one minute	***-0.028****	0.008	
Grade 7	Curriculum hours	**0.009*****	0.001	335,966
	Used at least one minute	**0.027*****	0.007	
Grade 8	Curriculum hours	**0.010*****	0.001	330,631
	Used at least one minute	**0.048*****	0.008	
Lagged STAAR-Reading Score		No		
Campus Fixed Effects?		Yes		
Campus Random Effects		No		
Propensity Score Weighting?		No		
Outcome Measure		Prior achievement decile-standardized gain score		

Source: Istation session history tables and State of Texas Assessments of Academic Readiness data, 2012-13-2013-14, Texas Education Agency.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in the analyses. Estimates derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. Statistically significant positive coefficients are denoted by bold font, and negative coefficients are denoted by bold and italicized font. * p<0.05, ** p<0.01, *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Guiding Research Question 7a.1: Summary of Findings

Istation assessments (i.e., ISIP) were strongly related to performance on the STAAR test, suggesting that the instrument was an important diagnostic tool for assessing students' preparation for STAAR. About half of students in the lowest ISIP performance tier met the STAAR-Reading phase-in 1 Level II (Satisfactory) standard in 2013-14, compared to more than 95% of students in the top performance tier.

Descriptive results demonstrate that, among students in the lowest two Istation performance tiers, more frequent Istation use was associated with higher rates of meeting STAAR-Reading state standards than less frequent, or non-use of the Istation system. These differences were largest in Grades 5, 7 and 8, of which Grades 5 and 8 are subject to SSI grade promotion requirements. In most grades, students who did not use Istation demonstrated larger performance gains in reading than students who used Istation. These differences remained even after students' performance in the previous year STAAR-Reading assessment was considered. Further, the performance differences persisted, and in some instances grew, as Istation system dosage increased.

Although these descriptive results are an important first step in understanding the relationship between Istation usage and student STAAR-Reading performance, they do not take into account other student, school, or district-level characteristics that the evaluation team can measure with available administrative data that may contribute to students' test performance. Key findings from statistical models that adjust for student, school, and district characteristics that may distort the descriptive mean difference are presented below:

- Results for students in elementary grades were consistent with descriptive results—that is, Istation use was associated with significantly smaller reading gains among students in Grades 3-6.
- Results were more mixed in Grades 7 and 8, where the direction of relationships between Istation use and reading gains was either not statistically significant, or small and significantly positive.
  - When the STAAR-Reading gain score outcome was measured in light of students' prior performance outcome (i.e., decile-standardized gains between 2012-13 and 2013-14), Istation usage was associated with small, positive associations with reading gains among students in Grades 7 and 8.

Guiding Question 7a.2:

Does the relationship between Istation participation and student performance on STAAR-Reading vary across student groups?

Key Findings:

- **There were few differences in associations between Istation use and students' STAAR-Reading performance gains across student groups of interest.** The effects were mostly stable across student ethnicity, economically disadvantaged status, and ELL status.
- **In Grades 4 and 7, among Istation Reading users, reading gains of students classified as ELL were slightly smaller than non-ELL students.** In this instance, the differences in reading gains were small, at less than one tenth of a standard deviation.
- **In Grade 5, among Istation Reading users, African American students' reading gains were slightly larger than Hispanic students.** Again, this difference was small, at less than one tenth of a standard deviation.

To investigate whether the association between Istation participation and reading outcomes differed across student groups, a slightly modified version of the multivariate model used to address the previous guiding question was used. Specifically, decile-standardized gain scores were used as the outcome variable. Further, measures of participation were confined to three variables: 1) whether the student used Istation for at least one minute compared to students who did not use the system;[46] 2) the total number of hours a student used the system in 2013-14; and 3) whether the student used Istation at the grade-level prescribed dosage compared to students who did not. Student group analysis was restricted to three student characteristics where disparities in state assessment performance were most pronounced: race, ELL identification, and economic disadvantaged status. More information about the statistical models used to address this research question is in Appendix A.

Outcome Results

Differences across student groups of interest, and across grades, were minimal and, largely statistically insignificant (see Table 2.19 and Table 2.20).[47] A few exceptions emerged. In Grades 4 and 7, students classified as ELL showed smaller decile-standardized gains as the number of hours of system usage increased compared to non-ELL students, although the difference in the relationship between these two

[46] Nonusers included students who were, and were not enrolled in the Istation system. An exception was made if the student was enrolled at a school that had no registered users. In that case, the student was not included in the analytic files.

[47] The results for each usage measure and for each student group are presented in the tables. Note that the table includes both statistically significant and insignificant coefficients.

groups was small (.003 and .004 standard deviations of the standardized gain score for each hour increment of usage). Small, inconsistent differences were also found among students identified as economically disadvantaged compared to non-economically disadvantaged students (the effect of increasing hours was positive and statistically significantly different for economically disadvantaged students compared to non-economically disadvantaged students in Grades 4 and 6), but, again, the magnitudes of the differences were substantively small (see Table 2.19). Few differences were found in the effect of dosage on student gain scores between ethnic groups, too, although African American Grade 5 students with higher system dosage demonstrated larger gains compared to Hispanic students in Grade 5 with comparable dosage (see Table 2.20).

While possible explanations for these findings are not evident in the quantitative data alone, information on students' Istation usage combined with information captured via interviews with district- and campus-level staff may provide some direction for further inquiry. Specifically, several interviewees noted that having access to a Spanish-language version of Istation would help them better serve students classified as ELL in upper-elementary (Grades 4 and 5) and middle school grades. These interviewees noted particular difficulties in using the system with ELL students for whom English language acquisition was still in beginning stages. These difficulties—albeit only noted by a handful of interviewees—seem to contradict quantitative Istation usage patterns (Guiding Question 2a), which demonstrated that Istation system usage was higher for students identified as ELL in Grades 4 and 5 (between 35 and 75 minutes more use among ELL students, respectively) and much higher among middle school students where, for example, Grade 7 ELL students used the system over twice as often as non-ELL students (122 minutes versus 58 minutes). Taken together, these findings may indicate that more Istation use is not necessarily more effective, particularly among students for whom the content may not be fully accessible.

Table 2.19. Estimates of the Different Effects of Istation Dosage on STAAR-Reading Test Scores Gains by ELL and Economically Disadvantaged Status, by Student Grade Level and Measure of Program Participation, 2013-14

	Grade 4		Grade 5		Grade 6		Grade 7		Grade 8	
	B	**SE**	**B**	**SE**	**B**	**SE**	**B**	**SE**	**B**	**SE**
ELL versus Non-ELL										
Curriculum system hours	***-0.003*****	0.001	0	0.001	0.001	0.002	***-0.004*****	0.002	-0.002	0.002
Used at least one minute (compared to no usage)	***-0.026****	0.013	-0.014	0.014	-0.000	0.016	-0.006	0.013	0.024	0.016
Used at threshold (compared to below)	***-0.031*****	0.013	-0.01	0.014	0.009	0.02	-0.031	0.02	-0.013	0.022
Economically Disadvantaged versus not Economically Disadvantaged										
Curriculum system hours	**0.002***	0.001	0.001	0.001	**0.003***	0.001	***-0.003****	0.001	-0.002	0.002
Used at least one minute (compared to no usage)	-0.002	0.009	0.01	0.009	0.010	0.009	0.007	0.01	0.018	0.01
Used at threshold (compared to below)	0.017	0.009	-0.001	0.01	0.018	0.013	-0.020	0.015	-0.022	0.021
Campus Fixed Effects?	Yes									
Campus Random Effects?	No									
Propensity Score Weighting?	No									
Outcome Measure	Decile-standardized gain score									

Source: State of Texas Assessments of Academic Readiness and Public Education Information Management System data, 2011-12 – 2013-14, Texas Education Agency. Istation session history table, 2013-14.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in the analyses. Hours were prior to the administration of the first assessment, which was contingent on grade level. Estimates were derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. Statistically significant positive coefficients are denoted by bold font, and negative coefficients are denoted by bold and italicized font. * p<0.05, ** p<0.01, *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Table 2.20. Estimates of the Different Effects of Istation Dosage on STAAR-Reading Test Score Gains by Racial Group, Student Grade Level, and Measure of Program Participation, 2013-14

Race Contrasts										
	Grade 4		Grade 5		Grade 6		Grade 7		Grade 8	
	B	SE	B	SE	B	SE	B	SE	B	SE
White compared to Hispanic										
Curriculum system hours	-0.002	0.001	0	0.001	***-0.004***	0.001	0.001	0.002	-0.000	0.002
Used at least one minute (compared to no usage)	-0.003	0.010	0.004	0.01	-0.004	0.01	-0.018	0.011	-0.018	0.012
Used at threshold (compared to below)	-0.016	0.011	0.006	0.012	-0.027	0.015	-0.003	0.02	-0.019	0.024
African American compared to Hispanic										
Curriculum system hours	0.002	0.001	**0.004****	0.001	0.001	0.002	0	0.002	-0.001	0.003
Used at least one minute (compared to no usage)	0.001	0.014	0.015	0.014	0.000	0.013	0.007	0.011	0.002	0.013
Used at threshold (compared to below)	0.004	0.014	**0.042****	0.014	-0.007	0.018	0.012	0.021	0.021	0.026
N	335,089		332,264		327,800		335,966		330,631	
Campus Fixed Effects?	Yes									
Campus Random Effects?	No									
Propensity Score Weighting?	No									
Outcome Measure	Decile-standardized gain score									

Source: State of Texas Assessments of Academic Readiness and Public Education Information Management System data, 2011-12 – 2013-14, Texas Education Agency. Istation session history table, 2013-14.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in the analyses. Hours were prior to the administration of the first assessment, which was contingent on grade level. Estimates were derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. Statistically significant positive coefficients are denoted by bold font, and negative coefficients are denoted by bold and italicized font. * p<0.05, ** p<0.01, *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Guiding Research Question 7a.2: Summary of Findings

To address whether relationships between reading gains and different measures of Istation usage varied by student groups, three separate statistical models were used. The models examined the relationship between decile-standardized gain scores (i.e., performance levels on prior year STAAR-Reading assessments) and students' ethnicity, ELL status, and economically disadvantaged status, and allowed the effect of these student group indicators to vary across Istation dosage levels (i.e., the number of minutes spent on Istation curriculum sessions). Differences across student groups of interest, and across grades, were minimal and largely statistically insignificant. A few exceptions emerged:

- In Grades 4 and 7, students identified as ELL showed smaller decile-standardized gains compared to non-ELL students as the number of hours of system usage increased, although the difference in the relationship between these two groups was small. This suggests that additional time on the system was associated with reading declines among ELL students more than non-ELL students. That said, the practical significance of these associations—represented in terms of standard deviations in decile-standardized scale scores—was quite minor (between -0.003 and 0.004 standard deviations).
- In Grade 5, African American students with higher system dosage demonstrated significantly larger, positive gains than Hispanic students in Grade 5 with comparable dosage. Qualitatively, some interviewees did note that the lack of a Spanish-language version of Istation in upper elementary and middle school grades was a barrier to serving students classified as ELL, which may be noteworthy given that ELL students typically used the system more frequently than their non-ELL peers.

Guiding Question 7a.3:

Is the relationship between Istation participation and student performance on STAAR-Reading conditioned on the intensity of usage?

Key Findings:

- **In Grades 3-5, students who used Istation more intensely demonstrated smaller STAAR-Reading gains compared to students who used the system at lower levels, or who did not use the system at all.** These differences increased as the number of minutes using the system increased, and remained after adjusting for student and school-level characteristics.
- **The effects of Istation dosage on STAAR-Reading gains were substantively very small.** In elementary grades, they were less than 5% of a standard deviation.
- **Istation dosage was associated with significantly larger STAAR-Reading score gains in Grades 7 and 8.** For Grade 7, the gains were largest for students who used the system 500 minutes or more, relative to students who did not use the system; and for Grade 8, the gains were largest for students who used the system 400 minutes or more, relative to students who did not use the system. These differences remained after adjusting for student and school-level characteristics. As with previously observed relationships between Istation use and reading outcomes, these effects remained substantively small. The size of the effect never exceeded 20% of a standard deviation of the decile-standardized gain scores.

To further explore associations between Istation use and reading gains, different—and more precise—measures of program participation were developed to reflect *incremental* increases and decreases in Istation usage. The benefit of these measures is that they allowed for differential relationships between different levels of Istation dosage and STAAR-Reading outcomes to be examined.

The evaluation team explored the relationship between dosage and STAAR-Reading performance using a variety of approaches, with Methods 3 and 4 inspired by Heinrich, Meyer, and Whitten (2010).[48]

1. Method 1: Met recommended usage threshold at or above the respective grade level (250 minutes for students in Grades 3-5, and 200 minutes for students in Grades 6-8) compared to students who did not meet the threshold
2. Method 2: Used the system 300, 400, 500, 600, or more than 700 minutes *compared to students who did not use the system*

[48] The results and discussion for Methods 3 and 4, which were fit to explore the robustness of the relationship between dosage level and STAAR-Reading outcomes, are provided in Appendix 7a.3. Additional technical errata outlining this method and its implementation in this evaluation are provided in Appendix 7a.1.

3. Method 3 (Appendix 7a.3): Met recommended usage threshold at the respective grade level (250 minutes for students in Grades 3-5, and 200 minutes for students in Grades 6-8) compared to students *who did not use the system*
4. Method 4 (Appendix 7a.3): Used the system 500 minutes or more compared to students *who did not use the system*
5. Method 5 (Appendix 7a.3): Met recommended usage threshold at the respective grade level (250 minutes for students in Grades 3-5, and 200 minutes for students in Grades 6-8) compared to students who did not meet the threshold

The results from Methods 1 and 2 are presented in Table 2.21, and the results from Methods 3, 4, and 5 are provided in Appendix 7a.3, in Table A5. The multivariate model used to estimate the relationship between Istation dosage and STAAR-Reading performance gains was identical to the technique used in the response to Guiding Research Question 7a, and employed decile-standardized gain scores as the outcome measure.

Outcome Results

Across all methods, in Grades 4 and 5, increasing Istation dosage levels did not produce larger STAAR-Reading gains relative to non-users. In Grades 4 and 5, each incrementally higher level of dosage was negatively related to student test score gains between 2012-13 and 2013-14, although the magnitude of the relationships was small, never exceeding one tenth of a decile-standardized gain score. This means that students in Grades 4 and 5 did not receive any academic benefits based on STAAR-Reading gains relative to non-users as their dosage increased; in fact, the gains of high-dosage students were weaker than non-users. For instance, in Grade 4, the decile-standardized performance gains of students who met the Istation usage at or above the threshold (250 minutes) were .025 standard deviations smaller than students below the recommended usage level. This negative difference grew to .076 standard deviations when comparing students who used the system 700 minutes or more to students who did not use the system. This pattern was weaker, although statistically significant, among Grade 5 students (.051 standard deviations lower).

Dosage effects for Grade 7 and 8 students, however, were positive and statistically significant across each measure of usage intensity, although the differences remained small. For instance, Grade 7 and 8 students who used the system at or above the recommended dosage threshold (200 minutes) experienced larger decile-standardized gains relative to students who used the system below the dosage threshold (.081 and .093 standard deviations higher). The magnitude of the difference between users and non-users grew as usage intensity increased among middle grade students. For instance, in Grade 7, the decile-standardized gain scores of students who used the system between 600 and 699 minutes were .135 standard deviations greater than students who did not use the system. Furthermore, in Grade 8, the impact of dosage on decile-standardized-gain scores nearly doubled from .071 standard deviations for students with between 300 and 399 minutes, to between .141 and .173 standard deviations among students with 400 or more minutes using the system. It is important to keep in mind, however, that the number of students who fall into the most extreme dosage categories in Grades 7 and 8 is a very small percentage of the total grade-level population. For instance, approximately 3% (around 14,000 students) recorded 300 or more minutes of curriculum session time during the 2013-14 school year. Consequently, these findings should be interpreted with this caveat in mind. (See Table 2.21.)

Table 2.21. Estimates of the Relationship between Intensity of Istation Usage and STAAR-Reading Decile-standardized Gain Scores, by Grade Level and Measure of Usage Intensity, 2013-14

	Method	Usage level	B	SE	N
Grade 4	Method 1	At or above threshold (compared to below)	***-0.025****	0.007	335,089
	Method 2	300 to 399 minutes (compared to none)	***-0.073****	0.011	
		400 to 499 minutes (compared to none)	***-0.063****	0.012	
		500 to 599 minutes (compared to none)	***-0.056****	0.013	
		600 to 699 minutes (compared to none)	***-0.075****	0.015	
		More than 700 minutes	***-0.076****	0.013	
Grade 5	Method 1	At or above threshold (compared to below)	***-0.025****	0.007	332,264
	Method 2	300 to 399 minutes (compared to none)	***-0.050****	0.012	
		400 to 499 minutes (compared to none)	***-0.054****	0.013	
		500 to 599 minutes (compared to none)	***-0.050****	0.015	
		600 to 699 minutes (compared to none)	***-0.064****	0.016	
		More than 700 minutes (compared to none)	***-0.051****	0.014	
Grade 6	Method 1	At or above threshold (compared to below)	-0.006	0.009	327,800
	Method 2	300 to 399 minutes (compared to none)	***-0.032****	0.015	
		400 to 499 minutes (compared to none)	-0.013	0.018	
		500 to 599 minutes (compared to none)	***-0.042****	0.019	
		600 to 699 minutes (compared to none)	***-0.051****	0.022	
		More than 700 minutes (compared to none)	-0.001	0.017	
Grade 7	Method 1	At or above threshold (compared to below)	**0.081*****	0.010	335,966
	Method 2	300 to 399 minutes (compared to none)	**0.069*****	0.017	
		400 to 499 minutes (compared to none)	**0.056****	0.020	
		500 to 599 minutes (compared to none)	**0.081*****	0.023	
		600 to 699 minutes (compared to none)	**0.135*****	0.024	
		More than 700 minutes (compared to none)	**0.164*****	0.019	
Grade 8	Method 1	At or above threshold (compared to below)	**0.093*****	0.012	330,631
	Method 2	300 to 399 minutes (compared to none)	**0.071*****	0.020	
		400 to 499 minutes (compared to none)	**0.173*****	0.027	
		500 to 599 minutes (compared to none)	**0.168*****	0.027	
		600 to 699 minutes (compared to none)	**0.141*****	0.030	

	Method	Usage level	B	SE	N
		More than 700 minutes (compared to none)	**0.154*** **	0.027	
	Campus Fixed Effects?		Yes		
	Campus Random Effects		No		
	Propensity Score Weighting?		No		
	Outcome Measure		decile-standardized gain score		

Source: State of Texas Assessments of Academic Readiness and Public Education Information Management System data, 2011-12 – 2013-14, Texas Education Agency. Istation session history table, 2013-14.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in the analyses. Hours were calculated prior to the administration of the first assessment, which was contingent on grade level. Estimates were derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. Statistically significant positive coefficients are denoted by bold font, and negative coefficients are denoted by bold and italicized font.* p<0.05; ** p<0.01; *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Guiding Research Question 7a.3: Summary of Findings

To further the exploration of associations between Istation use and reading gains, more precise measures of program participation were developed to reflect incremental increases and decreases in Istation usage. The benefit of these measures is that they allowed for differential relationships between different levels of Istation dosage and STAAR-Reading outcomes to be examined. In these analysis, Istation users were compared to students who had similar prior year STAAR-Reading performance levels.

Overall, in Grades 4 and 5, increasing Istation dosage levels did not produce larger STAAR-Reading gains relative to non-users, and students who met the recommended dosage level had smaller STAAR-Reading gains than students who surpassed this level. In these grades, each incremental increase in the number of minutes a student spent on the Istation system was negatively related to student test score gains between 2012-13 and 2013-14. However, it is important to keep in mind that differences were substantively small. Grade 4 and 5 students who used the Istation system did not receive academic benefits based on STAAR-Reading gains relative to non-users as their dosage increased.

Dosage effects for Grade 7 and 8 students, however, were positive and statistically significant across each measure of usage intensity, although the differences again remained small. The size of the difference in gain scores between users and non-users grew as usage intensity (i.e., minutes on the system) increased among middle grade students. For instance, in Grade 7, the decile-standardized gain scores of students who used the system between 600 and 699 minutes were .135 standard deviations greater than the scores for students who did not use the system. It is important to keep in mind, however, that the number of students who used the system for 300 minutes or more was very low (approximately 3% of the grade-level population). Thus, these findings should be interpreted with this in mind.

Guiding Question 8a:

What is the relationship between the timing of system usage (i.e., the proximity to the STAAR test) and usage levels (i.e., dosage as measured by number of minutes on the system) and performance on the first administration of the STAAR-Reading exam?

Key Findings:

- **Consistent Istation system usage throughout the school year yielded small, positive STAAR-Reading gains in Grades 7 and 8; more intermittent system use did not.** The more months in which students in Grades 7 and 8 recorded at least 30 minutes of Istation use, the larger the improvements in STAAR-Reading scores, although the differences were substantively small. Consistent Istation system usage was negatively associated with STAAR-Reading gains in elementary grades.
- **In middle grades, usage intensity in the month prior to the STAAR-Reading exam was weakly and positively associated with STAAR-Reading scaled score gains, although the relationship was not statistically significant.** In Grades 4, 5, and 6, the relationship was negative, although only statistically significant in Grade 4. These effects were small, however.

In this section, the relationship between the timing of usage throughout the school year (that is, whether the student used the system only in September 2013, or whether the student used the system in each month of the school year) and student gain scores is explored. Other measures of usage based on total time spent using the system, or whether the student met the usage threshold recommended by Istation, do not capture the proximity of usage to the STAAR test, nor do they reflect the continuity of usage throughout the school year. Student performance may be more sensitive to an intervention—such as Istation—that occurs more closely to when the outcome is measured.

Using monthly Istation usage data, three measures were created to investigate whether usage proximity to the test administration and continuity throughout the school year were associated with STAAR-Reading gains in 2013-14.

First, to capture **usage continuity**, whether a student recorded 30 minutes or more of Istation usage in a month was calculated for each grade and month. A 30-minute target was used because it, represents, approximately, the amount of time a student would need to use the system each month to meet the usage threshold recommended by Istation. The monthly flags were then summed to create a count of the total number of months in which the students' usage met or exceeded 30 minutes.

Second, to assess **usage proximity to the test**, the total number of minutes in the month *prior to the STAAR-Reading* test was used as a covariate in the multivariate model to investigate differences in the relationship between dosage proximity and changes in student performance between 2012-13 and 2013-14.[49] Monthly minute totals were then added to the multivariate models as covariates.

Third, for each month, as a measure of **intensity of usage,** the proportion of each student's total usage prior to the first administration of the STAAR-Reading exam that occurred in each month was calculated. For example, if Student A recorded 10 minutes of Istation usage throughout the entire school year, with 5 minutes logged in January and 5 minutes logged in March, the proportion of total usage that occurred in January and March would be 0.5. Then, the proportion of usage that occurred *three months prior to the test administration* (e.g., for students in Grades 5 and 8, this was January, February, and March 2014 and, using the contrived example for Student A, the total proportion would be 1.0) was summed to calculate the share of a student's total Istation usage that occurred near the test administration. This is used as covariate in the statistical models, and the association between this measure and changes in STAAR-Reading performance was estimated between students who met the grade-recommended usage threshold during the 2013-14 school year, and students who did not. In sum, this approach explored the joint relationship between proximity and usage intensity and accounted for the fact that the simple proportion measure disregards the total system dosage a student experienced.[50]

Outcome Results

Table 2.22 displays the results from the multivariate models.[51] Models were fit separately for each measure of usage proximity, continuity, and grade level. In Grades 4 and 5, as the number of months in which a student used the system at least 30 minutes rose, their decile-standardized gain scores were smaller (.009 and .007 standard deviations, respectively), suggesting that consistent usage in elementary grades led to lower average decile-standardized reading gains than inconsistent system usage. This relationship was statistically significant, although the magnitude of the relationship was small.[52] However, the relationship was reversed, and much larger and positive, for Grade 7 and 8 students. Students in Grades 7 and 8 who used the system consistently across the school year experienced larger decile-standardized gain scores than students who used the system less consistently across the school year. For each additional month, gain scores increased by 2% of a standard deviation (.021) for Grade 7 students, and 3% of a standard deviation (.026) for Grade 8 students.

In Grades 4 and 5, system usage intensity in the month immediately prior to the STAAR administration was negatively associated with student performance, although this relationship was weakly statistically

[49] For Grade 5 and 8 students, minutes were summed from August 2013 through March 2014, and for Grade 4, 6, and 7 students, minutes were summed from August 2013 through April 2014, because the STAAR-Reading first administration dates varied by grade level.

[50] Put another way, a student who only used the system 10 minutes in 2013-14, but who accumulated those minutes in March, the month before the STAAR administration date, would be weighted equivalently to a student who used the system 200 minutes, and who also recorded those minutes in March, because the proportion of usage that occurred in March would be identical (1.0) for each student.

[51] The technical matter for these statistical models is provided in Appendix A, Guiding Research Question 8.

[52] Significant at the $p < .001$ level.

significant (*p <.10)* for Grade 4 students, and not statistically significant among Grade 5 students. For Grade 7 and 8 students, the relationship was positive, but did not achieve conventional levels of statistical significance, nor did the joint association between usage proximity and dosage, measured by the proportion of usage that occurred in the three months prior to the test administration.

Table 2.22. Estimates of the Relationship between Istation Usage Continuity and Proximity to the STAAR Test and Decile-Standardized Gain Scores, by Grade Level, 2013-14

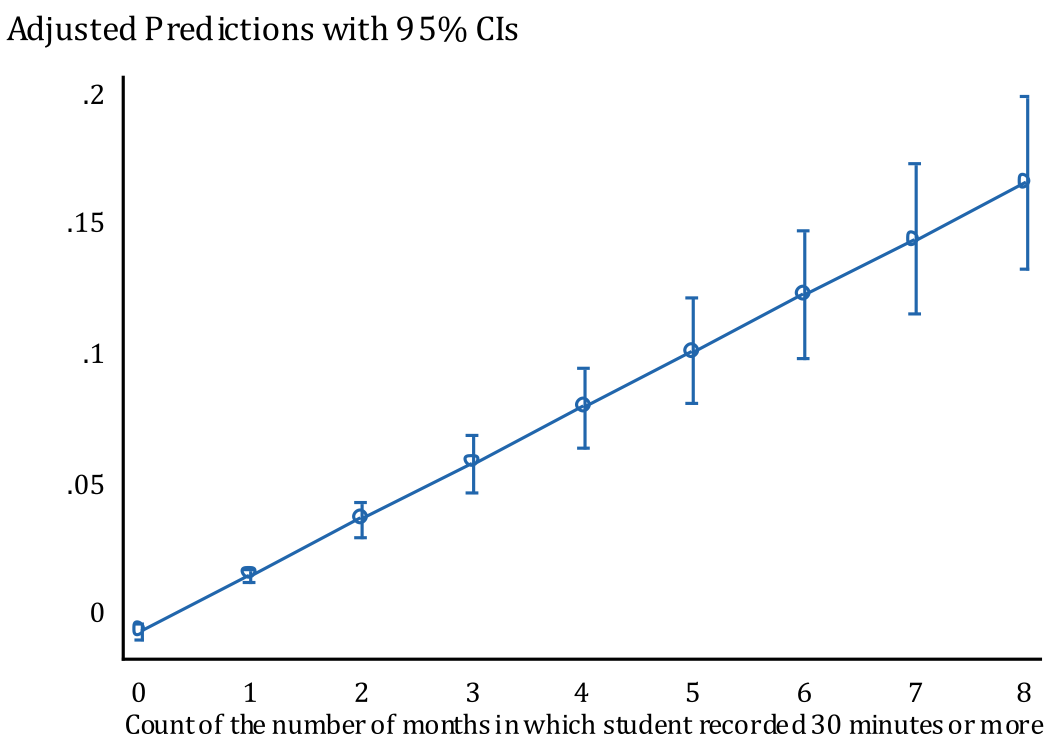
		B	SE	N
Grade 4	Number of months in which student recorded 30 minutes or more	***-0.009****	0.002	335,105
	System usage hours in April 2014	***-0.008****	0.004	
	Proportion of usage occurring in three months prior to STAAR test	-0.000	0.000	
Grade 5	Number of months in which student recorded 30 minutes or more	***-0.007*****	0.002	332,273
	System usage hours in March 2014	-0.007	0.004	
	Proportion of usage occurring in three months prior to STAAR test	0.000	0.000	
Grade 6	Number of months in which student recorded 30 minutes or more	0.001	0.002	327,804
	System usage hours in April 2014	-0.003	0.006	
	Proportion of usage occurring in three months prior to STAAR test	-0.000	0.000	
Grade 7	Number of months in which student recorded 30 minutes or more	**0.021*****	0.002	335,966
	System usage hours in April 2014	0.008	0.007	
	Proportion of usage occurring in three months prior to STAAR test	-0.000	0.000	
Grade 8	Number of months in which student recorded 30 minutes or more	**0.026*****	0.003	330,635
	System usage hours in March 2014	0.012	0.008	
	Proportion of usage occurring in three months prior to STAAR test	0.000	0.000	
Campus Fixed Effects?		Yes		
Campus Random Effects		No		
Propensity Score Weighting?		No		
Outcome Measure		Grade-year-decile standardized gain score		

Source: State of Texas Assessments of Academic Readiness and Public Education Information Management System data, 2011-12 –2013-14, Texas Education Agency. Istation session history table, 2013-14.

Note: For Grade 5 and 8 students, minutes were summed from August 2013 through March 2014, and for Grade 4, 6, and 7 students, minutes were summed from August 2013 through April 2014, because the STAAR-Reading first administration dates varied by grade level. Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in the analyses. Hours were calculated prior to the administration of the first assessment, which was contingent on grade level. Estimates were derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariate. Statistically significant positive coefficients are denoted by bold font, and negative coefficients are denoted by bold and italicized font. * p<0.05; ** p<0.01; *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Figure 2.9 illustrates the relationship between usage continuity across the school year and the predicted decile-standardized gain score for Grade 7 students. Predictive margins were derived from the statistical results used to populate Table 2.22. For each additional month in which a Grade 7 student used Istation for 30 minutes or more, the decile standardized gain score rose .022 standard deviations. Thus, the estimated decile-standardized gain score for Grade 7 students who used the system at least 30 minutes for seven months was .12, compared to .04 for students who used the system at this level of intensity for three months. Taken together, the results suggest that, in Grades 7 and 8, consistent usage of Istation throughout the school year was associated with small but positive gains in STAAR-Reading test scores.

Figure 2.9. Linear Prediction of Decile-Standardized Gain Scores by the Number of Months in which Grade 7 Students Used Istation for at Least 30 Minutes, 2013-14



Source: State of Texas Assessments of Academic Readiness and Public Education Information Management System data, 2011-12 – 2013-14, Texas Education Agency. Istation Session History table, 2013-14.

Guiding Research Question 8a: Summary of Findings

To assess relationships between the consistency and timing of Istation usage throughout the school year and reading outcomes, two additional measures of Istation usage were created: usage proximity to the STAAR test administration and use *continuity* throughout the school year. These measures were entered as independent variables in statistical models to explore joint relationships between the timing (i.e., proximity to STAAR test administration) and intensity of Istation, and reading outcomes. These analyses compared results of Istation users with other students with comparable prior year STAAR-Reading assessment scores.

With regard to findings related to usage continuity:

- In Grades 4 and 5, continuity was significantly related to student test score declines between 2012-13 and 2013-14, although the magnitude of the relationships was substantively very small.
- In Grades 7 and 8, associations between usage continuity and reading gains were positive and substantially larger than reading gains demonstrated in Grades 4 and 5. Specifically, for each additional month of Istation usage, gain scores increased by 2% of a standard deviation for Grade 7 students and 3% of a standard deviation for Grade 8 students.

With regard to findings related to usage proximity to the STAAR test:

- In Grades 4 and 5, system usage intensity in the month immediately prior to the STAAR administration was negatively associated with student performance, though neither relationship achieved conventional levels of statistical significance.
- For Grades 7 and 8, the relationship was positive, but did not achieve conventional levels of statistical significance, nor did the joint association between usage proximity and dosage, measured by the proportion of usage which occurred in the three months prior to the test administration.

Of the two Istation usage indicators captured here (i.e., consistency of usage and proximity to the STAAR assessment, consistent use was more substantially related to reading outcomes than the proximity of usage relative to the STAAR-Reading testing windows.

Guiding Question 9a:

To what extent does performance differ between Texas SUCCESS participants and non-participants among students at risk of being retained?

Key Descriptive Findings:

- **In Grades 5 and 8, students who failed the STAAR-Reading exam in a previous year used Istation at higher levels than students who never failed.** Consistent with patterns across all students (not just those who had previously failed the STAAR exam), usage was most intense among students in Grades 5 and 8. Students in Grade 5 who previously failed STAAR used Istation 87 minutes more (239 versus 152 minutes) than students who had not failed. These differences were larger in Grade 8: students who failed STAAR-Reading in a prior year averaged 53 minutes more (73 versus 20 minutes) than students who had not failed.
- **In Grades 5 and 8, students at risk of being retained experienced smaller gains between 2012-13 and 2013-14 than students not at risk, with the largest differences in Grade 8.** But, among students who were at risk of being retained in grade, Grade 8 students who used the system at higher dosage levels experienced larger gains relative to students who used the system at lower dosage levels, although the differences, while statistically significant, were substantively small.

Key Multivariate Findings:

- **Grade 8 students classified as at risk of being retained in grade, and who used Istation, demonstrated larger gains in reading than students who did not use the system.** Students who used the system longer showed larger gains than students who did not use the system, or who used the system below the level recommended by the vendor. Nonetheless, these differences were substantively small.

As shown in the descriptive analysis of usage patterns described in Research Question 2a, students who performed poorly on STAAR-Reading in prior years were more likely to use Istation, and use it more intensively, compared to students who performed well. This section presents relationships between Istation usage—and usage intensity—and changes in student performance on STAAR-Reading between students at risk of being retained and those not at risk of being retained. Addenda reports that further address outcomes for students at risk of being retained in grade will be provided to TEA in summer 2015. See Section 4 of this report (Additional FY 2015 Analyses) for descriptions of the forthcoming research.

At-risk is defined as students who, in 2011-12 or 2012-13, failed at least one STAAR-Reading assessment. The analysis was confined to students who were in Grades 5 and 8 in 2013-14 (i.e., grades where SSI grade promotion was conditional on passing STAAR exams), and who were enrolled at a school where at least one student was registered in the Istation system. In addition, only students with at least two valid test scores in 2012-13 and 2013-14 were included, eliminating students for whom the evaluation team had no information on prior test performance. Table 2.23 displays the count of students in each grade level who were identified as being at risk of being retained in grade between 2013-14 and 2014-15. In both grades, approximately 30% of students in 2013-14 were categorized as being at risk of being retained, and a slightly higher percentage of students in Grade 5 (34%) were at risk relative to students in Grade 8 (31%).

Table 2.23. Frequency Count and Percentage of Students at Risk of Being Retained between 2013-14 and 2014-15

Student Grade Level	Failed STAAR-Reading at Least Once in Prior Year	Count of Students	Percentage of Students
Grade 5	Did not fail	227,787	66.33
Grade 5	Failed at least once	117,152	34.11
Grade 8	Did not fail	237,697	69.22
Grade 8	Failed at least once	105,708	30.78

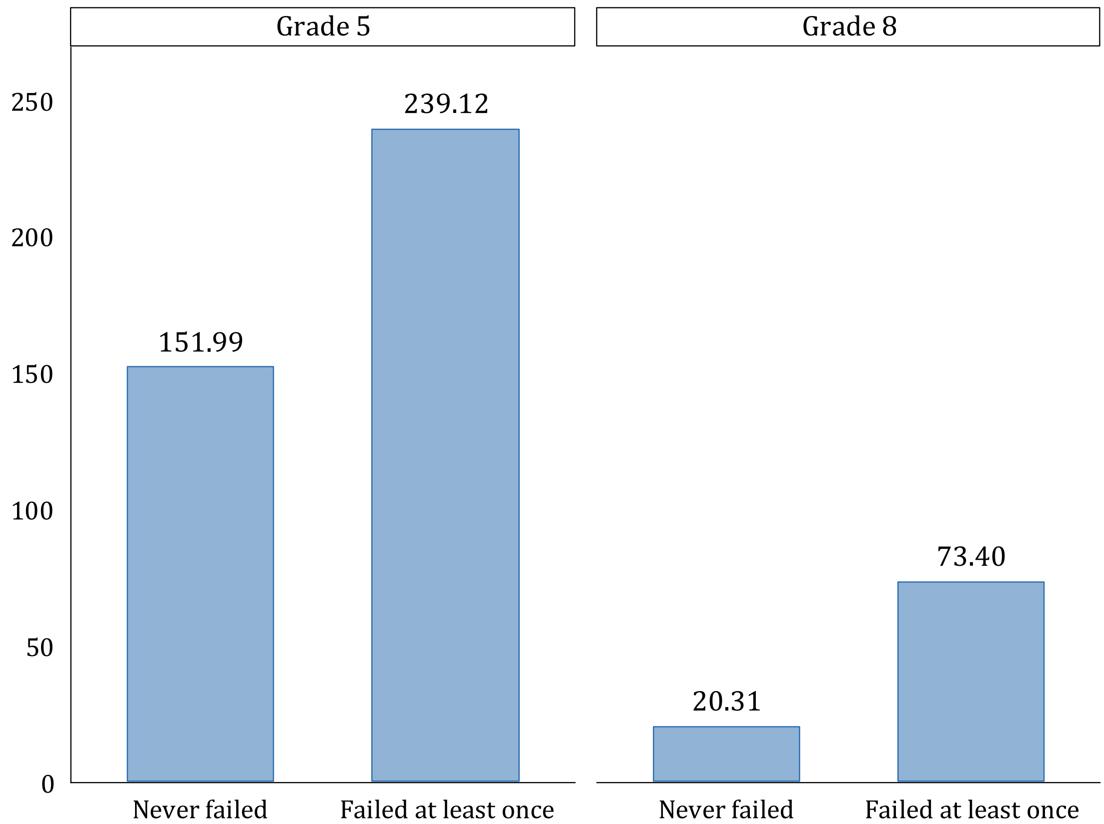
Source: State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: Frequency counts and calculations only include students who were enrolled at a campus where at least one student was registered for Istation in 2013-14 and had a valid STAAR-Reading test score in the prior year (2012-13). Only regular English and Spanish versions (i.e., not modified or alternate versions) were included. STAAR test data were only available from 2011-12 to 2013-14.

Outcome Results

Among Grade 5 and 8 students, students at risk of being retained used the system more intensively in 2013-14 than peers who were not at risk of being retained. As Figure 2.10 illustrates, Grade 5 students who failed STAAR-Reading at least once in a prior year used the system 87 minutes more than students who never failed in a prior school year (239 minutes compared to 152). In Grade 8, students who had failed a STAAR-Reading assessment in a prior year used the system 53 minutes more than students who had never failed (73 minutes compared to 20). This relationship was consistent with the patterns presented in previous sections (see for instance Research Question 2a), which demonstrated a link between prior academic performance and the amount of Istation system usage among middle grade students in 2013-14.

Figure 2.10. Average Number of Istation Minutes in 2013-14, by At-Risk Status and Grade Level, 2013-14



Source: State of Texas Assessments of Academic Readiness data, 2011-12 – 2013-14, Texas Education Agency. Istation session history table, 2013-14.

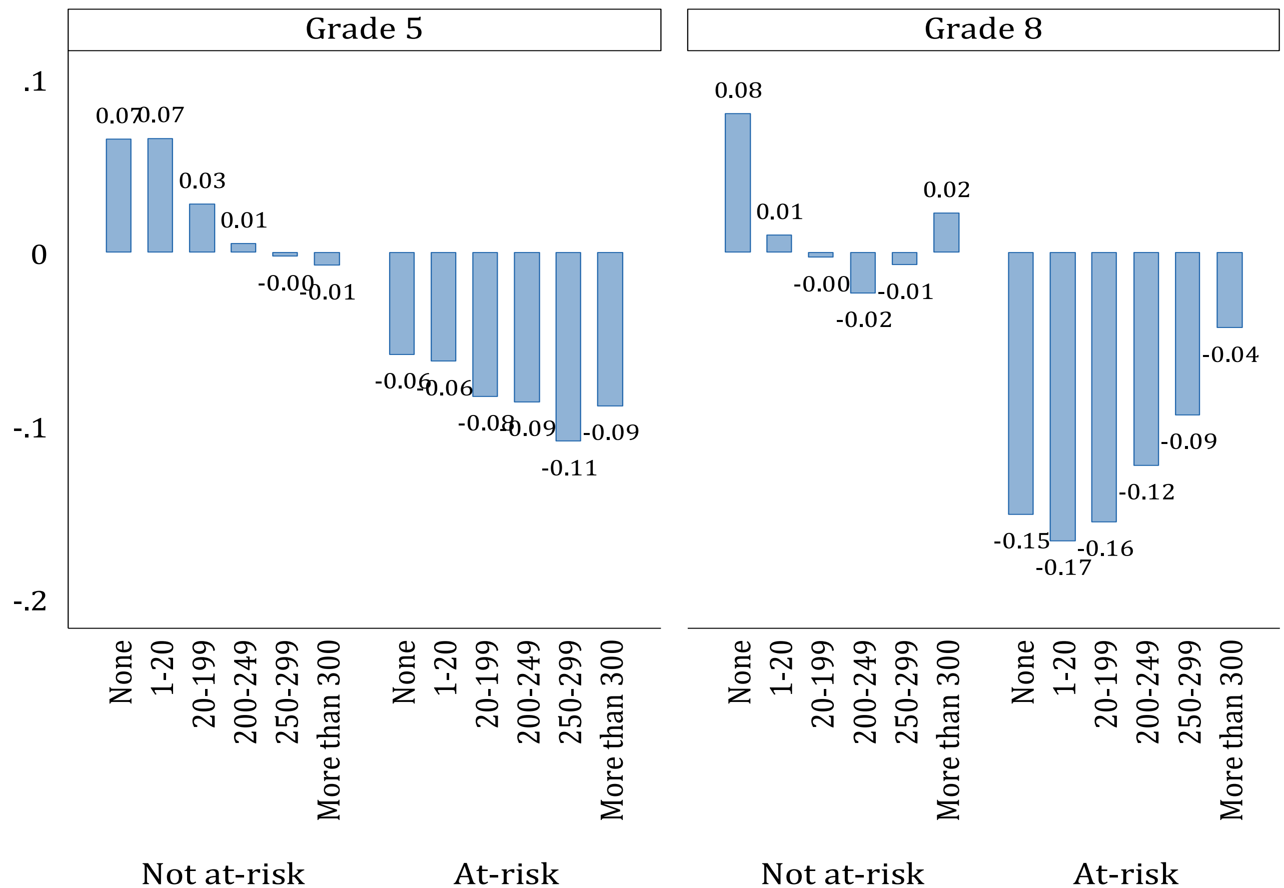
Note: Calculations only include students who were enrolled at a campus where at least one student was registered for Istation in 2013-14 and had a valid STAAR-Reading test score in the prior year (2012-13). Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in the at-risk indicator flag. Minutes were prior to the administration of the first assessment, which was contingent on grade level.

Figure 2.11 displays the relationship between student STAAR-Reading performance—measured as a decile-standardized score gain—and Istation system usage level in 2013-14 to provide a descriptive picture of the interaction between Istation usage in 2013-14, students' at-risk status, and students' STAAR-Reading performance changes between 2012-13 and 2013-14. Istation usage level is divided into five categories: No usage, between 1 and 20 minutes, between 20 and 199 minutes, between 200 and 249 minutes, between 250 and 299 minutes, and 300 or more minutes.

The calculation is disaggregated by an additional student-level attribute: whether the student was classified as being at risk of being retained in grade. In both Grades 5 and 8, students at risk of being retained experienced smaller decile-standardized gains between 2012-13 and 2013-14 compared to students who were not at risk, with the largest disparities occurring among Grade 8 students. Among the population of students who were not at risk of being retained, students who used Istation more intensively demonstrated weaker performance gains as usage intensity increased. This pattern was consistent across both grades.

Among students at risk of being retained, however, students in Grade 8 who used the system 200 or more minutes in 2013-14 experienced *larger* decile-standardized STAAR-Reading gains compared to students who did not use the system, or used the system between 1 and 199 minutes. For instance, the average decile-standardized gain score for students in Grade 8 who used the system more than 300 minutes was .04 standard deviations below the average, compared to .15 standard deviations below the average for students who did not use the system. In Grade 5, students with greater Istation dosage experienced weaker gains than students who did not use the system at all.

Figure 2.11. Mean Decile-Standardized Gain Scores between 2012-13 and 2013-14 on STAAR-Reading, by At-Risk Status, Student Grade Level, and Istation Curriculum System Usage



Source: State of Texas Assessments of Academic Readiness Data, 2011-12 – 2013-14, Texas Education Agency. Istation Session History table, 2013-14.

Note: Calculations only include students were enrolled at a campus where at least one student was registered for Istation in 2013-14 and had a valid STAAR-Reading test score in the prior year (2012-13). Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in the at-risk indicator flag. Minutes were prior to the administration of the first assessment, which was contingent on grade level.

To further investigate the differences in Istation system usage and dosage on student performance between students at risk of being retained and those not at risk, the evaluation team fit a multivariate regression that examined the relationship between separate measures of system dosage on students' decile-standardized STAAR-Reading gain scores between 2012-13 and 2013-14: student used the system at least once (compared to no usage); student used the system 300 minutes or more (compared to no usage); and the student used the system at the dosage recommended by Istation (compared to usage below the prescribed threshold).[53] The results are presented in Table 2.24.

Grade 8 students at risk of being retained who used the system, and who used the system at higher dosage levels, showed larger STAAR-Reading gains compared to students who did not use the system but who were at risk of being retained. This finding was consistent across different measures of system usage. For instance, Grade 8 students at risk of being retained who used the system 300 minutes or more had .105 standard deviations higher decile-standardized change scores compared to students who did not use the system.

Furthermore, across each measure of usage, students at risk of being retained showed larger STAAR-Reading gains compared to students who did not use the system. This finding suggests that, particularly among Grade 8 students, the supplementary instruction provided by Istation for students at risk of being retained in grade yielded greater benefits, as measured by changes in student test scores, relative to students who were not at risk of being retained. While students at risk of being retained in Grade 5 showed larger STAAR-Reading gains relative to their peers who were not at risk of being retained, the differences were only statistically significant among students who met the prescribed usage threshold.

[53] More technical detail about the estimation procedure and functional form is provided in Appendix 9.

Table 2.24. Estimates of the Differences in the Relationship between Students at Risk of Being Retained and Istation System Usage

Grade	Usage Measure	B	SE	N
Grade 5	Used at least one minute (compared to did not use)	0.004	0.01	332,264
	300 or more minutes (compared to zero)	0.022	0.01	
	Met Istation dosage threshold (compared to did not meet threshold)	**0.019***	0.01	
Grade 8	Used at least one minute (compared to did not use)	**0.050*****	0.01	330,631
	300 or more minutes (compared to zero)	**0.105*****	0.02	
	Met Istation dosage threshold (compared to did not meet threshold)	**0.089*****	0.02	

Source: State of Texas Assessments of Academic Readiness and Public Education Information Management System data, 2011-12 – 2013-14, Texas Education Agency. Istation Session History table, 2013-14.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in the analyses. Minutes were prior to the administration of the first assessment, which was contingent on grade level. Estimates derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. Recommended Istation dosage threshold is 250 minutes for Grade 5 and 200 minutes for Grade 8. * p<0.05; ** p<0.01; *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Guiding Question 9a: Summary of Findings

Analyses in this section explored relationships between Istation usage (and usage intensity) and changes in student performance on STAAR-Reading between students at risk of being retained and those not at risk of being retained. A student is classified as being at risk of being retained in grade if he or she failed at least one STAAR-Reading assessment in 2011-12 or 2012-13. Analyses were confined only to students in Grades 5 and 8 in 2013-14 (i.e., grades in which promotion is conditioned on passing STAAR exams based on SSI requirements) and students who were enrolled in schools in which Istation was used.

Descriptive analyses demonstrated that, in Grades 5 and 8, just over 30% of students in 2013-14 were at risk of being retained in grade. Students in Grades 5 and 8 who were at risk of being retained spent more time on Istation curriculum sessions (an average of 239 minutes for Grade 5 students and 73 minutes for Grade 8 students) in 2013-14 compared to peers who were not at risk of being retained (an average of 152 minutes for Grade 5 students and 20 minutes for Grade 8 students).

In both Grades 5 and 8, students at risk of being retained in grade experienced smaller reading gains between 2012-13 and 2013-14 than students not at risk of being retained. The largest differences were observed among students in Grade 8. It is also important to note that Grade 8 students who were at risk of being retained and who used the Istation system for more minutes (200 or more minutes) over the course of the 2013-14 school year experienced larger gains relative to students who used the system for fewer minutes (less than 200 minutes). The differences were statistically significant, but substantively small. For instance, the average decile-standardized gain score for students in Grade 8 who used the system more than 300 minutes was .04 standard deviations below the average, compared to .15 standard deviations below the average for students who did not use the system. In Grade 5, students with greater Istation dosage experienced weaker gains than students who did not use the system at all.

Istation users in Grade 8 classified as at risk of being retained in grade demonstrated larger gains in reading than students who did not use the Istation system at all. Students who used the system for more minutes over the 2013-14 school year showed larger gains than students who did not use the system, or who used the system below the level recommended by the vendor. However, these differences were substantively small. Taken together, these findings suggest that usage of Istation among students at risk of being retained does not meaningfully improve, or harm, students STAAR-Reading score gains compared to their peers who did not use the system.

Guiding Question 10a:

What is the relationship between program usage after failing the first administration of the STAAR-Reading assessment and the probability of passing subsequent administrations of the STAAR exam?

Key Descriptive Findings:

- **Roughly half of Grade 5 students who failed the first administration of STAAR-Reading recorded at least one curriculum session between the first and second test administrations.** Across all Grade 8 students, only 22% used Istation during the interim period between STAAR test administrations, compared to 50% of just those Grade 5 students who retook the STAAR.
- **Grade 5 students who failed the first administration of STAAR-Reading used Istation more intensively between the first and second test administrations than Grade 8 students.** Nearly 20% of Grade 5 students who failed the first administration of STAAR-Reading recorded 100 or more minutes of Istation usage between the first and second test administrations, compared to 7% of Grade 8 students. Nearly three out of four Grade 8 students (77%) recorded no usage in the time between the first and second administrations. This pattern of lower Istation use among middle grade students was consistent with usage differences in the overall student population.

Key Multivariate Findings:

- **Grade 5 students who failed the first STAAR-Reading administration, and who used Istation between the first and second STAAR administrations, were more likely to pass the STAAR on the second administration than students who did not use the system.** This effect was, however, small. Students who used the system 100 minutes or more were approximately two percentage points more likely than students who did not use the system to meet the passing standard.
- **No statistically significant effects of Istation use between the first and second STAAR-Reading administrations were found for Grade 8 students.**

The research team received session-day-level usage data from Istation. With this level of granularity, usage, and usage intensity, could be calculated for different time spans throughout the school year, including the period between failure of the first administration and the second administration of STAAR-Reading in SSI grades (Grades 5 and 8), in which passage of the assessments are mandated by statute in order to be promoted to the next grade. Among students who failed the first administration of Grade 5 and 8 STAAR, the association between Istation usage between the first and second assessment during the 2013-14 school year (April 2nd and May 14th) and the likelihood of passing the second administration was investigated by using a logistic regression, with covariates including the amount of time a student used

Istation between the first and second STAAR-Reading assessments. With this framework, the differential probability of passing the second administration STAAR-Reading assessment between students who used Istation at different levels between assessments was estimated. It is important to note a limitation with this analysis. Students who fail the first administration of the STAAR-Reading assessment are provided with a wide array of intensive reading interventions which vary by school district. Therefore, in addition to Istation, a variety of factors and interventions may be contributing to student performance on the second administration of the STAAR-Reading assessment.

Descriptive Analysis of Usage between First and Second Administration

Table 2.25 provides the dates for the first and second administration of STAAR-Reading in 2013-14 for Grades 5 and 8. The last column presents the dates for which Istation system usage was calculated. Time in the system was calculated by simply summing the total number of *curriculum* session minutes a student recorded during this period.

Table 2.25. STAAR-Reading Test Administration Dates for Grades 5 and 8, 2013-14

	First Administration	Second Administration	Inter-Administration Usage Dates
Grade 5	April 2, 2014	May 14, 2014	April 3rd-May 13th
Grade 8	April 2, 2014	May 14, 2014	April 3rd-May 13th

Source: Texas Education Agency, 2014.

Approximately 84,000 students who failed the first Grade 5 STAAR-Reading test, and 59,000 who failed the first administration of the Grade 8 test had a valid test score on the second administration in May 2014. The administration rates among these students was between 99% (Grade 5) and 98% (Grade 8). (See Table 2.26.)

Table 2.26. Frequency Count of Students Who Failed the First Administration of STAAR-Reading and Received a Valid Test Score on the Second Administration, by Grade, 2013-14

Student Grade Level	Indicator of Whether Student Took 2nd Administration Test	Number of Students	Percentage of Students
Grade 5	No 2nd administration test	1,088	1.81
Grade 5	Took 2nd administration test	83,650	98.72
Grade 8	No 2nd administration test	1,282	2.13
Grade 8	Took 2nd administration test	58,944	97.87

Source: State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level included the second administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions). Frequency counts only included students who failed the first administration of STAAR-Reading in April 2014.

Next, the system usage patterns of students who failed the first administration are described. This analysis includes *only* students in Grades 5 and 8 in 2013-14 who failed the first administration of STAAR-Reading and received a validly scored test on the second administration in May 2014. Usage patterns in the period between the first and second administration *among students who failed the first administration* are markedly different between Grade 5 and 8 students, which is consistent with usage pattern differences throughout the school year. Table 2.27 depicts the count and percentage of Grade 5 and 8 students who recorded at least one curriculum session minute between the first and second administrations. In addition, two usage metrics are presented: the average number of curriculum minutes between August 2013 and March 2014, and the average number of curriculum minutes recorded in the period between the first and second administrations.

Approximately 22% of Grade 8 students who failed the first administration used Istation in the period between the first and second administrations, compared to 50% of Grade 5 students. Furthermore, the intensity of usage among Grade 5 students who failed the first administration of Grade 5 STAAR-Reading was nearly three times higher than Grade 8 students (65 minutes compared to 21 minutes).

Table 2.27. Description of Istation System Usage between the First and Second Administrations of Grades 5 and 8 STAAR-Reading, 2013-14

Student Grade Level	Count of Students Who Logged at Least One Curriculum Session between the First and Second Administration	Percentage of Students Who Logged at Least One Curriculum Session between the First and Second Administration	Mean Number of Minutes Prior to First Administration	Mean Number of Minutes between First and Second Administrations
Grade 5	41,544	49.66	240.52	65.19
Grade 8	13,129	22.27	75.42	20.66

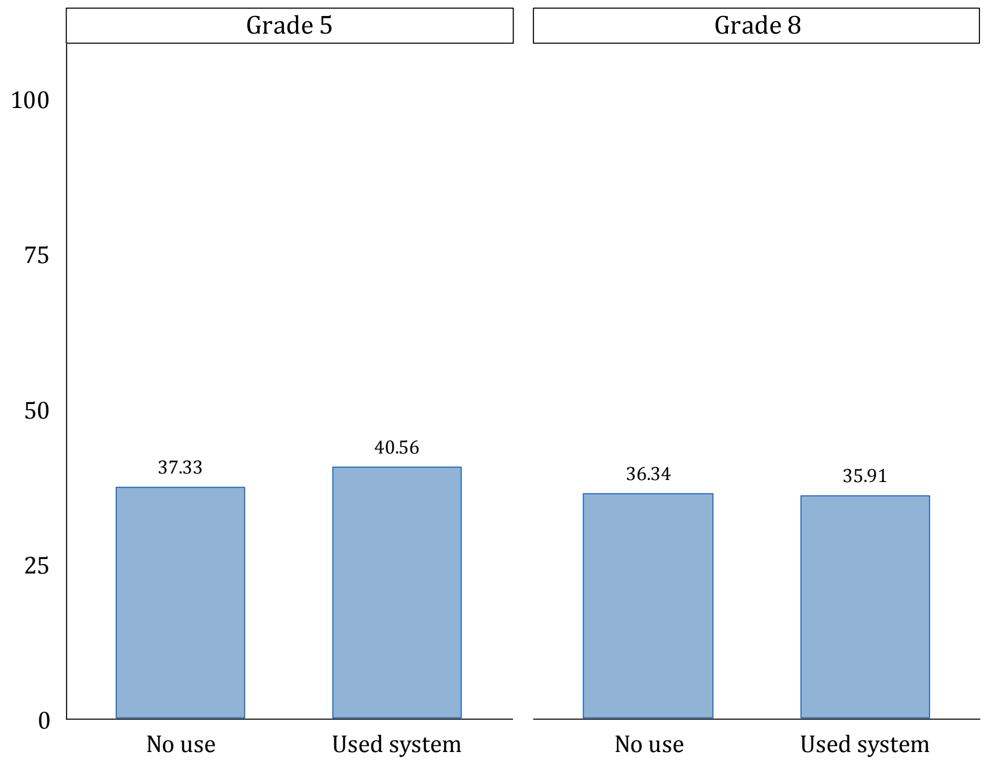
Source: Istation Session History table, 2013-14.

Note: Frequency counts and calculations only included students who failed the first administration of STAAR-Reading in April 2014 and had a validly scored test on the second administration. Only regular English and Spanish Versions (i.e., not modified or alternate versions) were included. Period between first and second administrations was April 3, 2014 to May 13, 2014. At least one curriculum session is defined as a curriculum session with greater than zero minutes of use.

Outcome Results

Figures 2.12 and 2.13 provide descriptive information about the association between system use in the interim period and the percentage of students who met the phase-in 1, Level II passing standard on the second administration of the Grade 5 or 8 STAAR-Reading test. Table 2.28 provides the distribution of students by usage intensity category. Figure 2.12 defines Istation usage narrowly, based only on whether the student recorded a curriculum session in the interim period between the first and second assessments, while Figure 2.13 classifies usage according to the number of minutes during this period. For Grade 5 students (Figure 2.12), the passing rate for students who used the system was four percentage points higher (41% compared to 37%) than students who did not use the system, and it was nearly identical among Grade 8 students.

Figure 2.12. Percentage of Students Who Met the Phase-In 1 Level II (Satisfactory) Standard on Second Administration of STAAR-Reading, by Interim Istation Usage, Grades 5 and 8, 2013-14 School Year

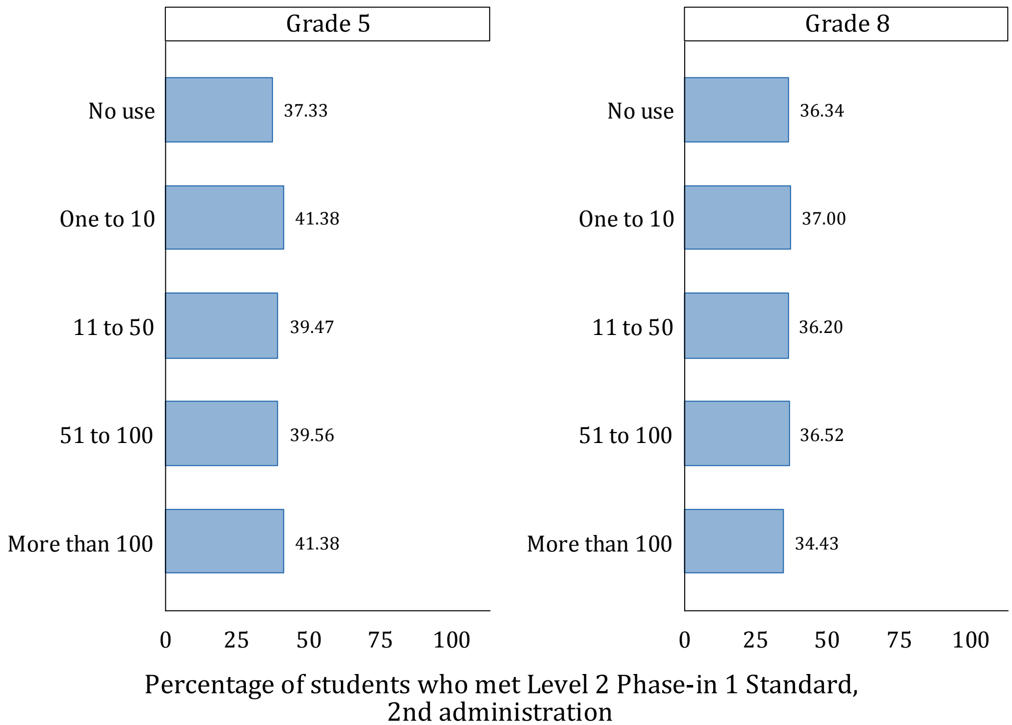


Source: Istation Session History table, 2013-14. State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: Calculations only included students who failed the first administration of STAAR-Reading (April 2014) and had a validly scored test on second administration. Only regular English and Spanish Versions (i.e., not modified or alternate versions) were included. The period between first and second administrations was April 3, 2014 to May 13, 2014. At least one curriculum session was defined as a curriculum session with greater than zero minutes of use.

Figure 2.13 indicates a positive relationship between usage intensity and the percentage of students who met the passing standard on the second administration among Grade 5 students. Students who recorded at least one curriculum session in the interim period between the first and second assessments met the passing standard on the second administration on the Grade 5 STAAR-Reading at a higher rate than students who did not use the system. Furthermore, the 16,937 students who used the system more than 100 minutes during this time had the highest passing rate (41.38%) on the second administration relative to any other usage intensity group. This pattern was not found among Grade 8 students, where students who used the system more than 100 minutes had the lowest passing rate (34.43%).

Figure 2.13. Percentage of Students Who Met the Phase-In 1 Level II (Satisfactory) Standard on Second Administration of STAAR-Reading, by Intensity of Istation Usage, Grades 5 and 8, 2013-14 School Year



Source: Istation Session History table, 2013-14. State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: Calculations only included students who failed the first administration of STAAR-Reading in April 2014 and had a validly scored test on the second administration. Only regular English and Spanish Versions (i.e., not modified or alternate versions) were included. Period between first and second administrations was April 3, 2014 to May 13, 2014. At least one curriculum session is defined as a curriculum session with greater than zero minutes of use.

These results, however, are purely descriptive, and do not account for differences in other characteristics between students who used the system and students who did not (and which may be correlated with changes in student performance), nor does it account for differences across campus types or environments. To account for these differences, the evaluation team fit a logistic regression with the response variable capturing whether the student passed the second administration STAAR-Reading test. In the model, the evaluation team adjusted for student attributes, including their performance on the *first* STAAR-Reading administration. Models were fit separately for Grades 5 and 8 to account for differences in the relationship between usage and usage intensity, and the likelihood of earning a satisfactory score, across grade levels. The results of this analysis are presented in Table 2.28.

Among Grade 5 students who failed the first STAAR-Reading assessment, students who used the system for 100 or more minutes in the period between the first and second administrations were more likely to pass the second administration (marginal effect of approximately two percentage points (36% compared to 38%) compared to students who did not use the system. In addition, the total number of minutes a student recorded during this period is also positive and statistically significant, suggesting Grade 5 students who used the system more intensively in the interim period were more likely to pass the second Grade 5 STAAR-Reading administration. Importantly, however, the differences in the estimated probabilities are small, ranging between 1.7 and 2 percentage points difference in the probability of passing the test on the second administration. The relationship between usage, and usage intensity, was negatively associated with the probability of passing the second administration among Grade 8 students, although the coefficients were not statistically significant, meaning the same effect was not found for Grade 8 students.

Table 2.28. Estimated Relationship between Istation Usage between the First and Second STAAR-Reading Administrations, Grades 5 and 8, 2013-14

Grade Level	Usage Measure	B	SE	Marginal Effect	N
Grade 5	Used at least once	0.035	0.020	NA	79,130
	100 or more minutes (compared to zero)	**0.072****	0.027	1.7 percentage points	
	Total minutes	**0.000*****	0.000	2 percentage points (zero minutes compared to 150 minutes)	
Grade 8	Used at least once	-0.021	0.028	NA	54,178
	100 or more minutes (compared to zero)	-0.084	0.047	NA	
	Total minutes	-0.000	0.000	NA	

Source: Istation Session History table, 2013-14. State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: Calculations only include students who failed the first administration of STAAR-Reading in April 2014 and had a validly scored test on the second administration. Only regular English and Spanish Versions (i.e., not modified or alternate versions) were included. Period between first and second administration was April 3, 2014 to May 13, 2014. Estimates were derived from a logistic regression with cluster-robust standard errors at the campus level. The covariates in the full functional form are provided in Appendix A. Marginal effects derived by holding all values at their means. Marginal effects for covariates that were not statistically significant were not calculated and are not presented. Statistically significant positive coefficients are denoted by bold font. * p<0.05, ** p<0.01, *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Guiding Research Question 10a: Summary of Findings

To assess the relationship between program usage after failing the first administration of the STAAR-Reading assessment and the probability of passing subsequent administrations of the STAAR exam, evaluators examined usage and usage intensity during the period between failure of the first administration and the second administration of STAAR-Reading in SSI grades (Grades 5 and 8).

Descriptive usage patterns varied by grade, with approximately 22% of Grade 8 students who failed the first administration using Istation in the period between the first and second administrations, compared to 50% of Grade 5 students. Intensity of usage among Grade 5 students who failed the first administration of Grade 5 STAAR-Reading was nearly three times higher than Grade 8 students (65 minutes compared to 21 minutes). Among students who used the system—relative to those who failed the first administration and did not use the system—the passing rate for Grade 5 Istation users was three percentage points higher than students who did not use the system (41% compared to 38%). Grade 8 Istation users had comparable retest passing rates than non-users (36%).

To control for other factors that may have impacted retest passing rates—outside of Istation use—the evaluation utilized a statistical model that explored whether system usage patterns in between STAAR-Reading assessment administrations was associated with a student passing the second administration of the STAAR-Reading test after controlling for differences in student attributes and prior reading results.

- Among Grade 5 students who failed the first STAAR-Reading assessment, students who used the Istation system for 100 or more minutes in the period between the first and second administrations were significantly more likely to pass the retest, compared to students who did not use the system. Intensity of use was also positive and statistically significant, suggesting that Grade 5 students who used the system more intensively in the interim period were more likely to pass the retest Grade 5 STAAR-Reading administration.
- Relationships between usage, and usage intensity, were negatively correlated with the probability of passing the second administration among Grade 8 students, although coefficients were not statistically significant.

These results imply a positive relationship between Istation usage and passing rates among students taking STAAR-Reading retests. Notably, while Grade 5 results were similar to positive relationships that have been demonstrated between Istation usage and reading gains in middle grades, the positive associations between Istation usage and passing retests did not extend to Grade 8 students.

Section 3 – Mathematics Outcomes

This section of the report is organized into two main parts: 1) implementation of the Think Through Math program (TTM) and how it is coordinated with other SSI-related mathematics interventions; and 2) relationship between TTM usage and student progress through TTM lessons and STAAR-Mathematics scores. A broader view of the SSI as it relates to mathematics is also examined in this section.

Implementation of TTM and other SSI-related Mathematics Interventions

The fidelity of TTM implementation can be assessed a variety ways. Similar to the Istation analysis presented in Section 2 of this report, the evaluation team chose to take a broad view of implementation and explore it in terms of students' usage patterns (and how usage levels corresponded with recommended minimum usage levels established by the vendor); district and campus participation in professional development and training required to use the TTM system more effectively; ways in which various districts and campuses used the TTM program and other SSI-related mathematics interventions to enable students to meet the grade promotion requirements for Grades 5 and 8; and perspectives of practitioners regarding fidelity of implementation and satisfaction with the TTM program.

The analysis of TTM program implementation serves to address the following four guiding research questions:

Guiding Question 1b: What are the program utilization rates across districts and campuses in Texas for TTM, and to what extent does TTM utilization vary by school characteristics and geographic regions (i.e., Title 1 school designation, ESC region, urbanicity, state accountability rating)?

Guiding Question 2b: To what extent is student usage within schools and districts reaching a level of fidelity with the intended implementation model, and how does it vary across various student groups (e.g., past academic performance, race/ethnicity, gender, English Language Learner status)?

Guiding Question 3b: To what extent are program supports for TTM, available through the online system vendors, and ESC 20 and other service centers, being utilized by participating districts and campuses?

Guiding Question 4b: How are school districts and campuses using TTM to address the SSI grade promotion requirements for mathematics, and what other mathematics resources are being used in combination with Texas SUCCESS programs to further support the learning needs of students to meet the SSI grade promotion requirements in mathematics?

First, the evaluation team examined TTM system usage patterns at the student level, and disaggregated usage results by grade level and the following school-level characteristics:

1. Title I campus
2. State accountability ratings
3. Status as *Priority* or *Focus school*[54]
4. Geographic region of the state (i.e., 20 ESC regions)
5. Urbanicity (e.g., urban, rural, suburban classifications)

The evaluation team also explored TTM system usage patterns by student groups, such as high/low performers on prior year STAAR-Mathematics assessment, race/ethnicity, gender, and ELL status. These analyses included assessments of the proportion of students using the systems at levels that approached a certain degree of fidelity (e.g., 5 or more attempted lessons, 10 or more attempted lessons, 5 or more passed lessons, 10 or more passed lessons). These recommended usage levels were determined by the TTM vendor.

Findings for each of the four research questions related to program implementation and system usage are presented below. Key findings are presented first for each research question. These analyses include assessments of the proportion of students using the systems at levels that approach a certain degree of fidelity. *These cut points are important because they reflect usage levels at which the TTM vendor feels that a student has reached a "gateway" point where he or she is more likely to continue using the system (5 or more lessons) and a "fidelity or impact" point where the student may start experiencing gains in mathematics scores (10 or more lessons).*[55]

Second, the evaluation team explored the extent to which staff at school districts and campuses across the state took advantage of training and technical assistance provided by the TTM vendor (either onsite or through regional sessions held at regional ESCs) or technical assistance provided through ESC 20. While data were limited, the results shed light on the extent to which TTM training was made available to district and campus staff across the state and the extent to which practitioners took advantage of this training.

Third, the evaluation team conducted interviews in spring 2014 with campus and district staff regarding the manner in which the TTM program was implemented at their schools, including their perspectives on the extent to which the program was implemented with fidelity. During the course of these interviews with district- and campus-level interventionists, the evaluation team captured additional information regarding how districts and campuses used the Texas SUCCESS programs in conjunction with other mathematics interventions to help students meet the SSI grade promotion requirements for Grades 5 and

[54] *Priority* schools represent Texas' lowest performing schools and account for at least 5% of the state's Title I schools. *Focus* schools account for at least 10% of the Title I schools in Texas and include those with the widest achievement gaps.

[55] Conference call with TTM and TEA staff (July 24, 2014).

8. Findings for each of the four research questions related to program implementation and system usage are presented below. Key findings are presented first for each research question.

Guiding Question 1b:

What are the program utilization rates across districts and campuses in Texas for the TTM program, and to what extent does TTM utilization vary by school characteristics and geographic regions (i.e., Title 1 school designation, ESC region, urbanicity, state accountability rating)?

Key Findings:

- **Just over a third of students in Grades 3-8 across the state completed at least one TTM lesson during the 2013-14 school year.** Two percent of students attempted, but did not complete, at least one TTM lesson in 2013-14, and the remaining 34% completed at least one lesson.[56]
- **Students in elementary grades (Grades 3-5) used TTM at higher rates than their middle school counterparts (Grades 6-8).** Between 43% and 44% of elementary school students completed at least one TTM lesson, while the same was true of 21% to 29% middle school students.
- **A slightly larger percentage of students in Title I campuses completed TTM lessons, while a slightly larger percentage of students in non-Title I campuses passed TTM lessons.** Roughly 78% of students in Title I campuses completed 5 or more lessons, versus 73% of non-Title I students; conversely, 47% of students in Title I campuses passed at least 5 lessons, while 52% of students in non-Title I campuses passed at least 5 lessons. Overall, while fewer students completed and passed 10 lessons, differences in completion and passing rates by Title I status were also observed at the 10 lesson usage threshold.
- **Substantive differences in TTM use were observed in different areas of the state.** Across TTM lessons attempted and passed, higher usage regions had roughly 20% to 30% more students engaged with the system than lower usage regions. Region 2 (Corpus Christi) had the highest proportion of students who attempted and passed TTM lessons while Region 8 (Mount Pleasant) had the lowest proportion of lessons attempted, and Regions 18 and 19 (Midland and El Paso) had the lowest passing percentage.
- **In most grades, students in schools rated as *Improvement Required* used the TTM system more often; while students in schools designated as *Met Standard* passed more TTM lessons.** Differences in TTM lessons attempted ranged from one to six percentage points and differences in lessons passed ranged from zero to eight percentage points between students at schools which *Met Standard* and those rated as *Improvement Required*.

While approximately 836,000 identifiable and unique students logged into the TTM system at least once during the 2013-14 school year, students varied in the intensity and consistency with which they used the system. It is important to note that a substantial proportion of students in Grades 3-8 contained in TEA's administrative database (PEIMS) could not be matched to the TTM user database. Of great interest, then, is the proportion of all students who logged into the system *and* the proportion of students who were meaningfully engaged in using it. This section addresses both the quantification of TTM usage and the extent to which TTM system usage rates varied across different types of schools (i.e., Title I schools, schools in different geographic regions of the state, rural/urban/suburban schools, and schools that did or did not meet state accountability standards in 2012-13).

TTM provided the evaluation team session-level (i.e., lesson-level), records for every student who used the system in 2013-14. Thus, the evaluation team was able to determine how many times a student used the system, and report basic descriptive information about the frequency with which the system was used during the 2013-14 school year.

Table 3.1 provides the number and percentage of Grade 3-8 students across the state who: 1) did not use the TTM system or could not be identified as having used the system; 2) logged into, or attempted, at least one TTM lesson but did not complete it; and 3) completed at least one TTM lesson during the 2013-14 school year. This table does not present the number or percent of students enrolled in TTM – a portion of those students are in each of the three columns in the table – instead it only shows the mutually exclusive category breakdown of students who never used the system, attempted but never completed a lesson, and attempted and completed at least one lesson. Data are further disaggregated by grade level. This table includes students who did not register or use the TTM system to reflect statewide Grade 3-8 estimates of usage.

Approximately 34% of Grade 3-8 students in the state completed at least one lesson during the 2013-14 school year.[57,58] Another 2% logged in but did not complete a lesson. Almost 64% of Grade 3-8 students in Texas could not be matched to TTM participation records. This may be a function of not using the system, as well as the inability to match students because of the enrollment platform employed and the lack of unique student identifiers reported.

The overall percentage of students in Grades 3-5 who completed one or more TTM sessions was substantially higher (43% to 44%) than among middle school students in Grade 6 (29%), Grade 7 (23%), and Grade 8 (21%). Similarly, higher percentages of middle school students were non-users of the TTM system when compared to students in Grades 3, 4, and 5. However, it is also important to note that the vast majority of students who logged onto the TTM system (94%) completed at least one mathematics lesson.

---

[56] The relatively low documented usage rate is a function of errors in the student IDs reported through the TTM enrollment platform, causing matching issues with TEA administrative data.

[57] Student records with 500 or more sessions were excluded from these tables as outliers.

[58] Only session records that occurred between August 2013 and June 2014 were retained.

Table 3.1. Frequency and Percentage of Students by TTM Usage, All Grade 3-8 Students, 2013-14

Student Grade Level	No Identifiable Record of TTM Usage		Logged In But Did Not Complete a TTM Lesson		Completed One or More TTM Lessons		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Grade 3	221,895	56.90%	661	0.17%	167,421	42.93%	389,977	100.00%
Grade 4	208,491	54.36%	5,519	1.44%	169,529	44.20%	383,539	100.00%
Grade 5	208,016	54.33%	7,078	1.85%	167,798	43.82%	382,892	100.00%
Grade 6	252,428	67.04%	13,234	3.51%	110,896	29.45%	376,558	100.00%
Grade 7	284,442	73.79%	10,605	2.75%	90,418	23.46%	385,465	100.00%
Grade 8	286,611	75.49%	13,691	3.61%	79,378	20.91%	379,680	100.00%
Total	**1,461,883**	**63.61%**	**50,788**	**2.21%**	**785,440**	**34.18%**	**2,298,111**	**100.00%**

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes all students in Grades 3-8 regardless of whether they logged onto the TTM system and attempted a lesson. In addition, the "No Identifiable Record of TTM Usage" column includes students who could not be matched between the PEIMS and TTM systems.

In all analyses that follow, the evaluation team confined usage analysis to students who logged in to the TTM program and attempted a TTM lesson (but did not necessarily complete the lesson) at least one time during the 2013-14 school year. Further, grade level information about TTM student users were not available through the system; therefore, student grade level was obtained by linking TTM files to PEIMS administrative files. The TTM-PEIMS link performed by TEA matched approximately 72% of students and excluded matching students who registered for TTM but never completed lessons.[59] This is important, because approximately 28% of students from the TTM registration roster could not be linked back to TEA administrative records. This will, most likely, lead to an underreporting of the number of students who used the system.[60]

Table 3.2 Illustrates that the majority of students (94%) who logged into a TTM lesson during the 2013-14 school year went on to complete one or more lessons in the online mathematics system. While this completion percentage was high for both elementary and middle school students, it was somewhat higher among students in Grades 3 to 5 (96% to over 99%) compared to students in Grades 6 to 8 (85% to 90%).

[59] Correspondence with TEA on 6/20/2014.

[60] While there is no particular reason to believe that the unmatched students are non-randomly distributed across the state, only students who logged into the TTM system were included in the outcomes analyses presented in Guiding Questions 7b through 10b.

Table 3.2. Frequency and Percentage of Students by TTM Usage, TTM Users Grades 3-8, 2013-14

Student Grade Level	Logged In But Did Not Complete a TTM Lesson		Completed One or More TTM Lessons		Total	
	Number	Percent	Number	Percent	Number	Percent
Grade 3	661	0.39%	167,429	99.61%	168,082	100.00%
Grade 4	5,519	3.15%	169,530	96.85%	175,048	100.00%
Grade 5	7,078	4.05%	167,798	95.95%	174,876	100.00%
Grade 6	13,234	10.66%	110,896	89.34%	124,130	100.00%
Grade 7	10,605	10.50%	90,422	89.50%	101,023	100.00%
Grade 8	13,691	14.71%	79,381	85.29%	93,069	100.00%
Total	**50,788**	**6.07%**	**785,456**	**93.93%**	**836,228**	**100.00%**

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

As noted previously, TTM system usage results were further disaggregated by the school's Title I status, geographic region of the state, urbanicity, state accountability rating, and *Priority/Focus* school status. The following TTM usage thresholds are presented for each disaggregation: 1) percent of students who attempted 5 or more TTM lessons; 2) percent of students who attempted 10 or more TTM lessons; 3) percent of students who passed 5 or more TTM lessons; and 4) percent of students who passed 10 or more TTM lessons. These cut points reflect usage levels at which the TTM vendor feels that a student has reached a "gateway" point where they are more likely to continue using the system (5 or more lessons) and a "fidelity or impact" point where students may start experiencing gains in mathematics scores (10 or more lessons). The results of these descriptive analyses are presented below.

Title I Status

As Table 3.3 shows, a slightly higher percentage of students enrolled at Title I campuses attempted 5 or more TTM lessons (78% versus 73%) or 10 or more TTM lessons (57% versus 51%) than their non-Title I counterparts. However, a smaller proportion of students at Title I schools *passed* 5 or more TTM lessons (47% versus 52%) or 10 or more TTM lessons (24% versus 31%).

Table 3.3. TTM System Usage Disaggregated by Campus Title I Status, 2013-14

Campus Title I Status	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
Not Title 1	231,840	72.91%	51.28%	52.14%	31.31%
Title 1	602,707	78.46%	56.69%	46.77%	24.28%
Total	**834,547**	**76.92%**	**55.19%**	**48.26%**	**26.23%**

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The "5 or more" and "10 or more" categories were not mutually exclusive (i.e., students included in "10 or more" percentages were also included in the "5 or more" percentages).

To gain a better understanding of how usage levels varied by grade levels for students at Title I and non-Title I schools, the evaluation team calculated usage by the four primary usage metrics (i.e., related to lesson attempts and lessons passed). There were minimal differences (less than six percentage points) in student usage by Title I status for each grade level. A slightly larger percentage of students in Title I schools attempted 5 or 10 TTM sessions, with the exception of Grade 8, where more non-Title I students attempted 5 or 10 lessons. (See Table D9 in Appendix D for further detail.)

Geographic Region

As Table 3.4 shows, there was substantial geographic variation in TTM system usage. For TTM usage attempts, the percent of students who attempted at least 5 lessons ranged from 68% to 89% across regions and the proportion of students attempting at least 10 ranged from 44% to 75%. For TTM lessons passed, the percent of students who passed at least 5 lessons ranged from 37% to 62% and the proportion of students passing at least 10 lessons ranged from 17% to 37%. Across all usage metrics—that is, TTM lessons attempted and passed—Region 2 (Corpus Christi) had the highest proportion of TTM use. In terms of lessons attempted, Region 8 (Mount Pleasant) had the smallest proportion, while Regions 18 (Midland) and 19 (El Paso) had the smallest percentages of students passing TTM lessons.

Table 3.4. TTM System Usage Disaggregated by Geographic Region, 2013-14

ESC Region	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
1 (Edinburg)	12,361	80.95%	61.14%	47.46%	24.19%
2 (Corpus Christi)	12,361	88.53%	74.52%	62.09%	36.64%
3 (Victoria)	8,113	82.47%	61.75%	51.42%	27.65%
4 (Houston)	174,759	77.23%	54.77%	49.17%	27.39%
5 (Beaumont)	16,628	78.70%	58.31%	48.80%	24.50%
6 (Huntsville)	17,813	76.57%	54.04%	45.51%	22.24%
7 (Kilgore)	28,195	73.43%	50.72%	42.47%	21.33%
8 (Mount Pleasant)	6,160	67.60%	43.73%	38.93%	18.99%
9 (Wichita Falls)	3,999	82.05%	62.29%	52.69%	25.66%
10 (Richardson)	137,540	75.38%	53.85%	48.74%	28.00%
11 (Fort Worth)	116,008	80.58%	60.21%	54.28%	31.37%
12 (Waco)	16,688	81.56%	59.87%	51.48%	26.48%
13 (Austin)	56,527	74.62%	50.78%	47.14%	25.15%
14 (Abilene)	8,563	77.04%	53.88%	48.20%	25.93%
15 (San Angelo)	5,210	76.18%	50.86%	42.57%	19.06%
16 (Amarillo)	10,599	77.20%	55.13%	49.40%	25.72%
17 (Lubbock)	10,590	79.23%	58.37%	49.45%	25.88%
18 (Midland)	13,762	75.92%	51.90%	39.60%	17.26%
19 (El Paso)	29,247	71.10%	45.78%	37.45%	16.62%
20 (San Antonio)	89,470	72.26%	50.57%	44.49%	23.86%
Total	**836,228**	**76.93%**	**55.21%**	**48.30%**	**26.26%**

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The "5 or more" and "10 or more" categories were not mutually exclusive (i.e., students included in "10 or more" percentages were also included in the "5 or more" percentages).

Urbanicity

Modest variation in TTM usage rates during in the 2013-14 school year was observed when urbanicity was taken into consideration. Data were disaggregated by the following eight categories captured by TEA: 1) Charter School; 2) Independent Town; 3) Major Suburban; 4) Major Urban; 5) Non-metropolitan Fast Growing; 6) Non-metropolitan Stable Growth; 7) Other Central City; 8) Other Central City Suburban; and 9) Rural.

In terms of lessons attempted, a smaller percentage of students enrolled in schools located in Major Urban areas attempted 5 or more and 10 or more TTM lessons (72% and 49%, respectively) compared to students in any other jurisdiction, such as students enrolled in Charter Schools (82% and 60%, respectively), Rural (82% and 61%, respectively) or Non-Metropolitan Stable Growth (81% and 60%, respectively) and Fast Growth (82% and 56%, respectively) areas. Similar trends emerged when looking at the percentage of students who passed TTM lessons. A smaller percentage of students enrolled in schools in Major Urban areas passed 5 or more and 10 or more TTM lessons (41% and 21%, respectively) relative to students in other areas, while students enrolled in Rural area schools passed TTM lessons at higher rates (55% and 31% for 5 or more and 10 or more lessons, respectively). (See Table D10 in Appendix D.)

State Accountability Rating

To assess whether TTM system usage was related to prior year (2012-13) campus academic performance (as measured by the state accountability rating), the number of lessons attempted and passed were analyzed for schools designated as *Improvement Req*uired or *Met Standard*. As Table 3.5 shows, minor differences in usage levels were observed between students at campuses designated as *Improvement Required* and campuses designated as *Met Standard*. A slightly larger percentage of students at *Improvement Required* campuses attempted 5 or more (79% versus 77%) or 10 or more (59% versus 55%) TTM lessons in 2013-14 than students at campuses which *Met Standard* in 2012-13. Conversely, more students at campuses which *Met Standard* in 2012-13 passed 5 or more (49% versus 44%) or 10 or more (27% versus 22%) TTM lessons in 2013-14.

Table 3.5. TTM System Usage Disaggregated by Campus Accountability Rating, 2013-14

2012-13 Accountability Rating	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
Improvement Required	61,917	79.34%	58.78%	43.92%	22.05%
Met Standard	764,273	76.70%	54.89%	48.69%	26.64%
Total	**826,190**	**76.90%**	**55.18%**	**48.33%**	**26.30%**

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The "5 or more" and "10 or more" categories were not mutually exclusive (i.e., students included in "10 or more" percentages were also included in the "5 or more" percentages).

Students in Grades 3 to 5 at campuses with 2012-13 *Improvement Required* state accountability ratings attempted 5 or more and 10 or more TTM lessons more frequently than students at schools with *Met Standard* ratings, with a 10 percentage point differential on the 10 or more attempted metric (68% versus 58%) for Grade 5 students, who are subject to SSI grade promotion requirements (see Table 3.6). In two of the middle school grades (Grades 6 and 8), students at schools that met state accountability standards in 2012-13 were slightly more likely to attempt 5 or more and 10 or more TTM lessons in 2013-14. With the exception of Grade 5 (where grade promotion requirements are in place), more students enrolled in campuses that met state standards in 2012-13 passed 5 or more and 10 or more TTM lessons during the 2013-14 school years than their peers in schools rated as *Improvement Required*.

Table 3.6 TTM Usage Disaggregated by State Accountability Rating and Grade Level, 2013-14

Grade Level	2012-13 Accountability Rating	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
Grade 3	*Improvement Required*	14,178	87.49%	65.19%	43.50%	19.54%
Grade 3	*Met Standard*	151,938	86.44%	62.92%	50.30%	24.35%
Grade 4	*Improvement Required*	15,094	83.54%	62.18%	45.89%	24.36%
Grade 4	*Met Standard*	158,186	82.54%	60.39%	52.23%	29.81%
Grade 5	*Improvement Required*	13,023	86.68%	67.58%	54.11%	27.90%
Grade 5	*Met Standard*	160,426	81.11%	58.42%	54.67%	30.05%
Grade 6	*Improvement Required*	8,875	66.38%	48.52%	38.67%	20.35%
Grade 6	*Met Standard*	113,304	67.94%	48.52%	44.62%	25.99%
Grade 7	*Improvement Required*	5,442	70.38%	47.37%	36.35%	17.88%
Grade 7	*Met Standard*	93,806	69.16%	47.45%	44.23%	24.82%
Grade 8	*Improvement Required*	5,305	58.45%	39.23%	30.97%	15.00%
Grade 8	*Met Standard*	86,613	60.38%	40.59%	38.46%	21.39%

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The "5 or more" and "10 or more" categories were not mutually exclusive (i.e., students included in "10 or more" percentages were also included in the "5 or more" percentages).

Overall, little difference in system usage was observed among students at *Focus* schools (77% attempted lessons, 41% passed lessons) and *Priority* schools (78% attempted lessons, 40% passed lessons) in terms of either attempting or passing 5 or more TTM mathematics lessons.[61] In Grade 8, larger proportions of students in *Focus* schools attempted and passed TTM lessons, while the opposite was true among Grade 7 students, where larger proportions of *Priority* school students attempted and passed TTM lessons. (See Table D11 in Appendix D.)

[61] *Priority* schools represent Texas' lowest performing schools and account for at least 5% of the state's Title I schools. *Focus* schools account for at least 10%of the Title I schools in Texas and include those with the widest achievement gaps.

Guiding Research Question 1b: Summary of Findings

Approximately 36% of the Grade 3-8 students in Texas could be identified as TTM system users, though it is important to note that many students' records could not be linked to TTM even though some may have used the system. TTM usage decreased by grade level, students in Grades 3-5 having the highest proportion of system users (43%-46%) and Grade 8 having the lowest (25%). Of those students who logged into TTM, the vast majority (94%) completed at least one lesson; again, with use of the system decreasing by grade level.

School-level characteristics related to TTM system usage include:

- A higher percentage of students in Title I schools attempted TTM lessons while a higher percentage of non-Title I students *passed* TTM lessons.
- By educational region, in terms of TTM lessons attempted, the following usage patterns were observed:
  - The percent of students attempting 5 or more lessons ranged from 68% to 89%.
  - The percent of students attempting 10 or more lessons ranged from 44% to 75%.
  - Region 2 (Corpus Christi) had the largest proportion of students who attempted TTM lessons and Region 8 (Mount Pleasant) had the smallest proportion of TTM lessons attempted (across both the 5 or more and 10 or more usage cut-points).
- By educational region, for TTM lessons passed, the following passing rates were observed:
  - The percent of students who passed 5 or more lessons ranged from 37% to 62%.
  - The percent of students who passed 10 or more lessons ranged from 17% to 37%.
  - Region 2 (Corpus Christi) had the highest percentage of students passing TTM lessons and Regions 18 (Midland) and 19 (El Paso) had the lowest percentage of students who passed TTM lessons.
- Among all geographic regions, a lower percentage of students attending schools in Major Urban areas used the system at recommended levels compared to students in Rural or Non-Metropolitan Stable Growth areas.
- Students in Grades 3-5 at campuses with 2012-13 *Improvement Required* state accountability ratings attempted 5 or more and 10 or more TTM lessons more frequently than students at schools with *Met Standard* ratings, with a 10 percentage point differential on the 10 or more attempted metric (68% versus 58%) for Grade 5 students, who are subject to SSI grade promotion requirements. In two of the middle school grades (Grades 6 and 8), students at schools that met state accountability standards in 2012-13 were slightly more likely to attempt 5 or more and 10 or more TTM lessons in 2013-14.

Guiding Question 2b:

To what extent is student usage within schools and districts reaching a level of fidelity with the intended implementation model, and how does it vary across various student groups (e.g., past academic performance, race/ethnicity, gender, English Language Learner status)?

Key Findings:

- **When considering Grade 3-8 students who attempted at least one TTM lesson, roughly one quarter of students used it minimally in 2013-14, while three quarters of students used it at the recommended threshold levels of 5 or more attempted lessons.** Of the students who used the TTM system in 2013-14, approximately 23% attempted 1 to 4 lessons, 22% between 5 and 9 lessons, and 55% attempted 10 or more lessons.
- **There were usage differences between elementary and middle grade students, particularly among students who used TTM minimally.** Of students using TTM in 2013-14, 13% of Grade 3 students attempted fewer than 5 lessons, compared to 40% of Grade 8 students. Similarly, 63% of Grade 3 TTM users attempted 10 or more lessons during the 2013-14 school year compared to 41% of Grade 8 students.
- **It does not seem that students were identified for TTM based on prior academic achievement in mathematics.** There was little or no descriptive relationship between 2012-13 STAAR-Mathematics test scores and the number of TTM lessons attempted during the 2013-14 school year. The same was true for other student populations, such as 2012-13 grade retention status, ELL status, and other demographic student characteristics. These trends suggest that schools and teachers did not assign students to TTM based on their prior achievement or other observable characteristics.
- **Over three quarters of campus staff interviewed indicated that they experienced some barriers in implementing TTM at their schools.** The most commonly cited barriers were not having enough computers (35%), not having enough time in the daily schedule (26%), and technology issues (23%).

Table 3.7 presents the number and proportion of students using the TTM system at various levels (as measured by categories of lessons attempted and lessons passed). To illustrate important differences in usage levels at the elementary and middle school levels, the data were further disaggregated by grade level in this table. The calculations included only students who logged into a TTM program lesson during the 2013-14 school year and attempted a TTM lesson. As Table 3.7 illustrates, elementary students were less likely than their middle school counterparts to attempt 1 to 4 TTM lessons, but comparable percentages of elementary and middle school students completed between 5 and 9 TTM lessons. It is important to note that a substantially larger percentages of Grade 3-5 elementary school students attempted 10 or more lessons than their middle school counterparts in Grades 6 to 8.

Table 3.7. Percent of Students Attempting Various Numbers of TTM Lessons by Grade, 2013-14

Grade Level	Percent of Students with 1 to 4 Lessons Attempted	Percent of Students with 5 to 9 Attempted Lessons	Percent of Students with 10 or More Attempted Lessons
Grade 3 (n=168,082)	13.43%	23.41%	63.17%
Grade 4 (n=175,048)	17.32%	22.10%	60.58%
Grade 5 (n=174,876)	18.44%	22.45%	59.11%
Grade 6 (n=124,130)	32.11%	19.36%	48.53%
Grade 7 (n=101,023)	30.74%	21.80%	47.46%
Grade 8 (n=93,069)	39.58%	19.72%	40.70%
Total (n=836,228)	**23.07%**	**21.73%**	**55.21%**

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students.

TTM Usage Patterns Disaggregated by Student Groups

To better understand the characteristics of students using the TTM system across the state of Texas, TTM usage patterns were further disaggregated by a number of characteristics, including:

- Students' prior year's performance on the STAAR-Mathematics exam (i.e., organized by the quartile of their STAAR scale score)
- Students' prior year grade retention status (i.e., retained in grade in 2012-13)
- Students' race/ethnicity
- Student's gender
- Students' English language learner (ELL) status

To assess the extent to which students were using the system with fidelity (based on vendor recommendations of lesson attempt levels), the following metrics were calculated for each student group: 1) percent of students attempting 5 or more TTM lessons; 2) percent of students attempting 10 or more TTM lessons; 3) percent of students passing 5 or more TTM lessons; and 4) percent of students passing 10 or more TTM lessons. As noted in Guiding Question 1b, TTM suggests that students who use 5 or more sessions in a given school year are at the "gateway level" and they are more likely to continue use of the system; and that students who complete 10 or more lessons in a school year are more likely to experience gains in mathematics scores than other students.

Prior Performance of STAAR-Mathematics

While data are commonly reported for Grades 3-8 in this report, the disaggregation of TTM usage by prior STAAR performance was reported for Grades 4-8 because prior data were not available for students in Grade 3 in 2013-14. For this analysis, each student was categorized into quartiles of performance on first administration of the 2012-13 STAAR-Mathematics exam.

As Table 3.8 shows, 2013-14 TTM system usage levels for students in the bottom quartile and the top quartile of 2012-13 STAAR-Mathematics scores were comparable for both of the lessons attempted thresholds:

- 76% of students in the bottom quartile attempted 5 or more TTM lessons compared to 75% of students in the top mathematics performance quartile.
- The same percentage of students (56%) in the bottom and top mathematics performance quartiles attempted 10 or more TTM lessons.

These findings strongly suggest that students were not identified for the TTM intervention based on past performance on the STAAR-Mathematics assessment.

However, prior performance on state assessment for mathematics was related to students actually passing more TTM lessons. Students in the top quartile on the 2012-13 STAAR-Mathematics assessment were more likely to pass 5 or more (63% versus 37%) and 10 or more (43% versus 16%) TTM lessons compared to students in the bottom quartile of 2012-13 STAAR-Mathematics scores.

Table 3.8. 2013-14 TTM Usage Disaggregated by Prior Year (2012-13) STAAR-Mathematics Performance

Prior Year STAAR-Mathematics Performance	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
Bottom quartile	169,903	76.41%	55.73%	36.89%	16.19%
25th-50th	153,063	73.50%	50.81%	45.44%	22.54%
50th-75th	146,943	73.28%	50.84%	52.93%	30.24%
Top quartile	139,129	75.30%	55.56%	63.41%	43.42%
Total	**609,038**	**74.67%**	**53.27%**	**48.97%**	**27.40%**

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The "5 or more" and "10 or more" categories were not mutually exclusive (i.e., students included in "10 or more" percentages were also included in the "5 or more" percentages).

Overall, there appears to be little relationship between performance on the 2012-13 STAAR-Mathematics assessment and the propensity to use the TTM system in 2013-14. However, some grade level differences were observed. For example, 75% of bottom quartile students in Grade 7 attempted 5 or more TTM lessons (compared to 67% of top quartile students), and 55% of bottom quartile students completed 10 or more TTM lessons (compared to 47% of top quartile students). These differences tended to disappear in Grade 8, which is one of the grades subject to SSI grade promotion requirements (see Table D12 in Appendix D).

Differences in the percent of top and bottom quartile students passing 5 or more and 10 or more TTM lessons in 2013-14 were consistent across grades with higher percentages of top quartile students passing than bottom quartiles, but more profound in the elementary grades. For example, in Grade 4, 72% of top quartile students passed 5 or more TTM lessons compared to just 31% of bottom quartile students. (See Table D12 in Appendix D.)

Prior Year Grade Retention Status

Whether a student had been retained in grade in a prior school year does not appear to have much impact on TTM usage. This suggests that students were not identified for TTM usage based on prior grade retention status. As Table 3.9 shows, a higher proportion of elementary school students were attempting 5 or more and 10 or more TTM lessons compared to students in the middle school grades; however, differences in usage rates between students who were retained in grade in 2012-13 and those who were promoted to the next grade were small and mixed in terms of which group was completing more TTM lessons in 2013-14. For example, in Grades 3, 6, 7 and 8, a higher percent of students who were promoted to the next grade attempted 5 or more TTM lessons, and the reverse was true for students in Grade 5.[62] Comparable TTM usage rates were observed in both retained and promoted students in Grades 4 and 7.

Table 3.9. TTM Usage Disaggregated by Prior-Year Grade Retention Status and Grade Level, 2013-14

Grade Level	Prior Year Grade Retention Status	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
Grade 3	Promoted	158,422	86.73%	63.32%	50.00%	24.14%
Grade 3	Retained	3,300	81.82%	58.52%	35.03%	12.33%
Grade 4	Promoted	166,829	82.66%	60.62%	51.84%	29.42%
Grade 4	Retained	2,000	82.45%	58.85%	36.10%	17.15%
Grade 5	Promoted	166,804	81.46%	58.99%	54.71%	29.97%
Grade 5	Retained	2,314	84.66%	63.40%	45.72%	20.18%
Grade 6	Promoted	119,254	67.71%	48.36%	44.10%	25.42%
Grade 6	Retained	767	65.32%	47.20%	32.99%	18.90%
Grade 7	Promoted	96,872	69.04%	47.26%	43.72%	24.33%
Grade 7	Retained	729	68.04%	41.98%	31.28%	15.36%
Grade 8	Promoted	89,567	60.23%	40.53%	38.10%	21.06%
Grade 8	Retained	677	55.54%	34.56%	26.59%	11.52%

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The "5 or more" and "10 or more" categories were not mutually exclusive (i.e., students included in "10 or more" percentages were also included in the "5 or more" percentages).

Race/Ethnicity

With the exception of Asian/Pacific Islander students, for which a larger percentage of students attempted 5 or more (83%) and 10 or more (66%) TTM lessons, comparable proportions of all other student groups attempted 5 or more (76% to 78%) and 10 or more (54% to 56%) lessons through the TTM system. A substantially higher percentage of Asian/Pacific Islander students passed 5 or more (69%) and 10 or more (50%) TTM lessons than any of the other student groups, which ranged from 44% to 54% passing 5 or more TTM lessons and 23% to 31% passing 10 or more TTM lessons. (See Table D13 in Appendix D.)

[62] Comparable percentages of students, retained in grade and promoted to the next grade, attempted 5 or more TTM lessons in Grade 4.

Gender

Overall, small differences in TTM system usage were found between male and female students. A total of 78% of male students attempted 5 or more TTM sessions in 2013-14 compared to 76% of female students. Similarly, 51% of male students passed 5 or more TTM lessons in 2013-14 compared to 48% of female students. Only minor differences in system usage for male and female students were observed across grade levels, with larger variances observed for the 10 or more lessons attempted in elementary grades (see Table D14 in Appendix D).

English Language Learner Status

At Table 3.10 shows, for students in Grades 3 to 5, subtle differences in the percentage of students classified as ELL and non-ELL students attempted 5 or more and 10 or more TTM lessons were observed – with ELL students somewhat more inclined to use the system at these levels. The differences in TTM system usage between ELL and non-ELL status disappear for middle school grades (and even reverse in Grade 8). A higher proportion of non-ELL students across all grades passed more TTM lessons than students identified as ELL, with the largest gaps observed in Grade 6 (10 to 11 percentage points depending upon the metric), Grade 7 (10 to 11 percentage points), and Grade 8 (10 to 12 percentage points depending upon the metric).

Table 3.10. TTM System Usage Disaggregated by ELL Status and Grade Level, 2013-14

Grade Level	ELL Status	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
Grade 3	Non ELL	124,114	86.34%	62.54%	51.53%	25.05%
Grade 3	ELL	43,968	87.22%	64.93%	44.53%	20.72%
Grade 4	Non ELL	135,895	82.51%	60.11%	53.60%	30.75%
Grade 4	ELL	39,153	83.24%	62.21%	44.83%	24.21%
Grade 5	Non ELL	140,407	80.98%	57.97%	55.92%	31.04%
Grade 5	ELL	34,469	83.92%	63.76%	49.10%	25.05%
Grade 6	Non ELL	105,269	67.82%	48.54%	45.80%	27.03%
Grade 6	ELL	18,861	68.30%	48.50%	34.82%	16.97%
Grade 7	Non ELL	88,125	69.01%	47.35%	45.20%	25.65%
Grade 7	ELL	12,898	70.96%	48.18%	33.94%	15.63%
Grade 8	Non ELL	83,519	60.66%	40.89%	39.40%	22.17%
Grade 8	ELL	9,550	58.35%	39.03%	27.20%	11.77%

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The "5 or more" and "10 or more" categories were not mutually exclusive (i.e., students included in "10 or more" percentages were also included in the "5 or more" percentages).

Facilitators and Barriers to TTM Implementation

To gain a better understanding of the factors that may be impacting TTM usage at campuses across the state, staff responsible for implementing TTM and other mathematics interventions at their schools were interviewed in spring 2014. While a number of different facilitators for TTM usage were shared by interviewees, the most commonly noted items included differentiated and supportive content (18%), monitoring student use (18%), student enjoyment in using the TTM system (13%), having a clear plan and goals for implementation and system usage (10%), and training on how to use the TTM system (10%).

The campus staff interview sample was stratified by the intensity and breadth (across grade levels) of TTM usage. High, moderate, and low usage campuses were identified for staff interviews. Staff at campuses with high system usage rates were much more inclined than staff at moderate or low usage TTM campuses to indicate that having a clear plan and goals for system implementation/usage and monitoring student use were key facilitators for effective implementation. (See Table 3.11.)

Table 3.11. Primary TTM Implementation Facilitators by Usage and School Level

Facilitators	Overall (n=77)	High Usage (n=28)	Moderate Usage (n=27)	Low Usage (n=22)	Elementary School (n=38)	Middle School (n=39)
Differentiated/supportive content	18%	14%	19%	23%	18%	18%
Monitoring student use	18%	25%	15%	14%	18%	18%
Motivation/enjoyment	13%	18%	15%	5%	16%	10%
Clear plan/goals	10%	21%	7%	0%	8%	13%
Training	10%	18%	0%	14%	16%	5%
Consistent use	9%	4%	19%	5%	8%	10%
Support for implementation	9%	14%	11%	0%	5%	13%
Computers/ technology	8%	11%	4%	9%	5%	10%
Understanding TTM	6%	7%	7%	5%	8%	5%

Source: Spring 2014 Interviews with campus-level staff.

Over three quarters (78%) of campus staff interviewed reported that some barriers to implementation were encountered for TTM. This finding was consistent across high, moderate, and low usage campus, as well as elementary and middle schools. Across all campuses, not enough computers (35%), not having enough time in the daily schedule (26%), and technology issues (23%) were the most frequently mentioned barriers to implementing the online mathematics program. Training and knowledge of the program, and lack of staff buy-in to the TTM program were each noted as implementation barriers by 11% of interviewees.

Supports for Implementation

Staffing to Support TTM Implementation

About one third of interviewees (34%) indicated that no additional support staff other than the teacher (or TTM interventionist) were used to help implement the TTM program at their campuses. The lack of support staff was most evident at low TTM usage campuses, where 65% shared that no additional staff were used to implement TTM. Educational aides (used by 26% overall, but 43% of high TTM usage campuses), and interventionists/coaches (used by 26% overall, and 44% of moderate TTM use campuses) were most commonly used to support the TTM program. Educational aides were also somewhat more prominent at elementary school (30%) campuses than middle schools (21%). (See Table 3.12.)

Table 3.12. Percent of Campuses Indicating TTM Support Personnel by Usage and School Type

	Overall (n=70)	High Usage (n=23)	Moderate Usage (n=27)	Low Usage (n=20)	Elementary School (n=37)	Middle School (n=33)
No other staff assists	34%	17%	26%	65%	38%	30%
Educational aides	26%	43%	19%	15%	30%	21%
Interventionists/Coaches	26%	17%	44%	10%	27%	24%
Computer lab staff	13%	17%	15%	5%	19%	6%

Source: Spring 2014 Interviews with campus-level staff.

Adequate Resources to Support TTM Implementation

Interviewees were asked about whether their campuses had adequate resources to effectively implement the TTM program. They were queried about instructional staff, educational aides, computers/laptops, and internet connectivity. High percentages of interviewees across all groups indicated that they had enough instructional staff (82% to 90%) and internet connectivity (80% to 96%) to implement TTM well. Even within these high ratings, differences were observed. A lack of computers was a critical issue for campuses that used the TTM program at moderate or minimal levels. High usage campuses (73%) were much less likely than moderate (44%) or low (48%) TTM usage campuses to mention that they did not have enough computers to effectively implement the online mathematics program. (See Table 3.13.)

Table 3.13. Enough Resources Available to Effectively Implement TTM, 2013-14

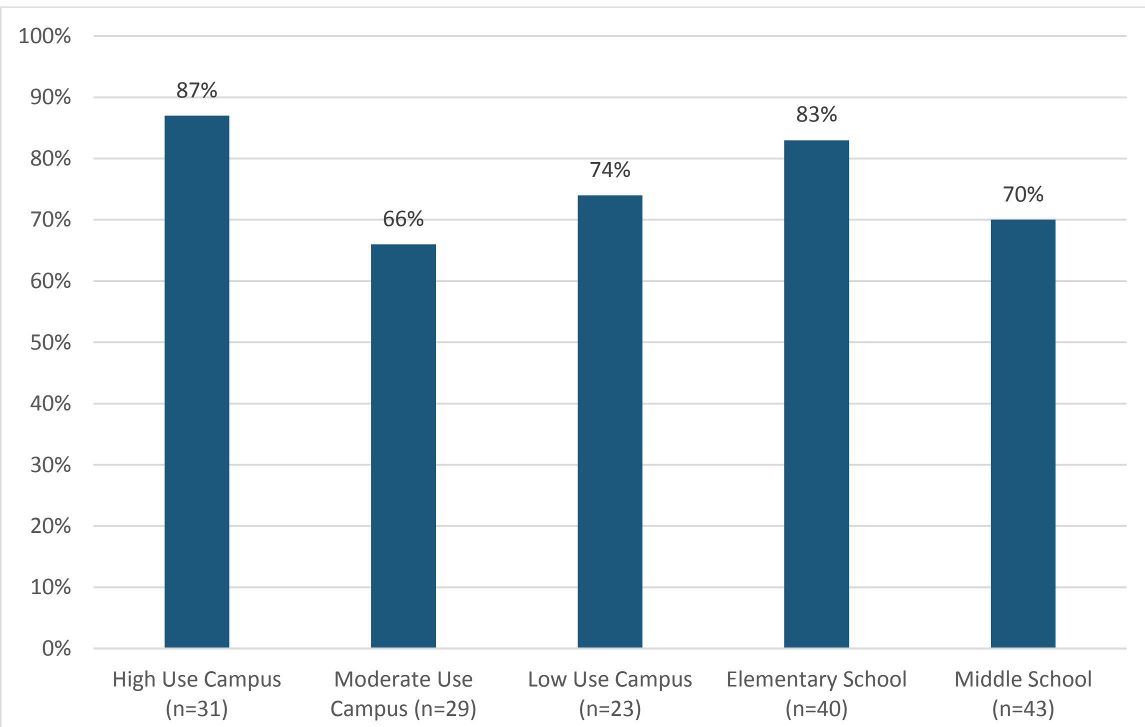
	Overall (n=79)	High Usage (n=30)	Moderate Usage (n=28)	Low Usage (n=21)	Elementary School (n=39)	Middle School (n=40)
Instructional staff	87%	90%	82%	90%	85%	90%
Educational aides	68%	79%	57%	65%	65%	71%
Computers/laptops	56%	73%	44%	48%	55%	58%
Internet connectivity	87%	80%	96%	86%	87%	88%

Source: Spring 2014 Interviews with campus-level staff.

Training

The majority of interviewees (76%) indicated that staff at their campuses were trained on how to use the TTM program. However, as Figure 3.1 illustrates, high TTM usage campuses (87%) were more likely to indicate their staff had been trained on the system than moderate (66%) or low TTM usage (74%) campuses. Staff at elementary schools were somewhat more inclined to report that they had been trained in the online mathematics program (83%) than middle school staff (70%).

Figure 3.1. Percent of Campuses Indicating Staff Were Trained on How to Use the TTM System



Source: Spring 2014 Interviews with campus-level staff.

Overall, 60% of the interviewees said that the training their campus received was sufficient for them to effectively use the TTM system. Staff at middle schools (69%) were more likely than elementary school staff (52%) to feel that the TTM training they and their campus staff received was sufficient.

Interviewees were asked to describe any additional training on the TTM system that they would like to receive. Of the 81 interviewees who provided responses to this item, the largest proportions sought information about best practices in using the TTM program (31%), general overview of the program and its content (20%), how to create or change pathways for students (17%), and how to create reports (14%).

Guiding Research Question 2b: Summary of Findings

Based on the TTM usage data (i.e., percent of students attempting 5 or more and 10 or more TTM sessions, and the percent of students passing 5 or more and 10 or more TTM lessons), it does not appear that schools and teachers were systematically identifying students for the TTM intervention. There was little or no descriptive relationship between 2012-13 STAAR-Mathematics assessment scores (i.e., quartile ranking) and the number of lessons attempted (e.g., five or more, 10 or more) during the 2013-14 school year. The same was true for other student populations, such as 2012-13 grade retention status, ELL status, and other demographic student characteristics.

To gain a better understanding of reasons for variation in TTM usage across Texas campuses, the evaluation team conducted telephone interviews with mathematics interventionists responsible for implementing TTM and other mathematics programs in their respective districts and schools. Over three quarters (78%) of campus staff interviewed indicated that they experienced some barriers when attempting to implement the TTM program at their schools. The most common barriers included:

- Not having enough computers (35%)
- Not having enough time in the daily schedule (26%)
- Technology issues (23%)

Training does not appear to be a major issue for TTM implementation as the majority of interviewees (60%) felt that the training staff at their campus received was sufficient for them to effectively use the system. However, staff at middle schools (69%) were more likely than staff at elementary schools (52%) to feel that the TTM-related training was sufficient.

Guiding Question 3b:

To what extent are program supports available through the TTM vendor, and ESC 20 staff, being utilized by participating districts and campuses?

Key Findings:

- **It appears as though campus and district staff were provided ample opportunities to attend training or receive technical assistance through ESC 20, and a substantial proportion of districts had staff attend face-to-face training at the regional ESCs on the TTM system.**
  - ✓ **Staff at under half of Texas school districts attended trainings on TTM held at regional ESCs.** Thirty-five different training opportunities were provided by TTM staff at the 20 regional ESCs in 2013-14. At least one training was held at each of the service centers. Over 1,600 staff members (typically campus staff) from 438 different school districts (43%) attended TTM training at regional ESCs during the 2013-14 school year.
  - ✓ **Service center staff provided technical assistance and support to district and campus staff during the 2012-13 and 2013-14 school years.** Of the 3,443 technical assistance inquiries received over the September 2012 to March 2014 period, 1,901 (55%) were exclusively related to TTM and 576 (17%) involved questions about both Istation and TTM. Most calls (88%) to ESC 20 were inquiries related to signing students up for or logging in to the online programs. The majority of campus mathematics interventionists interviewed in spring 2014 (76%) indicated that staff at their campuses were trained on how to use the TTM program.
- **Overall, campus-level staff felt their TTM training was sufficient for implementing the program at their campuses, though staff at elementary campuses and campuses with lower usage rates were less inclined to feel this way.** Staff at 60% of campuses said that the training they received was sufficient for them to effectively implement TTM. However, just over half of staff at elementary campuses and campuses with low system usage rates felt training was sufficient, compared to 69% of middle school staff.

Telephone-Based Technical Assistance Provided by ESC 20 Staff

As noted in Section 2 of this report, staff at ESC 20 provided technical telephone support for the Istation and TTM programs during the 2012-13 and 2013-14 school years. Based on data reported by the service center, ESC 20 received a total of 3,443 support calls (2,581 in 2012-13 and 862 in 2013-14) for both the Istation and TTM programs over the September 2012 to March 2014 period.

Of the technical assistance inquiries received over the September 2012 to March 2014 period, 1,901 were exclusively related to TTM and 576 involved questions about both the Istation and TTM programs. The remainder of the calls were related to TTM or some other unrelated issue. Most of the calls (88%) fielded by ESC 20 staff were related to how to sign students up for or log in to the programs. These figures are not disaggregated by whether in inquiry was related to Istation or TTM.

In-Person Professional Development Provided by TTM Staff at Regional ESCs

Data provided by TTM and the service centers show that 35 different training opportunities were provided by TTM staff at the 20 regional ESCs in 2013-14. At least one TTM training was held at each of the service centers. As Table 3.14 shows, staff from 438 school districts (typically campus staff) were documented as having attended training related to the online mathematics program. Overall, excluding charter schools, approximately 43% of the school districts in Texas had staff from one or more of their campuses attend the training. The proportion of districts attending TTM training was highest in Region 19 (El Paso), Region 1 (Edinburg), Region 4 (Houston), and Region 13 (Austin). District-level attendance at TTM training was lowest in Region 15 (San Angelo), Region 16 (Amarillo), and Region 17 (Lubbock). Over 1,600 school district and campus staff participated in the face-to-face TTM training during the 2013-14 school year. (See Table 3.14.)

Table 3.14. TTM Professional Development Provided by Vendor at Regional ESCs, 2013-14

ESC	Date(s) of Training	Number of Districts Served in the Region	Number of Districts Represented at Training	Percent of Districts Represented at Training	Number of Attendees at Training
1 (Edinburg)	10/28/13	37	26	70.3%	111
2 (Corpus Christi)	11/21/13 & 2/11/14	42	14	33.3%	43
3 (Victoria)	2/11/14	39	17	43.6%	69
4 (Houston)	11/12/13 & 2/13/14	51	34	66.7%	135
5 (Beaumont)	2/4/14	32	18	56.3%	71
6 (Huntsville)	11/1/13 & 5/15/14	56	23	41.1%	90
7 (Kilgore)	2/10/14	96	38	39.6%	104
8 (Mount Pleasant)	10/23/13	47	16	34.0%	36
9 (Wichita Falls)	1/13/2013 & 1/24/14	37	14	37.8%	37
10 (Richardson)	2/13/14 & 4/16/14	80	32	40.0%	157
11 (Fort Worth)	11/11/13	76	30	39.5%	93
12 (Waco)	10/28/13 & 12/10/13	77	25	32.5%	69
13 (Austin)	2/11/2014	56	35	62.5%	116
14 (Abilene)	10/31/13 & 2/28/14	42	23	54.8%	87
15 (San Angelo)	11/18/13 & 2/12/14	42	8	19.0%	13
16 (Amarillo)	10/17/13 & 3/21/14	62	14	22.6%	50
17 (Lubbock)	10/9/13 & 2/11/14	57	10	17.5%	34
18 (Midland)	10/16/13	33	19	57.6%	107
19 (El Paso)	2/4/14	12	12	100%	52
20 (San Antonio)	10/15/13 & 2/11/14	52	30	57.7%	139
Total		**1,026**	**438**	**42.9%**	**1,611**

Source: Texas Education Agency, 2014.

Note: District counts do not reflect charter school organizations, but TTM training participation counts do include charter organizations, resulting in inflated participation rates in some cases.

Guiding Research Question 3b: Summary of Findings

Statewide efforts to support widespread usage of the TTM system included technical phone support provided by ESC 20 in San Antonio, webinars sponsored by the TTM vendor, and face-to-face professional development provided by TTM at each of the 20 regional ESCs across the state. While data are not available on the number of onsite trainings conducted, it has also been reported that TTM staff served district and campus staff through onsite training and technical support. It appears as though campus and district staff were provided with adequate and varied opportunities to learn how to use the TTM system to support students in mathematics in their respective jurisdictions.

Of the 3,443, technical assistance inquiries received over the September 2012 to March 2014 period, the majority were related to the TTM program. In total 1,901 were exclusively related to TTM and 576 involved questions about both the Istation and TTM programs. Over 1,600 staff members from 440 of the 1,026 school districts in Texas attended TTM training at regional ESCs during the 2013-14 school year. Another 526 staff members from 201 districts participated in TTM-related webinars.

The majority of campus mathematics interventionists interviewed in spring 2014 (76%) indicated that staff at their campuses were trained on how to use the TTM program. Not surprisingly, campuses using the TTM system heavily were more likely (87%) to indicate their staff had been trained on the system than moderate (66%) or low TTM usage (74%) campuses. Staff at elementary schools were somewhat more inclined to report that staff had been trained in the online mathematics program (83%) than middle school staff (70%).

Overall, 60% of the interviewees said that the training their campus received was sufficient for them to effectively use the TTM system. Staff at middle schools (69%) were more likely to feel that the TTM training their campus staff received was sufficient than their elementary school counterparts (52%). District staff also appeared to be generally satisfied with the quality of training received from TTM; however, they felt the training could have been more targeted to specific system features that could have benefitted teachers more.

Based on data provided to the evaluation team, it appears as though there were ample opportunities for school districts to participate in training opportunities for the TTM program, and that the training was generally satisfactory. However, the training seems to have better met the needs of middle school staff than elementary school staff, and may have benefited from greater specificity for teachers.

Guiding Question 4b:

How are school districts and campuses using TTM to address the SSI grade promotion requirements for mathematics, and what other mathematics resources are being used in combination with Texas SUCCESS programs to further support the learning needs of students to meet the SSI grade promotion requirements?

Key Findings:

- **Initial decisions regarding TTM adoption typically occurred at the district level, though campus staff made the majority of the decisions regarding how TTM was used at particular campuses.** Some district-level staff also provided recommendations on TTM use and monitored TTM implementation, but decisions regarding how TTM was used for instruction and assessment purposes were made by campus-level staff.
- **Some commonalities in usage were evident across schools.** Almost three out of four interviewees reported that TTM had been in use in their schools for both the 2012-13 and 2013-14 school years. Staff at both elementary and middle schools reported that TTM was used similarly in all tested grades. In terms of instruction, schools most commonly reported using TTM during regular classroom instruction (72%), as a pullout intervention for struggling students (59%), and/or as part of their after-school program (39%).
- **Most campuses began using TTM in early fall, but the consistency of usage throughout the school year varied across campuses.** Many campuses (68%) began using TTM in the fall (August—October, 2013), and used the program for the duration of the 2013-14 school year. This was particularly true of more frequent TTM users, with 90% of high usage campuses—versus 52% of low usage campuses—reporting that they were still using the program at the time of campus interviews (May—June, 2014).
- **TTM was rarely the only mathematics intervention program utilized at campuses to support students in meeting SSI grade promotion requirements.** With regard to all mathematics interventions and programs used, only 18% of school staff reported using only TTM, while 82% indicated that they used at least one other mathematics program.
- **Variability in TTM use was tied to levels of system use and school level.** Almost half of middle schools used TTM exclusively in computer labs, while less than 33% of elementary schools did so. In elementary schools, TTM was typically used in blended classroom settings (i.e., classrooms with dedicated computers). Middle schools were more likely to report identifying students for the TTM intervention based on prior performance on standardized assessments. Staff at high TTM use campuses were more likely than staff at moderate or low usage campuses to report that they monitored TTM use, that use was consistent throughout the year, and that the system was implemented with fidelity.

The data used to address this research question were collected from telephone interviews with campus- and district-level staff responsible for either implementing TTM and other mathematics interventions at the campus level, or coordinating mathematics interventions and other SSI-related mathematics programs across campuses in their respective districts. Data presented in this section are based on campus-level interviews with 82 staff members and 30 district-level interviews. The interviews included a series of questions related to approaches used by school districts and campuses to implement the online TTM program and mathematics-related interventions to address the grade promotion requirements of SSI (see Appendices F and G for the interview protocols). The following sub-questions are explored in this section:

1) How did school districts and campuses use TTM to address SSI requirements for mathematics?
2) What resources and staff were used by campuses to address SSI grade promotion requirements?
3) How was the use of TTM coordinated with other mathematics interventions?

How Districts and Campuses Used TTM to Address SSI Requirements

As with interviews conducted with district- and campus-level staff regarding various aspects of TTM implementation and use, school-level TTM usage (high, medium, or low), was used to select the interviewee sample. Also similarly, district-level interviews provided contextual information regarding district-wide policies and practices related to TTM implementation, while campus-level results are presented in aggregate, by school-level and TTM usage patterns. The paragraphs below summarize findings from both district- and campus-level interviews, starting with a description of district-level findings related to TTM implementation. Following that, the remainder of this section focuses on patterns of TTM implementation and use at the campus-level.

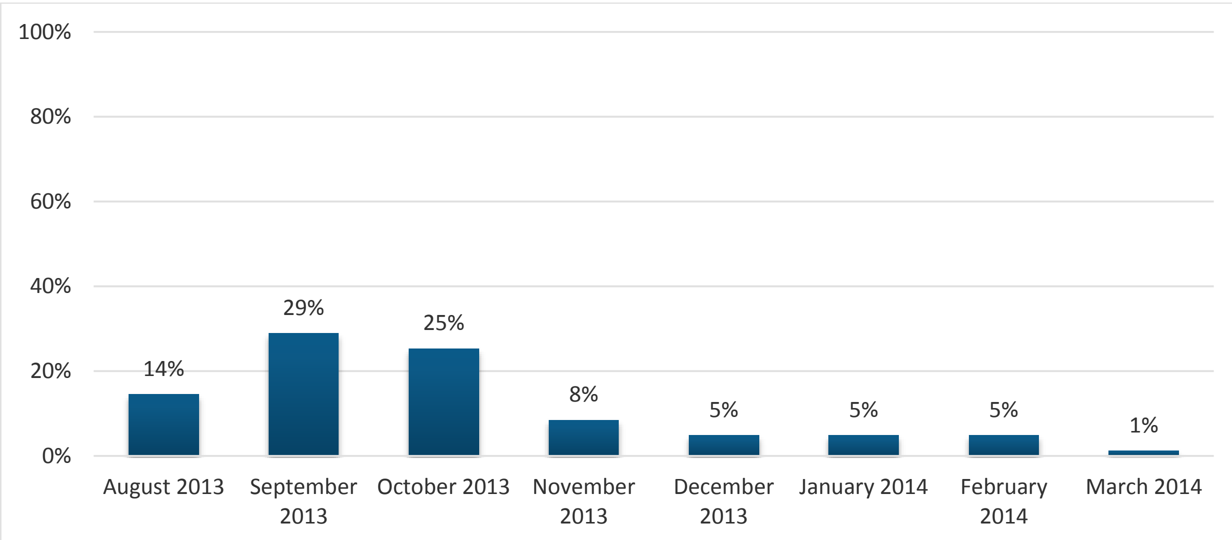
District-wide Policies and Practices Related to TTM

Most initial decisions about implementing TTM were either completely centralized or made via discussions between districts and schools—only eight districts (28%) reported decentralizing the initial decision-making process with regard to implementing TTM. Outside of the role districts played in initial decisions regarding TTM implementation, centralized mandates about how the system should be used were few. For example, several district-level staff mentioned that they set minimum expectations for use (e.g., minimum number of lessons completed). Others reported that they set requirements about which students should be served by the TTM intervention (e.g., students in specific grades and/or at certain performance levels). Besides these district-level expectations, specific choices about TTM—that is, how, when, where, and with whom the program would be used—were largely left to schools.

Campus-Level Practices Related to TTM

Generally, TTM program use varied across schools, with differences noted between elementary and middle schools as well as differences by schools' TTM usage pattern. There were, however, commonalities regardless of school-level and usage pattern. On average, schools have been using TTM for about two years with 71% of schools reporting that they used the program prior to the 2013-14 school year. Also, in 2013-14, most campuses began implementing the program in the fall (Figure 3.2) and reported that they were still using TTM at the time of the interview. This was particularly true of high usage TTM campuses, where high use schools were more likely to still be using the program at the time of the interviews (90%) compared to campuses designated as moderate or low usage schools (79% and 52%, respectively).

Figure 3.2. Month Campuses Began Using TTM, 2013-14 School Year



Source: Spring 2014 Interviews with campus-level staff.

Note: N=83. Seven percent of campus staff indicated that TTM usage for 2013-14 began sometime after March 2014.

Campus staff indicated a number of different ways in which they used TTM to help improve their students' mathematics skills. It was most common for schools to use TTM in the regular classroom (72%), as a pullout intervention for struggling students (59%), and/or as part of their after-school program (39%). And, similar to Istation, most interviewees (71%) reported that the targeted number of hours students used TTM was variable depending upon a variety of factors (e.g., student performance, student group, and grade-level).

With regard to grade-level, specifically, high percentages of elementary school staff indicated that TTM was used with students in Grades 3, 4, and 5 (82%, 85%, and 85%, respectively). In other words, TTM use was relatively evenly distributed across grades at the elementary level. This was also true in middle school grades, with most middle school staff reporting that TTM was used in Grades 6, 7, and 8 (84%, 88%, and 86%, respectively). Generally, reported TTM use was fairly equal across grade levels. (See Table D15 in Appendix D.)

Campus staff were also asked about whether certain students were identified for the TTM intervention or if it was made available to all students at their schools. The evaluation team queried campus interviewees as to whether students were identified for TTM by: 1) grade level; 2) the results of district benchmark assessments; 3) STAAR assessment results; and 4) student subpopulations (e.g., students identified as at-risk or ELL). Relatively few campus staff indicated that they focused on specific grade levels for the TTM intervention, and there was little variation in identification of students by grade-level (see Table 3.15).

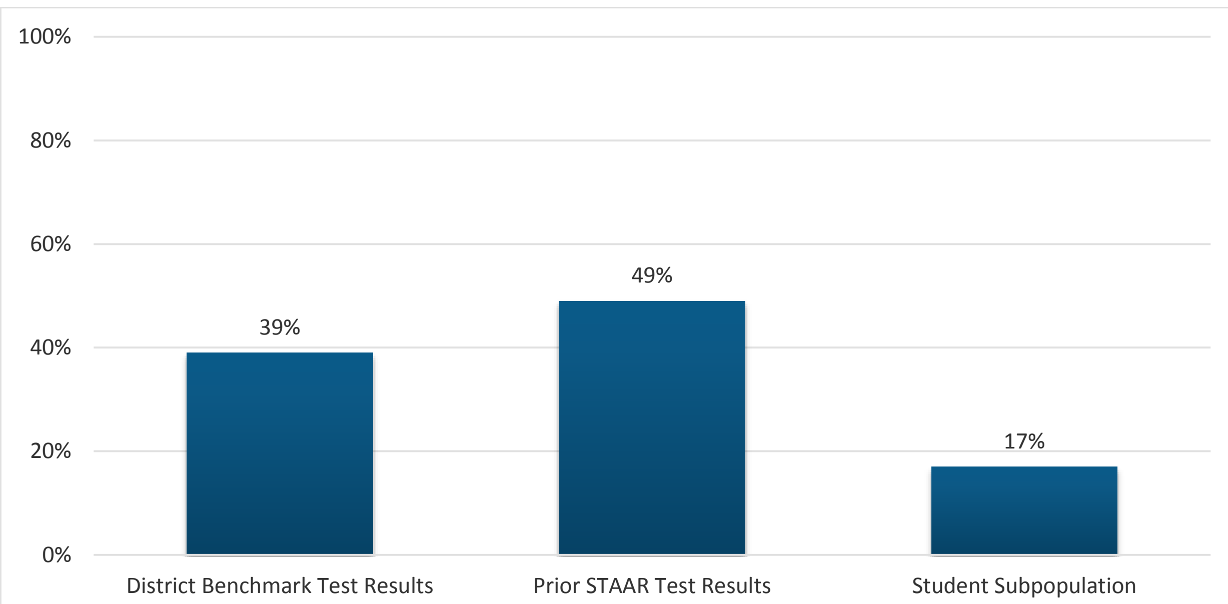
Table 3.15. TTM Identification by Grade Level, 2013-14

Grade Level	Percent of Respondents Identifying Students for the Intervention by Grade Level (N=83)
3	18%
4	18%
5	13%
6	13%
7	13%
8	18%

Source: Spring 2014 Interviews with campus-level staff.

Figure 3.3 presents the percentage of campuses reporting targeting students for TTM usage based on prior performance on STAAR, interim district benchmark tests, or by student subpopulations. About half of campuses reported targeting students for the TTM intervention based on prior STAAR results (49%), and 39% identified students for the TTM program based on the results of district benchmark tests. A smaller proportion of campus staff reported that their schools identified students for the TTM intervention based on student subpopulations (17%).

Figure 3.3. Methods Used by Campuses for Identifying Students for TTM Use, 2013-14



Source: N=83. Spring 2014 Interviews with campus-level staff.

Different Patterns of TTM Usage Across Campus Types

Elementary versus Middle Schools

The TTM program was delivered in a variety of different school settings (i.e., regular classroom, computer lab, or a combination of those settings). Most commonly (45%), staff reported that the TTM program was used in both the classroom and a computer lab setting. Forty percent of the interviewees noted that the program was used exclusively in computer labs, and 15% said that students used TTM in their regular classrooms. When considering differences in setting by school-level, middle schools tended to use TTM more in computer labs (46%) than elementary schools (33%). Elementary schools were more likely than middle schools to utilize a blended approach of regular classroom/computer lab setting (51% versus 39%, respectively). While not collected in the interviews, it is important to note that TTM has a sizable proportion of students using the program out of regular school time, at home. See Table D17 in Appendix D for further detail.

In terms of identifying students for the TTM intervention, middle schools tended to use prior STAAR results for identifying students for the program (56% versus 41% of elementary schools); while elementary schools tended to use benchmark tests to identify students for TTM usage (45% versus 33% of middle schools).

High, Medium, and Low Usage Campuses

When asked about how the TTM system was being used at the campus level (i.e., for assessment purposes, curriculum, or both), less than half of interviewees noted that TTM was used for both assessment and curriculum (48%). The same proportion of interviewees reported using the mathematics program for curriculum only (48%). When responses about the use of different TTM functions were disaggregated by usage patterns, all high usage campuses noted they used the program for both assessment and curriculum. By contrast, 68% of moderate use campuses and 71% of low usage campuses noted using both functions.

Other differences by TTM usage pattern were reported in terms of changes in schools' use of the system over time—that is, differences between TTM use before 2013-14 and during the 2013-14 school year. Changes in TTM use were reported more frequently among low usage campuses, with 29% of staff from low use schools reporting declines in TTM usage in 2013-14. (See Table D16 in Appendix D.)

Overall, 79% of all interviewees indicated that they had some type of system for monitoring TTM usage. High usage TTM campuses (90%) were more inclined to have monitoring systems in place for the programs than moderate (76%) or low (67%) usage campuses. Of those that did monitor usage, TTM use was most commonly monitored by teachers (16%) or jointly by teachers and administrators (26%). When asked what was being monitored, overall, the largest proportion of campuses reported that student progress on the system was most commonly tracked (36%). Middle schools were more likely to monitor for student progress (43%) than elementary schools (28%). (See Table D18 in Appendix D.)

Campus interviewees also indicated whether the TTM program was available for student use consistently throughout the 2013-14 school year (i.e., availability was not discontinued at any time during the year). Overall, 63% of interviewees reported that TTM was made available to students for the entire school year. However, consistent availability of the system was more prevalent among high TTM usage campuses (83%) than moderate or low TTM usage campuses (46% for both). (See Table D19 in Appendix D.)

Overall, 54% of the interviewees indicated that they felt the program was implemented with fidelity. However, a great deal of variation was reported by system usage level. The majority of high usage campuses (80%) noted the program was implemented with fidelity compared to 39% of moderate and 33% of low usage campuses. (See Table D19 in Appendix D.)

Resources and Staff Used by Campuses to Address SSI Requirements

In addition to examining the implementation of the online TTM mathematics program, the evaluation team also sought to understand other supports used to help students struggling with mathematics at the campus-level. Results from interviews with school-based staff revealed that TTM was rarely the only mathematics intervention utilized by campuses to help students meet the SSI grade promotion requirements. A total of 50 different mathematics programs were being used in addition to TTM across the 83 campuses included in the interview sample. Approximately four out of five campus-level interviewees reported that they used a variety of mathematics supports and programs, along with TTM, to help students meet SSI requirements in 2013-14. In other words, staff at less than 20% of campuses indicated that TTM was the only mathematics program used at their school. Staff at 43% of schools indicated that they used one additional mathematics program, 19% said that they used TTM and two additional mathematics programs, and almost another 20% reported that they utilized three or more other programs in addition to TTM at their campuses (see Table 3.16).

Table 3.16. Number and Percent of Other Resources Used to Address SSI Requirements, 2013-14

	Number of Other Resources	Percent
No programs other than TTM	15	18%
One other program	36	43%
Two other programs	16	19%
Three other programs	8	10%
Four other programs	7	8%
Five or more other programs	1	1%
Total	**83**	**100%**

Source: N=83. Spring 2014 Interviews with campus-level staff.

As Table 3.17 shows, interviewees shared a variety of ways (besides TTM assessment data) they determined which mathematics interventions were appropriate for students. Those most commonly noted included district or campus benchmark tests (25%), prior STAAR results (19%), teacher knowledge of students (14%), and class performance (11%). The TTM assessment was used in only 6% of the schools to determine which intervention might be appropriate for students.

Table 3.17. Major Themes, Respondents' Descriptions of How Interventions Were Determined by Usage and School Level, 2013-14

	Overall (n=80)	High Usage (n=27)	Moderate Usage (n=29)	Low Usage (n=24)	Elementary School (n=39)	Middle School (n=41)
Campus or district benchmark tests	25%	19%	31%	25%	26%	24%
Prior STAAR results	19%	22%	14%	21%	23%	15%
Teacher knowledge of students	14%	15%	21%	4%	10%	17%
Class performance	11%	15%	10%	8%	10%	12%
Online program assessment	10%	15%	14%	0%	10%	10%
TTM assessment	6%	7%	7%	4%	5%	7%
Response to Intervention (RTI)	5%	7%	3%	4%	8%	2%
Teacher assessments	5%	4%	7%	4%	5%	5%

Source: Spring 2014 Interviews with campus-level staff.

Note: Percentages may not total to 100% because not all responses could be categorized into reported themes.

Interviewees mentioned different instructional strategies and settings used to help students meet SSI requirements in mathematics, including in- and out-of-school strategies, such as:

- Other online mathematics programs as homework or used outside of regular class time (39%)
- Tutoring (30%)
- Use of other mathematics curricula or programs during class time (28%)
- Small group mathematics instruction (20%)
- Saturday or summer school (20%)
- Pulling students out of their regular classes for additional mathematics assistance (16%)

Minor differences were observed between elementary and middle schools in the areas listed above. Staff from moderate usage campuses reported using in-class strategies and out-of-school strategies more often than high or low usage campuses.

About two thirds of interviewees (66%) indicated that, aside from classroom teachers, additional support staff were used to help implement the TTM program at their campuses. Lower availability of support staff was most evident at low TTM usage campuses, where 65% shared that no additional staff were used to implement TTM, versus 17% of high usage campuses and 26% of moderate use schools. Overall, educational aides and interventionists/coaches were most commonly cited as supporting TTM implementation (26% for each). However, taken together, high and moderate usage campuses had more access to educational aides and interventionists/coaches (see Table 3.18).

Table 3.18. Percent of Campuses Indicating TTM Support Personnel by Usage Level

	Overall (n=70)	High Usage (n=23)	Moderate Usage (n=27)	Low Usage (n=20)
No support staff assists	34%	17%	26%	65%
Educational aides	26%	43%	19%	15%
Interventionists/ Coaches	26%	17%	44%	10%
Computer lab staff	13%	17%	15%	5%

Source: Spring 2014 Interviews with campus-level staff.

Taking into account differences in access to support staff, it is perhaps unsurprising that smaller proportions of staff at low and moderate usage campuses reported that their schools had adequate support staff to effectively implement TTM (see Table 3.19). Campus-level interviewees were also asked about whether their campuses had adequate access to computers/laptops and internet connectivity to implement TTM effectively. High percentages of interviewees across all groups indicated that they had sufficient internet connectivity to implement TTM well (80% or better). There were differences, however, between reports of adequate computer resources. It appears that the lack of computers was an issue for campuses that used the program at moderate and low levels. Low and moderate usage campuses were more likely than high usage campuses to mention that they had enough computers to effectively implement the TTM program (48% and 44% versus 73%, respectively). District-level staff also cited computer access and technological resources as a factor limiting implementation success.

Table 3.19. Percent of Campuses with Enough Resources to Effectively Implement TTM

	Overall (n=79)	High Usage (n=30)	Moderate Usage (n=28)	Low Usage (n=21)	Elementary School (n=39)	Middle School (n=40)
Instructional staff	87%	90%	82%	90%	85%	90%
Educational aides	68%	79%	57%	65%	65%	71%
Computers/laptops	56%	73%	44%	48%	55%	58%
Internet connectivity	87%	80%	96%	86%	87%	88%

Source: Spring 2014 Interviews with campus-level staff.

Coordination of TTM with Other Mathematics Interventions

Considering whether other programs were coordinated with TTM, less than half of interviewees (42%) indicated that this coordination was occurring. In other words, most campus-level interviewees did not report systematic coordination of resources designed to support students in meeting grade-level requirements. The most commonly cited issue was related to logistics and time—37% of interviewees who reported coordinating interventions cited that planning and scheduling among different teachers and students who needed access to the resources was a challenge. Coordinating interventions that addressed students' mathematics needs as well as other areas in which they were struggling was cited as another obstacle. Garnering student and teacher buy-in was cited as a challenge, though less frequently than challenges such as time, scheduling, and coordinating multiple interventions for students.

Among those who reported coordinating services, interviewees commonly noted using different resources and programs during out-of-class learning activities, including tutoring or afterschool programs. Interviewees also noted that schools modified students' schedules if their data indicated that they were having trouble with mathematics. It was during these non-traditional, out-of-class types of settings that coordination of interventions was reported most often.

Guiding Research Question 4b: Summary of Findings

At the district-level, staff were most commonly involved in initial decisions about whether to implement TTM. Some districts also set expectations for how schools would use the system and/or monitored TTM use. Generally, campuses made most decisions regarding TTM use. Across all campuses, almost half of interviewees noted using just TTM curricular resources and almost three quarters of interviewees noted that the system was typically used to support regular classroom instruction.

Variability in TTM use was tied to levels of system use and school-level:

- High usage campuses were more likely than moderate or low usage campuses to report consistent usage, monitoring of TTM use throughout the year, and that TTM was being implemented with fidelity, compared to moderate or low use campuses.
- Almost half of middle schools used TTM exclusively in computer labs, while the same was true in 33% of elementary schools. In elementary schools, TTM was typically used in blended classroom settings (i.e., classrooms with dedicated computers).
- Middle schools were more likely to report identifying students for the TTM intervention based on prior performance on standardized assessments.
- Low TTM usage campuses were more likely to identify students for TTM use based on grade-level.
- Low TTM use campuses were less likely to monitor system usage.

Among those campuses implementing TTM for two or more years (i.e., prior to 2013-14), low usage schools were more likely to report changes over time, and were more likely than high or moderate usage campuses to report those changes as reductions in use. Outside of TTM, most district- and campus-level staff reported using other programs to support students in mathematics, with only 18% of schools reporting that they only used TTM and 82% indicating that they used at least one other mathematics program. Less than half of interviewees indicated that coordination between TTM and other mathematics interventions occurred, citing a lack of time and scheduling difficulties as reasons for the lack of coordination. Teachers most often supported students in using these mathematics programs, though educational aides and instructional coaches also played a role. There were differences reported in access to support staff, with over 70% of staff at high and moderate usage campuses reporting access to support staff, while only 35% of low use campuses reported the same. Low and moderate usage campuses were also less likely than high usage campuses to report that their schools had sufficient access to computers to implement mathematics programs.

Student Outcomes Related to TTM and Other SSI-Related Mathematics Interventions

Relationships between TTM implementation and mathematics outcomes were explored in a variety of ways. First, to explore the extent to which students showed progress within the system, the evaluation team considered whether students who used the system showed meaningful progress as evidenced by performance data contained in the online program. The extent to which this relationship varied by the fidelity of TTM implementation in different schools and districts was also considered (Guiding Question 5b). Second, the evaluation team explored relationships between school-level mathematics achievement and schools' implementation of a broader complement of SSI mathematics interventions—including TTM (Guiding Question 6b). Next, to explore relationships between TTM use and mathematics achievement on STAAR exams, the evaluation team used several analytic strategies to provide a multi-faceted view of those associations. For Guiding Question 7b, the evaluation team explored whether mathematics gains were related to categorical thresholds of TTM usage and TTM passing rates. In other words, analyses explored whether different dosages and whether students' success rate at passing TTM lessons resulted in stronger associations between TTM participation and mathematics achievement.

For Guiding Question 8b, the evaluation team explored relationships between STAAR test performance on the first administration of the STAAR-Mathematics exam and the continuity and the proximity of TTM usage to the test administration date. In sum, by exploring both the continuity and timing of system use, these analyses were designed to build on those conducted for the previous question by examining whether associations between students' STAAR performance and TTM use were stronger when use increased in proximity to the test.

To explore whether relationships between TTM use and mathematics achievement would vary among students whose academic performance classified them as at risk for academic difficulties in 2013-14, the evaluation team explored:

- For Guiding Question 9b, the extent to which relationships between TTM use and mathematics performance differed among students at risk of being retained versus those not at risk.
- For Guiding Question 10b, the evaluation team explored the relationship between program usage after failing the first administration of the STAAR-Mathematics test and the probability of passing a second administration of the exam.

Guiding Question 5b:

To what extent are students participating in the Texas SUCCESS program showing meaningful progress as evidenced by performance data contained in the TTM online program, and to what extent was program implementation fidelity related to student growth on the TTM online program?

Key Findings:

- **Most TTM users passed at least one lesson in 2013-14; however, of all lessons attempted, less than half were passed.** Passing rates were highest among students who attempted the most TTM lessons and lowest among students who attempted the least number of lessons.
- **Passing rates varied by grade level, student groups, and the month in which lessons were attempted.** A higher percentage of elementary students passed at least one lesson, versus middle grade students (96% to over 99% versus 85% to 89%, respectively). Passing rates were lower among students categorized as ELL and economically disadvantaged. In all but Grade 7, passing rates were highest in the beginning of the school year (August, 2013).
- **Most students progressed along their prescribed TTM pathway; and, of those who did get off path by being required to take a remedial lesson or failing a TTM lesson, most were able to get back on path.** Two out of three lessons completed in 2013-14 were "on path" (i.e., not remedial or retaken lessons). And, of those students who deviated from their prescribed pathways, roughly 95% of elementary and 85% of middle school students were able to get back on path.
- **There was no clear relationship between TTM lesson passing rates, usage levels, and the different types of lessons.** This was particularly true among students taking pathway or remedial lessons; that is, passing rates did not seem to vary as a function of the number of pathway or remedial lessons attempted.

This section of the report examines students' use of, and progress through, the TTM system. To do so, patterns of student usage and passing rates across the year are described along with several metrics for student progression through the system.

TTM Assessment Passing Rates, Overall and By Student Groups

Most students (91%) who attempted at least one TTM lesson during the 2013-14 school year completed at least one lesson. However, completing a lesson did not mean that a student passed that lesson; students were able complete a lesson that they did not pass. After failing a lesson, students were directed to either retake a lesson or receive remedial instruction. To accommodate this function of the system, rather than focusing on whether students attempted or passed five or more or ten or more lessons, this section describes student progression through the system as the ratio of the number of lessons a student passed to the number of lessons attempted.[63]

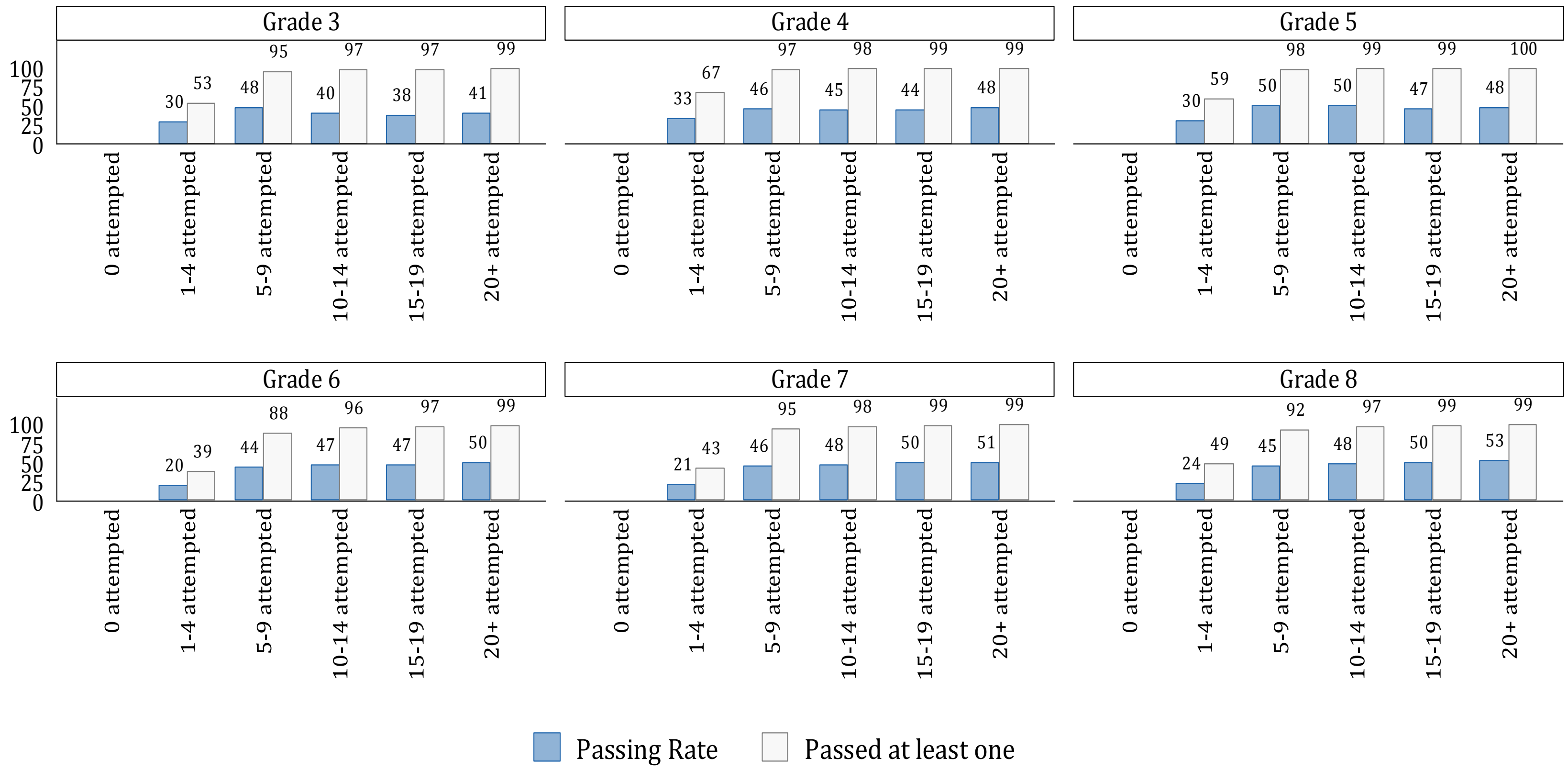
Along with passing rates, this section explores progression through the system by examining changes in attempts and passing rates across the school year as well as student persistence and progression through the TTM system after passing or failing lessons.

Overall, about 94% of TTM users passed at least one lesson in 2013-14. That rate was higher for students in Grades 3- 5 (96.0 to 99.6%) than Grades 6-8 (85.3 to 89.3%). The overall average passing rate for all TTM users was about 43%, meaning just under half of attempted lessons were passed. This rate was slightly higher for students in Grades 3-5 (43.9%) than students in Grades 6-8 (39.4%).

Figure 3.4 shows the percent of students who passed at least one lesson and passing rates by lessons attempted. For students who attempted five or more lessons, the passing rate remained around the 50% mark for all grade levels. However, for students in Grade 3, the passing rate declined from 48% for students with 5-9 lessons attempted to 38% for students with 15-19 attempts and 41% for students with 20 or more lessons attempted.

[63] For analyses conducted for this section, TTM passing rates were calculated by dividing the total number of lessons passed by the number attempted.

Figure 3.4. Percentage of Students Who Passed at Least One TTM Lesson and Average Student Passing Rate, by Grade Level and Number of TTM Lessons Attempted, 2013-14



Source: Think Through Math session history table and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Table 3.20 illustrates passing rates by student ELL status. Across all grades, passing rates were higher among non-ELL students, with the gap between non-ELL and ELL students increasing with grade level. In Grade 3, there was an 8 percentage point passing rate difference between non-ELL and ELL students (45% and 37%, respectively), while Grade 8 had the largest gap in passing rates with a 15 percentage point difference (43% and 28%, respectively).

Students classified as ELL in middle school did not perform as well on the TTM lessons as students in elementary school; however, this pattern did not hold as strongly for non-ELL students. Students classified as ELL in elementary school passed 36% of TTM lessons, compared to 28% in Grade 6, 29% in Grade 7, and 27% in Grade 8. In contrast, non-ELL students passed between 44% and 49% of lessons in elementary school grades, and 42% to 43% of lessons in middle school grades. This finding suggests that the TTM program may not have been particularly effective with students identified as ELL at the middle school level.

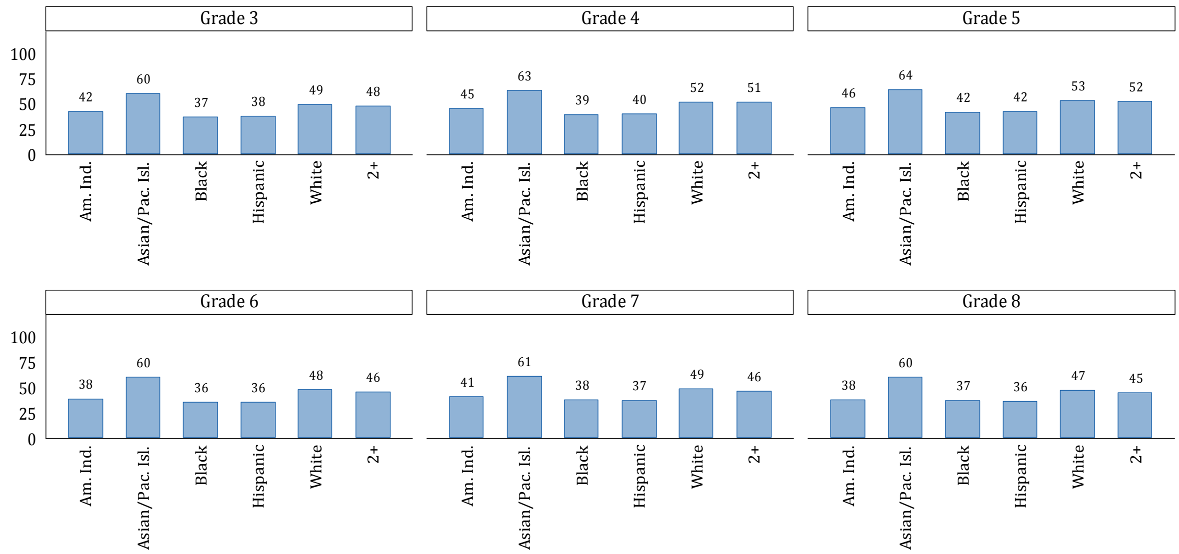
Table 3.20. Student Passing Rate by Student ELL Status, by Grade Level, 2013-14

Grade Level	Non-ELL Students	ELL Students
Grade 3	44%	36%
Grade 4	47%	36%
Grade 5	49%	36%
Grade 6	42%	28%
Grade 7	43%	29%
Grade 8	42%	27%

Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Figure 3.5 depicts the passing rate by race category. Across all grades, passing rates were highest among Asian students, with an over 60% passing rate in Grades 3 to 8. The gap between the highest passing rates and the lowest was over 20 percentage points across all grades, with passing rates around 40% among African American and Hispanic students across grades. Within race/ethnicity categories, passing rates were relatively stable across grades, with the largest fluctuation occurring between Grades 5 to 6, where declines in passing rates ranged from 3 to 8 percentage points (among Asian and American Indian students, respectively).

Figure 3.5. Average Student Passing Rate by Student Race/Ethnicity Category, by Grade Level, 2013-14



Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Table 3.21 presents the TTM lesson passing rate by student economic disadvantaged status. Across all grades, passing rates were higher among students not classified as economically disadvantaged compared to economically disadvantaged students, across all grades. The gap between non-economically disadvantaged and economically disadvantaged students' passing rates ranged from 12 percentage points (Grade 7) to 15 percentage points (Grade 3).

Table 3.21. Student TTM Lesson Passing Rate by Student Economic Status, by Grade Level, 2013-14

Grade Level	Not Economically Disadvantaged	Economically Disadvantaged
Grade 3	51%	36%
Grade 4	54%	39%
Grade 5	54%	41%
Grade 6	49%	35%
Grade 7	49%	37%
Grade 8	48%	35%

Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

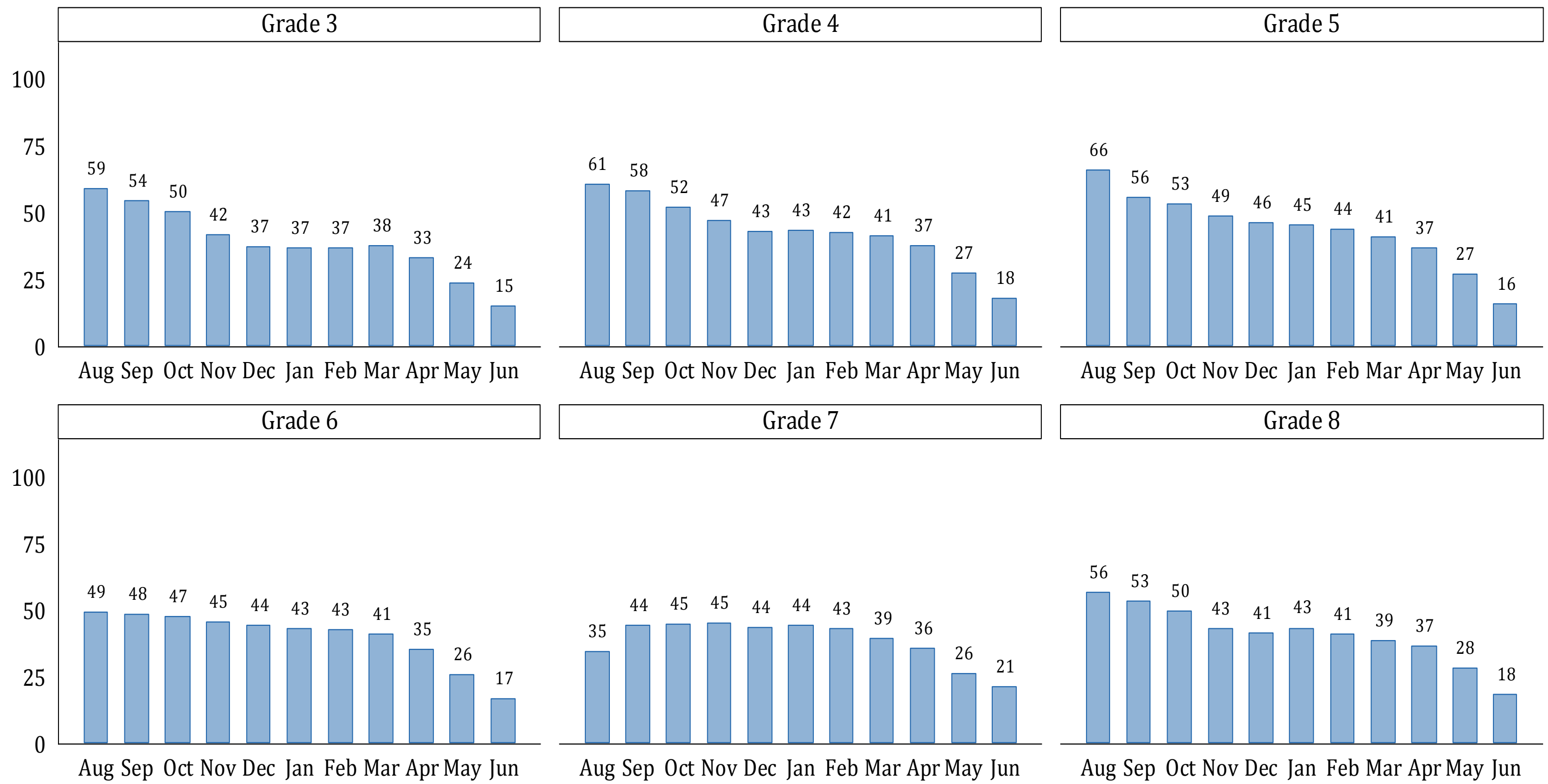
Student Progression in the TTM System

With passing rates described as a function of passed relative to the number of attempted lessons, student progression in the TTM system was explored in two ways: 1) average passing rates by month, and 2) the contrast between passing rates in the first and last months of TTM system use. Passing rates were only included for students who had TTM usage in at least two months. Passing rates for August and June were included; however, these rates should be interpreted with caution as the total number of attempts and students attempting lessons in those months was very low compared to the rest of the year.

Figure 3.6 shows the average student passing rate by month. With the exception of Grade 7, passing rates were highest in August and September, with rates declining or plateauing in subsequent months during the 2013-14 school year. In elementary grades, passing rates declined by over 20 percentage points between August and March, and typically leveled off from December to March. In Grade 6, passing rates declined by 8 percentage points from August to March and by 17 percentage points in Grade 8 over the same period. Passing rates plateaued in these grades between December and February. In Grade 7, passing rates were 10 percentage points lower in August than in the three subsequent months (35% versus 44%-45%).[64] This finding may suggest that students encountered more difficult material as the school year progressed, resulting in lower passing rates across most grades.

[64] Note that the reason for the drop in passing rates after April may be due to the fact that there were fewer students taking lessons, and many of those took lessons as a result of failing the first administration of the STAAR. This will be explored in Research Question 10b.

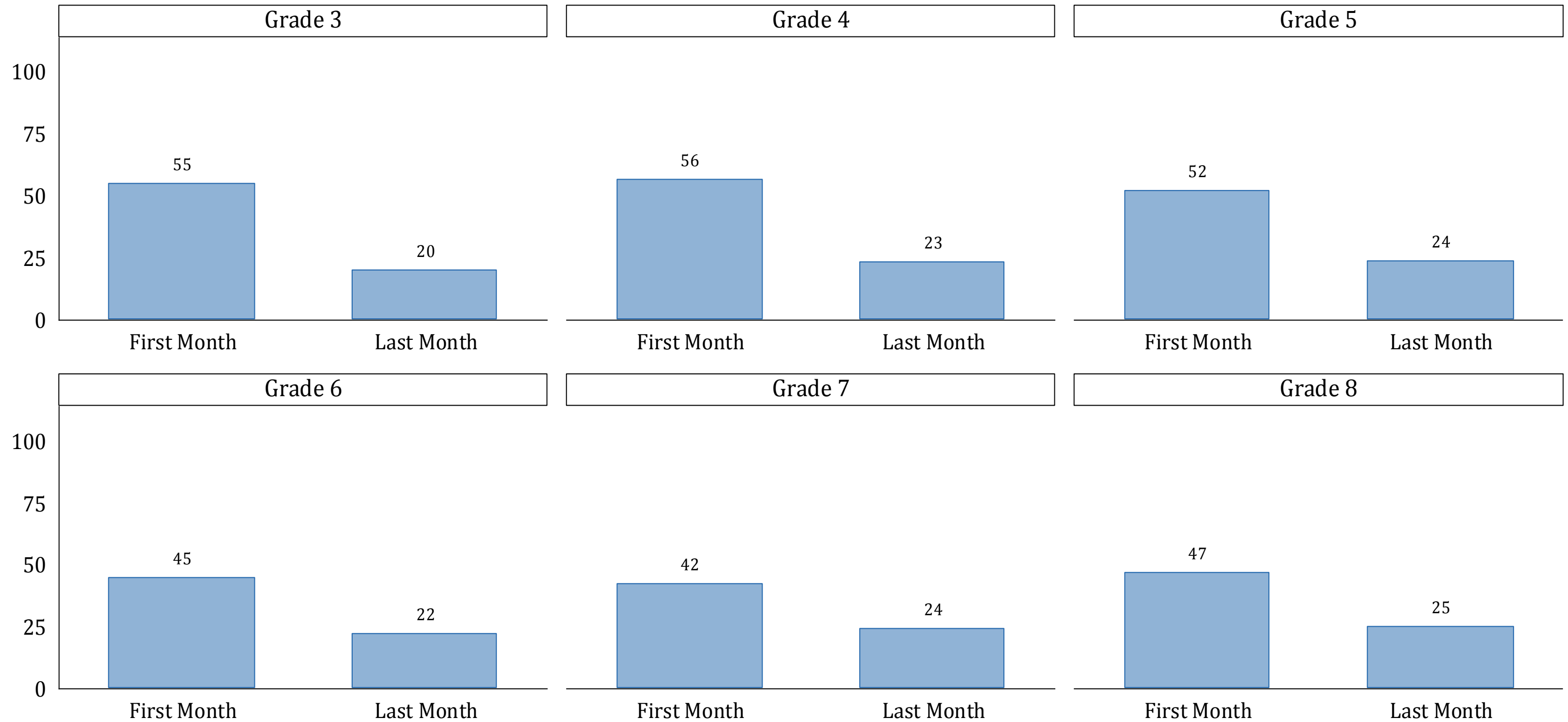
Figure 3.6. Average Student Passing Rate, by Grade Level and Month, 2013-14



Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Passing rates for each student's first and last month, for students with at least two months of TTM usage, are presented in Figure 3.7. While there were no parameters on when students were able to start and stop taking TTM lessons during the school year, on average, students attempted about 4.1 lessons in their first month of TTM use and 3.5 lessons in their last month of TTM. Across elementary grades, passing rates were above 50% in the first month that students used TTM, and passing rates ranged from 42% to 47% in the first month of use in middle grades. By students' last month of use, passing rates had fallen to 20%-25% across all grade levels. Again, passing rates for students' last month of use should be interpreted with caution given that the number of lessons attempted in those months was very low compared to the rest of the year. As noted earlier, this decline in passing rates for lessons attempted later in the school year may be a function of students encountering more difficult mathematics content at that time.

Figure 3.7. Average Student Passing Rate for the First and Last Month a Student Used the System, by Grade Level, 2013-14



Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Notes: Only students with at least two months of TTM usage in 2013-14 were included in the calculation. July was excluded because there were no TTM lessons records for that month.

Given the range in the number of lessons initiated and/or completed by students, examining differences in passing rates over time—alone—provides an imperfect representation of students' exposure to, and progress through, the system. Students who took different numbers of lessons may have the same passing rates (e.g., a student who passed one out of one lessons would have the same passing rate as a student who passed 20 out of 20 lessons). Also related to the customized nature of TTM, the system does not administer calibrated tests that assess all students on the same grade-level content. Unlike systems where the same assessments are administered periodically to all students in a particular grade (e.g., benchmarks or interim assessments), TTM lessons and assessments are tailored to individual students' unique level of content knowledge. The assessments are testing students on the content to which they were exposed in the lesson they attempted, and not necessarily the students' grade-level mathematics learning objectives. While there are pedagogical advantages to this differentiated approach, one downside is the inability to produce aggregate snapshots that gauge students' mastery of the grade-level content throughout the year.

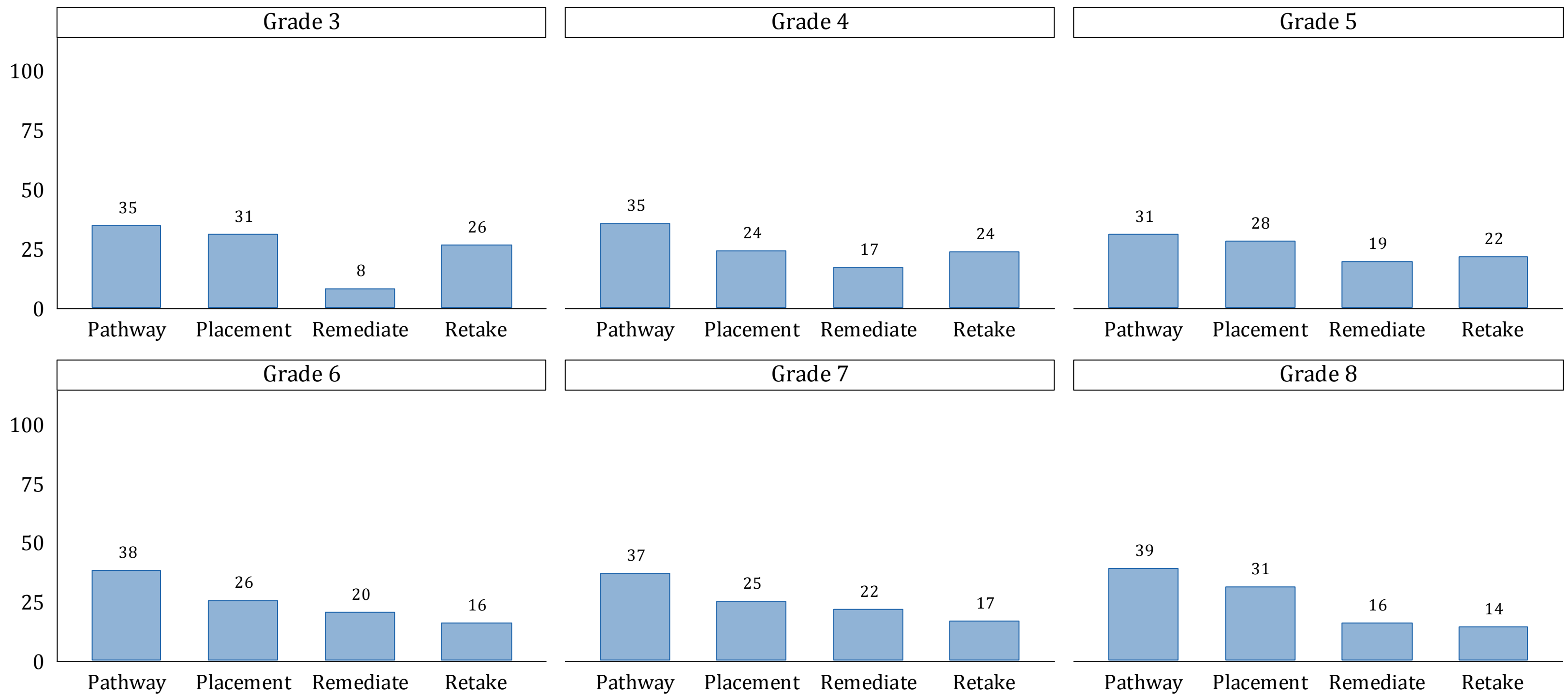
To accommodate this issue, the evaluation team developed other measures that tracked students' progress in learning TTM mathematics content based on their own unique progression through the system. All students entered a TTM path via a lesson starting point designated by the system or a teacher. Students who did not fail or were not remediated for lessons remained on that pathway and progressed to the next lesson determined by the TTM software or pre-determined by a teacher who placed a student on a customized pathway. When a student failed a lesson or was remediated, he or she was taken off the pathway to either one or more lessons. Given these pathways, student progression was assessed via the type of lesson students took. Lesson type was operationalized in four different ways:

1. *pathway lessons*, or lessons taken by students who never got off their pathway;
2. *placement lessons*, or lessons assigned to students by their teachers, usually a starting point on a student pathway;
3. *remedial lessons*, or lessons in which foundational knowledge was inserted to assist students in passing lessons on their current pathway;
4. *retake lessons*, or entire lessons in a student's pathway that have to be repeated.

Overall, approximately 66% of lessons taken in 2013-14 were pathway or placement lessons—with 35% of lessons taken on the TTM system's prescribed pathway and 31% of lessons completed as prescribed by teachers. About 35% of lessons were either remedial lessons (16%) or lessons retaken (19%).

As Figure 3.8 illustrates, the percentages of on path lessons, either via TTM pathways or placement pathways, were fairly consistent across elementary and middle grades, as were the percentage of remedial lessons taken (with the exception of Grade 3). The largest difference was found among retaken lessons. On average, 24% of lessons at the elementary level were retaken versus 16% of lessons taken by middle grade students.

Figure 3.8. TTM Lesson Types Attempted, by Grade Level, 2013-14



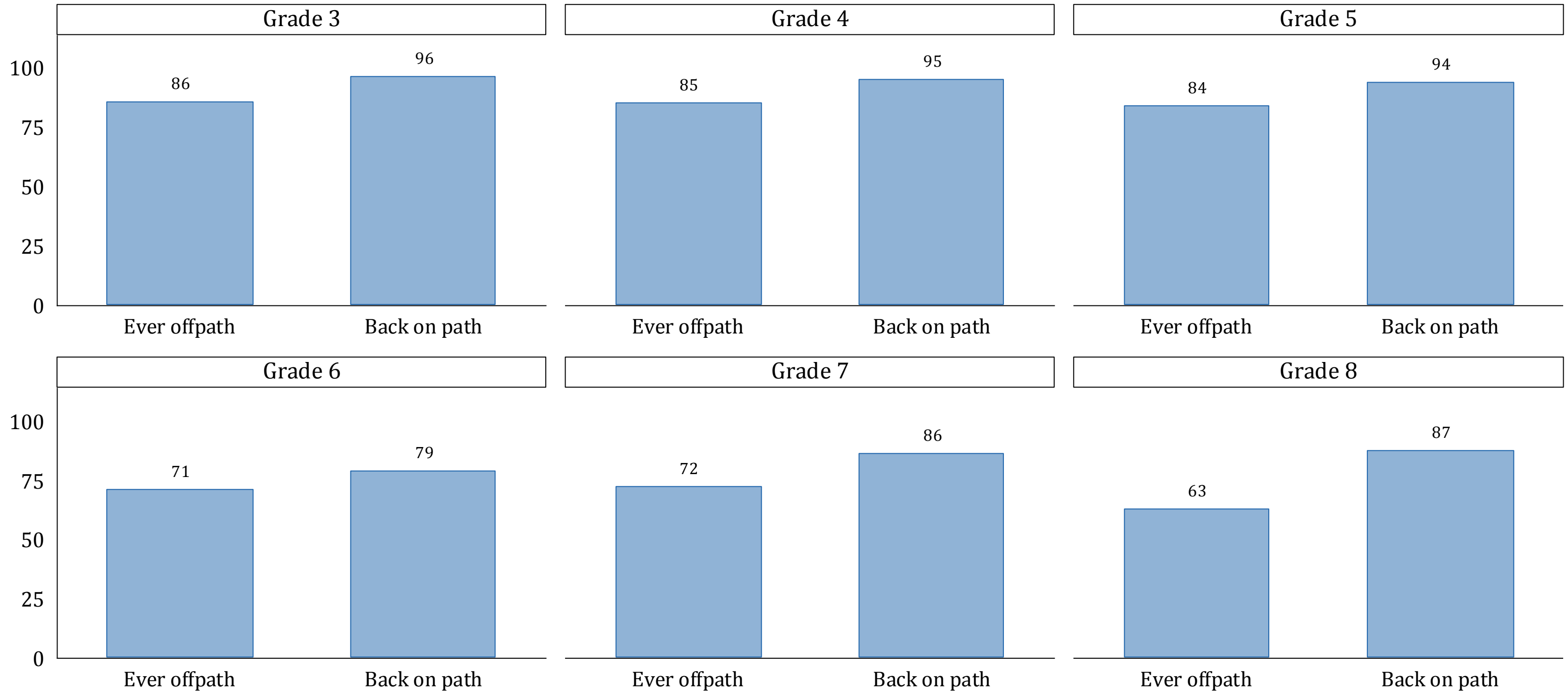
Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Next, considering the various lesson pathways, the evaluation team focused on whether students were able to get back on track after not passing TTM lessons (classified as *Ever Off Path*). *Ever Off Path* refers to students who at some point, after their initial pathway or placement TTM lesson, fail, retake, or are remediated for a lesson. *Back On Path* refers to students who got off path but who eventually got back to pathway lessons (i.e., persisted beyond retake and remedial lessons).

Figure 3.9 shows the percentage of students who at any time, after their initial pathway or placement TTM lesson, failed, retook, or were remediated for a lesson and of those students, the percentage that were able to get *Back On Path* by subsequently passing the lesson and progressing through the TTM program.

Figure 3.9 illustrates that, across all grades, most students got off path at some point during their work on TTM, but the vast majority were able to recover and get back on track with their TTM mathematics lessons. This was especially true in elementary grades, where roughly 95% of off-track students were able to get back on track. In middle grades, on average, 84% of students were able to get back on track. This finding suggests that the TTM system is fairly adept at accommodating students who got off path in their TTM lesson to support continued use of the system.

Figure 3.9. Percentage of Users Who Got Off TTM Lesson Pathway and Returned to Path, by Grade Level, 2013-14



Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: *Ever Off Path* refers to students who at some point, after their initial pathway or placement TTM lesson, fail, retake, or are remediated for a lesson. *Back On Path* refers to the percent of the subset of students who got off path but who eventually got back to pathway lessons (i.e., persisted beyond retake and remediation lessons).

TTM Usage and Student Progression in the TTM System

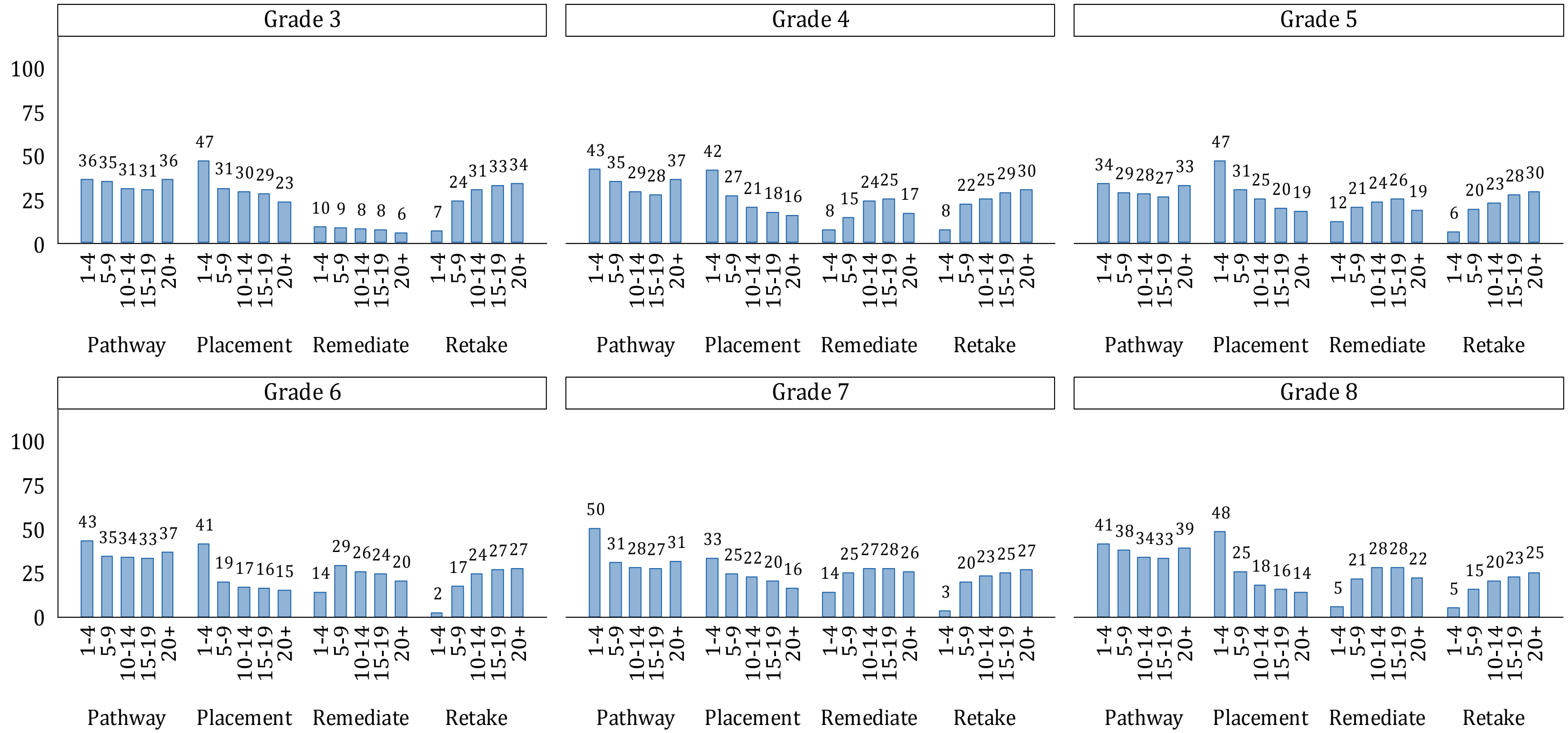
In this section, student progression through TTM lessons by passing rate and usage is examined. TTM system usage was categorized into five groups: 1-4 lessons, 5-9 lessons, 10-14 lessons, 15-19 lessons, and 20 or more lessons. These categories represent different levels of usage consistency, particularly use corresponding to levels recommended by TTM. Analyses were further disaggregated by grade level and lesson type (see Figure 3.10) and by whether students who fell off track, were able to get back on track (see Figure 3.11).

Figure 3.10 illustrates passing rates among students by grade level, lesson type, and students' TTM system usage. In terms of lesson type, non-linear relationships emerged between passing rates and TTM usage among students taking pathway or remedial lessons.[65] Across all grades, students with either the lowest or highest number of pathway attempted had the highest percentage of pathway lessons passed. For example, in Grade 3, students who registered 1-4 pathway lessons attempted and students who attempted 20 or more pathway lessons passed 36% of lessons (with students attempting 5- 9, 10-14, and 15-19 lessons attempted passing between 31% and 35%).

The opposite pattern was mostly true among students taking remedial lessons, with students who registered the fewest and the most TTM remedial lessons passing the lowest percentage of lessons. Conversely, linear relationships emerged between passing rates and TTM usage among students taking placement or retake lessons.

[65] A nonlinear relationship exists when change in one variable (grade level) does not correspond with constant change in the other variable (number of lessons attempted.

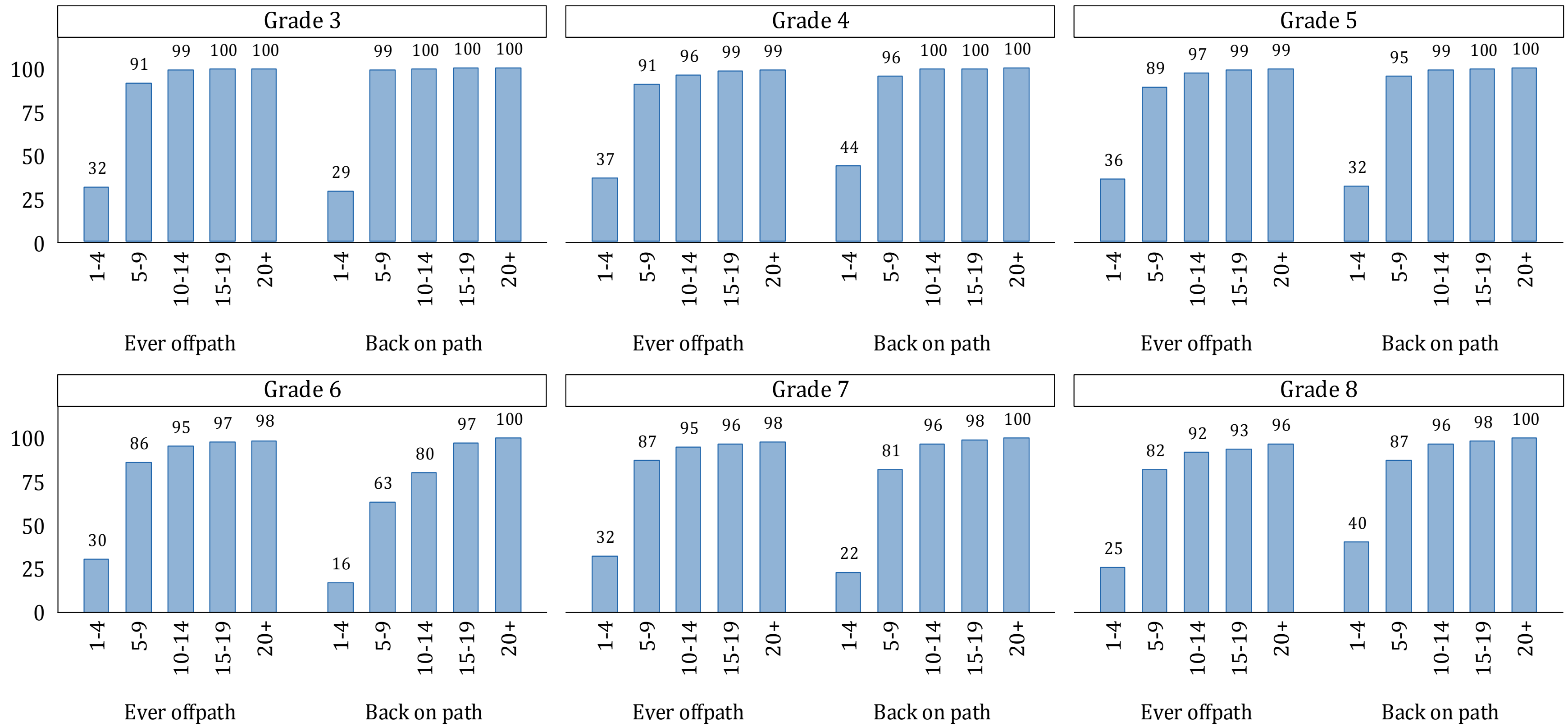
Figure 3.10. TTM Lesson Type by Number of Attempts, by Grade Level, 2013-14



Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Figure 3.11 illustrates that most students were able to get back on track after falling off track on the TTM system. In all grades, the vast majority of students with 5 or more lessons attempted were able to get back on track after failing a lesson; however, this is not the case for students attempting just 1-4 lessons (where 16% to 44% of students were able to get back on track, depending upon grade level).

Figure 3.11. Percentage of Users Who Got Off TTM Lesson Pathway and Returned to Path, by Grade Level and Number of Lessons Attempted, 2013-14



Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Guiding Research Question 5b: Summary of Findings

To examine relationships between TTM usage levels and progress within the TTM system, the evaluation team first assessed the TTM lesson passing rates, overall and by different student groups. Of all TTM lessons attempted in 2013-14, roughly 43% were passed, with similar passing rates across elementary and middle grades. Average passing rates across all grades were lowest among students with the fewest number of TTM lessons attempted (1-4 lessons) and highest among students attempting 20 or more lessons.

Analyses also considered passing rates by student groups. Across all grades, TTM lesson passing rates were lower among students classified as ELL and economically disadvantaged. There were also gaps in passing rates by race/ethnicity, with Asian students having the highest passing rates and African American and Hispanic students having the lowest. Across all student groups, with the exception of Grade 7, passing rates were highest in August. In subsequent months, passing rates either declined or plateaued, suggesting that students may have encountered more difficult mathematics content as the year progressed.

Next, students' progress in using TTM was examined as a function of whether students remained on prescribed TTM lesson pathways or had to repeat lessons or take remedial lessons. Of those students who had to take at least one remedial lesson—or retake a lesson—the majority were able to get back on path (roughly 95% of elementary and 85% of middle school students). This is an important finding because it suggests that the TTM program is creating an online environment where students can retake lessons and get back on track, continuing to benefit from the system, as opposed to becoming frustrated and disengaging from usage.

Last, to get a sense of whether varying levels of TTM implementation were associated with students' progress in TTM, TTM usage levels were examined for associated passing rates and pathways of students' progression through the system. A curvilinear relationship emerged between varying levels of TTM usage and progress through the system among students taking pathway or remedial lessons. The relationship for pathway and remedial is not linear but rather it is highest for the highest and lowest categories of pathway (or lowest for the highest and lowest categories of remedial). Generally, among students who fell off track, passing rates were substantially higher among those who initiated 5 or more lessons in the system. Low levels of TTM use (i.e., between 1-4 lessons) were associated with much lower TTM lesson passing rates among students classified as *Ever Off Path*, across all grades.

Guiding Question 6b:

What is the relationship between broader implementation of Texas SUCCESS and other SSI interventions and student outcomes for mathematics?

Key Findings:

- **Low TTM usage campuses used fewer other mathematics interventions than campuses classified as high TTM users.** Three out of four low TTM users reported using only TTM or TTM along with one other program (versus 58% of high TTM usage campuses).
- **Staff from low TTM usage campuses were more likely to cite barriers to TTM usage.** A larger proportion of staff at low TTM usage campuses reported barriers to TTM implementation, such as having insufficient support staff and less access to computers.
- **Inconsistent patterns emerged between school-level mathematics achievement and schools' combined use of mathematics programs and interventions.** Schools with higher average mathematics achievement used a greater number of mathematics programs than schools with lower average achievement, though schools using fewer mathematics programs had descriptively better gains between the two study years in mathematics achievement. These findings are descriptive only; they do not imply statistically significant relationships or causality.

While most of this section of the report has focused on relationships between mathematics outcomes for TTM users, this section includes exploratory analyses of other contextual factors that may have influenced whether students met SSI grade promotion requirements in mathematics in 2013-14. That is, this section explores relationships between mathematics achievement and contextual and instructional factors in schools by considering TTM use along with other mathematics interventions and programs used in schools. As described previously, TTM was not the only mathematics program used on most campuses in 2013-14. In fact, less than 20% of school-level interviewees reported that TTM was the only mathematics program used at their schools. In addition to TTM, school staff most commonly reported using either one or two other mathematics programs, with an average of two other mathematics programs used.

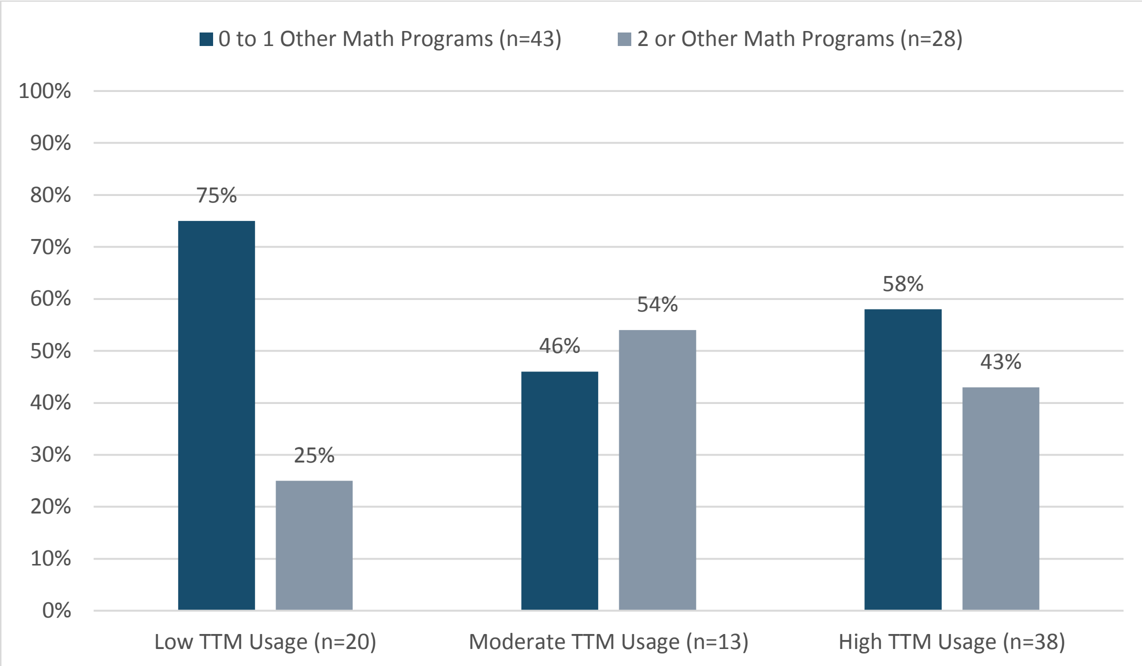
Given that over 80% of school-level interviewees reported using TTM and at least one other mathematics program, analyses described in this section sought to provide perspective on relationships between the use of different resources and mathematics achievement. Data elements used in the analyses included information on TTM usage, data on other programs used to support mathematics instruction in 2013-14—collected via interviews with school staff—and school-level STAAR-Mathematics outcomes. Before describing the findings, it is important to note several limitations to these data. For one, while fine-grained data on TTM usage patterns were available, data on schools' use of other programs were collected via interviewee responses regarding questions about their schools' use of "other mathematics interventions or other mathematics online learning programs." In other words, other mathematics programs were cited as either used or not used, while information on the frequency, timing, and duration of use, or use with particular grades or types of students, was not reported systematically. As such, relationships between use of other mathematics programs and mathematics outcomes could not be disaggregated by the nature of how those programs were used (e.g., frequency, intensity, etc.), nor could relationships be explored between other mathematics program use and mathematics achievement among students in specific grades or groups. In addition, information related to other mathematics program use was only available from schools selected for interviews (n=83), unlike TTM data, which were collected from all users of that system in public schools in the state. Given these features of the data, findings reported in this section provide a purely descriptive and exploratory view of school-level interventions geared toward improving student mathematics outcomes. These descriptive analyses were guided by two primary questions:

1) What was the nature of cumulative, school-level efforts to help students meet SSI grade promotion requirements, operationalized as the combination of TTM usage and the use of other mathematics programs?
2) Were there relationships between these cumulative efforts and school-level mathematics outcomes?

Descriptive Analysis of Programs Implemented for SSI Requirements

To explore the nature of combined TTM and other mathematics program use, the distribution of school-level TTM usage patterns was compared to the number of other mathematics programs used. Analyses focused on whether relationships between TTM and other mathematics program use were positive or negative—that is, whether high TTM using campuses used more or fewer other mathematics programs compared to low TTM usage campuses. Figure 3.12 provides results of this comparison, contrasting TTM use levels (i.e., low, moderate, or high usage) with the percentage of schools that were either above or below the mean number of programs used (i.e., fewer or more than two other programs). As the figure shows, descriptive levels of program use show a mixed pattern of TTM and other mathematics program usage. Three quarters of low TTM users used zero or one other program, meaning that only 25% of low TTM users also used two or more other mathematics programs. These trends imply that low TTM users were also lower users of other mathematics interventions. Conversely, the breakdown of other mathematics program use among moderate and high TTM users was split more evenly across the number of other mathematics programs used. Specifically, 46% of moderate TTM users cited use of zero or one other mathematics program in their schools, while the remaining 54% reported at least two other mathematics programs. Among high TTM users, 58% of school-level staff interviewed for this report cited use of zero or one other mathematics program, while the remaining 42% reported at least two other mathematics programs. Generally, low TTM users were three times more likely to report low use of other mathematics programs than use of two or more other mathematics programs. Conversely, the percentage of moderate and high TTM users were split more evenly across usage categories for other mathematics programs.

Figure 3.12. Usage Patterns between TTM and Other Mathematics Programs (N=71 Schools)



Source: Spring 2014 Interviews with campus-level staff.

In practical terms, these findings imply that a larger percentage of high and moderate TTM users utilized other mathematics programs relative to low TTM users. It is difficult to draw upon available data to determine possible explanations for these patterns. Considering differences between low, moderate, and high TTM users presented in previous sections, a larger percentage of low TTM users reported that they did not have sufficient support staff helping with TTM implementation compared to high or moderate TTM using campuses (65% versus 17% and 26%, respectively). A lack of computers was also cited as a barrier to TTM implementation at low use schools more often than at high usage schools, with less than half of low users reporting that they had sufficient computers to implement TTM, versus 73% of high users. That said, moderate TTM users also cited lack of access to computers as a challenge to TTM implementation, with almost the same frequency as low TTM users (48% versus 44%, respectively). This trend challenges possible assumptions about the role computer access may have played in schools' ability to implement mathematics interventions because, while low and moderate TTM users reported similar computer access challenges, they did not report similar use of other (online or face-to-face) mathematics programs. In sum, while different patterns emerged in terms of schools' TTM and other mathematics program use, clear and plausible explanations for those differences did not emerge in the interview or TTM usage data.

Relationships between Programs Implemented to Meet SSI Requirements and School-Level Mathematics Outcomes

Building on the exploration of mathematics interventions used to promote SSI grade promotion requirements, descriptive patterns in mathematics achievement—from both 2012-13 and 2013-14—were explored in schools by their TTM and other mathematics program usage (see Table 3.22).[66] In general, mathematics achievement was lower in schools reporting use of one or no other mathematics program compared to schools reporting use of at least two other mathematics programs, in both 2012-13 and 2013-14. That said, school-level mathematics achievement was higher in 2013-14 than in the previous year in schools reporting zero or one other mathematics program, while, on average, schools using TTM and at least two other mathematics programs had lower 2013-14 achievement than they had in the previous year.

While achievement was lower in schools using fewer mathematics programs, two-year differences in mathematics achievement were higher in schools using fewer mathematics programs. It is important to note that these comparisons involve different students each year—average school-level mathematics achievement was calculated from different cohorts of elementary and middle school students in 2012-13 and 2013-14. Conducting these types of cross-sectional analyses does not allow for conclusions to be drawn about whether school-level achievement differences from one year to the next may have been expected given the performance of previous student cohorts. These findings are provided simply to allow for the consideration of possible trends in schools' use of mathematics interventions and achievement. To those ends, it may be instructive to consider two trends: (1) schools with higher average mathematics achievement—(i.e., higher school-level STAAR mathematics results)—used more mathematics programs than schools with lower average achievement; and (2) schools using fewer mathematics programs had better 2013-14 achievement than in the previous year, while the opposite was mostly true for schools using TTM and other mathematics programs with greater frequency. Further exploration would be needed to uncover possible reasons for this trend.

Table 3.22. Average School-Level STAAR Mathematics Scale Scores, by TTM and Other Mathematics Program Use

Zero to One Other Program (N=39)			Two or More Other Programs (N=26)		
Campus System Usage Level	**2012-13**	**2013-14**	**Campus System Usage Level**	**2012-13**	**2013-14**
Low TTM Usage (n=14)	1519	1527	Low TTM Usage (n=5)	1623	1621
Moderate TTM Usage (n=6)	1477	1493	Moderate TTM Usage (n=7)	1507	1512
High TTM Usage (n=19)	1579	1583	High TTM Usage (n=14)	1647	1636

Source: State of Texas Assessments of Academic Readiness data, 2012-13 and 2013-14, Texas Education Agency. Spring 2014 Interviews with campus-level staff.

Note: Ns represent the number of schools in each category.

[66] While this report focuses on relationships between TTM use and mathematics outcomes in 2013-14, analyses throughout have considered prior achievement when examining that relationship. As such, 2012-13 achievement is included here to show, descriptively, average prior mathematics achievement in schools included in the analysis.

Guiding Research Question 6b: Summary of Findings

TTM was not the only mathematics program used on most campuses in 2013-14. Less than 20% of school-level interviewees reported that TTM was the only mathematics program used at their schools. Mathematics intervention staff at three quarters of low TTM usage campuses indicated that they used TTM only or just one other mathematics program, meaning that only a quarter of low TTM users also used two or more other mathematics programs. These trends imply that low TTM users were also lower users of other mathematics interventions. Conversely, the breakdown of other mathematics program use among moderate and high TTM usage campuses was split more evenly across the number of other mathematics programs used. Among high TTM users, 58% of school-level staff interviewed for this report cited use of zero or one other mathematics program, while the remaining 42% reported at least two other mathematics programs.

A larger percentage of staff at low TTM usage campuses reported that they did not have sufficient support staff helping with TTM implementation, compared to high or moderate TTM usage campuses. A lack of computers was also cited as a barrier to TTM implementation at low TTM usage schools more often than at high usage schools, with less than half of staff from low usage schools reporting that they had sufficient computers to implement TTM, versus 73% of high users.

In general, mathematics achievement was lower in schools reporting use of one or no other mathematics program relative to schools reporting use of at least two other mathematics programs, in both 2012-13 and 2013-14. It is important to note that conducting these types of cross-sectional analyses does not allow for conclusions to be drawn about whether school-level achievement differences from one year to the next may have been expected given the performance of student cohorts in the prior year.

Research questions (7-10) which follow explore the relationship between TTM usage and STAAR-Mathematics related outcomes.

Guiding Question 7b:

To what extent do student performance results differ for students participating in the Texas SUCCESS program (i.e., using TTM) and non-participating students, and how do results differ by student characteristics, prior academic performance, and other key variables?

Key Descriptive Findings

- **In Grades 3 and 4, students attempting more TTM lessons experienced higher passing rates on the 2013-14 STAAR-Mathematics assessment.** For students in Grade 3, 64% with 1-4 TTM lessons attempted passed the STAAR-Mathematics test, compared to 72% of students who attempted 10 or more TTM lessons. Comparable results were observed for students in Grade 4 (66% versus 72%).
- **In Grades 5, 7, and 8, a negative relationship between the number of TTM lessons attempted and STAAR-Mathematics passing rates was observed in 2013-14.** For students in Grades 5, 7, and 8, students who attempted 1-4 lessons passed the STAAR-Mathematics assessment at higher rates (i.e., 3-8 percentage points, depending upon grade) compared to students attempting 10 or more TTM lessons during the 2013-14 school year. In Grade 6, STARR-Mathematics passing rates were flat across TTM usage levels.

Key Multivariate Findings

- **TTM users, particularly those who attempted higher numbers of TTM lessons during the 2013-14 school year, had statistically significant higher gain scores on the STAAR-Mathematics assessment (i.e., decile-standardized gain scores).** For instance, Grade 6 students who used the system at least once had positive gain scores compared to non-users; however, students attempting 20 or more TTM lessons had positive effects sized over three times as large.
- **In lower grades (Grades 4 and 5) only high levels of TTM use were associated with significant and sizable gains, while gains were detected for other grades at lower thresholds or categories of TTM usage (as low as 5-9 attempts).**

In this section, the evaluation team explored the degree to which students' mathematics performance varied across TTM program participation status, given their prior performance and other student- and school-level factors. To do so, student-level longitudinal state administrative and STAAR data from school years 2011-12 to 2013-14 were linked to student-level TTM system usage in 2012-13 and 2013-14.

Relationships between TTM Assessment Performance, TTM Usage, and STAAR-Mathematics Performance

Because it is feasible that non-TTM users were systematically different from TTM users, key student characteristics were taken into account in both the descriptive and inferential analyses of differences in student outcomes between TTM participants and non-participants. The evaluation team linked a number of metrics from the TTM system usage data— including student TTM lesson usage and passing rates—to TEA administrative and assessment files. This allowed the evaluation team to assess the relationship between students' TTM lessons attempted and passed with their subsequent performance on STAAR-Mathematics in spring 2014. These descriptive analyses provide information on associations between TTM usage and mathematics performance in ways that inform findings in statistical models described later in this section.

Figure 3.13 depicts the percentage of students, at each grade level, who met the phase-in 1 Level II (Satisfactory) standard on the first administration of STAAR-Mathematics assessment in the spring of the 2013-14 school year, disaggregated by the number of lessons attempted in 2013-14.[67] In terms of the number of lessons attempted, it is important to recall that key cut-points for user engagement with TTM and recommended usage levels (as defined by the vendor) are 5 or more and 10 or more attempted lessons, respectively. Students with zero lessons attempted are displayed for comparison; these are students who never participated in TTM.

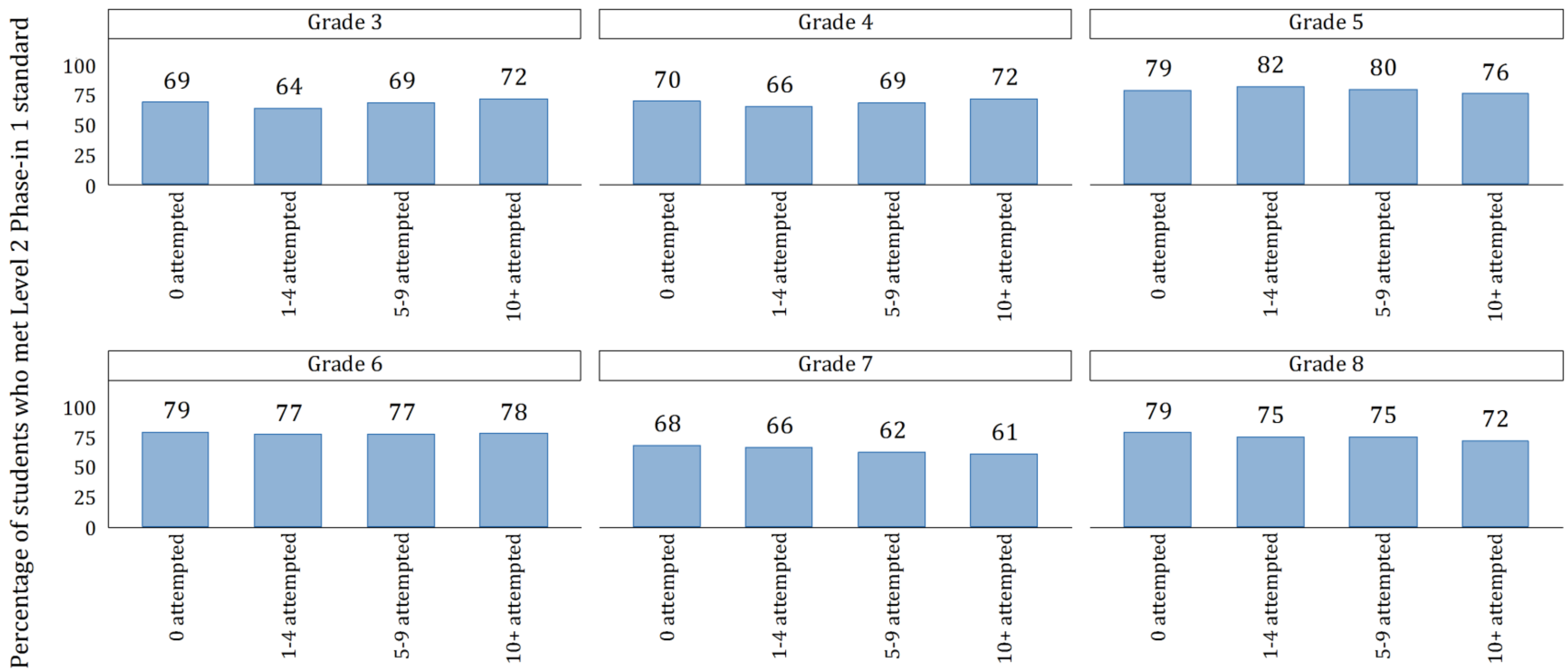
As the figure shows, the proportion of students who attempted zero lessons (non-TTM students) met the STAAR-Mathematics standard with roughly the same frequency as the proportion of students who attempted 10 or more lessons. In Grades 7 and 8, a larger proportion of students with zero lessons attempted met the standard than students with 10 or more lessons (differences of 8 and 6 percentage points for Grades 7 and 8, respectively). One takeaway from these results may be that, in Grades 7 and 8, students who never used TTM met STAAR-Mathematics performance standards with greater frequency than their peers who attempted 10 or more TTM lessons. Another takeaway could be that, in lower grades, students who attempted 10 or more lessons performed roughly the same or slightly better than students who did not attempt a lesson. For students in Grades 5, 7, and 8, students who attempted 1-4 lessons passed the STAAR-Mathematics assessment at higher rates (i.e., 3-8 percentage points, depending upon grade) compared to students attempting 10 or more TTM lessons during the 2013-14 school year. In Grade 6, STARR-Mathematics passing rates were flat across TTM usage levels.

Although these descriptive results are an important first step in understanding the relationship between TTM usage and student STAAR-Mathematics performance, they do not adjust for other observable student, school, or district-level characteristics that may have contributed to students' test performance. The multivariate models that follow seek to unpack these patterns in ways that explore whether students who had higher levels of TTM use (e.g., higher lessons attempted) performed better on STAAR-

[67] Henceforth, the phase-in 1 Level II (Satisfactory) standard will be called "STAAR passing standard" for brevity.

Mathematics exams, or whether those students with lower prior performance on standardized testing were targeted for higher levels of TTM usage.

Figure 3.13. Percentage of Students Meeting the Phase-In 1 Level II (Satisfactory) Standard in Mathematics, by Grade Level and TTM Lessons Attempted (2013-14)



Source: Think Through Math assessment and session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level included the first administration only, and regular English and Spanish versions only (i.e., not modified or alternate versions).

Relationship between TTM Usage and Changes in Student Performance on STAAR-Mathematics

The evaluation team calculated standardized STAAR-Mathematics test-score gains between the first test administration in 2012-13 and 2013-14. This is the measure of student performance that will be used in the multivariate outcomes portion of this section because it classifies how students' STAAR-Mathematics scores changed between 2012-13 and 2013-14.

Table 3.23 provides the number of students, by grade level, for which both grade and year decile-standardized gain scores could be calculated. Gain scores could not be calculated for Grade 3 students because STAAR-Mathematics tests were only administered in Grades 3-8 and no comparable prior-year measure of student performance was available. Consequently, any measure comparing 2012-13 to 2013-14 could not be calculated for Grade 3.

Table 3.23. Counts of Students with Valid STAAR-Mathematics Decile-Standardized Gain Scores between 2012-13 and 2013-14, by Grade Level

Student Grade Level	Count of Students without Valid Gain Scores	Count of Students with Valid Gain Scores	Total
Grade 4	69,042	314,346	383,388
Grade 5	69,881	312,777	382,658
Grade 6	76,109	300,178	376,287
Grade 7	98,724	286,170	384,894
Grade 8	146,776	232,876	379,652
Total	**460,532**	**1,446,347**	**1,906,879**

Source: State of Texas Assessments of Academic Readiness data, 2012-13 and 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level included the first administration only, and regular English and Spanish versions only (i.e., not modified or alternate versions). Only students with consecutive grade level assessments (e.g., three in 2012-13 and four in 2013-14) were included in the calculations.

Multivariate Analysis of the Relationship between TTM Usage, Usage Intensity, and Student Performance

The outcome analysis for this research question focused on students' performance on 2013-14 STAAR-Mathematics. The intensity of usage (or dosage) was defined as the number of TTM lessons attempted (i.e., 0 attempts, 1-4 attempts, 5-9 attempts, 10-14 attempts, 15-19 attempts, and 20 or more attempts).

The evaluation team utilized a number of different statistical models to explore the research question, all of which are not presented in the body of the report, but are provided in Appendix B.[68] The approach was to build sequentially more complex statistical models that carried different assumptions about the processes by which students chose, or were selected, to participate in the program. This is important, because the ways in which students were identified for TTM use were not random across districts and

[68] All of the models are fit using ordinary least squares (OLS) regression, because the outcome measure across research questions is continuous.

schools, nor were they measurable through administrative and testing files that were available in state education databases. The fundamental research problem was that it was not possible to observe the same student under two conditions: participating in the program and not participating in the program (Gelman & Hill, 2007; Rubin, 1974). To overcome this missing data problem, the evaluation team used descriptive analyses, qualitative interview data, and complex statistical models to minimize the observable and unobservable differences between students who participated and those who did not.[69]

In addition, it was not possible to measure, with the same degree of precision, other supports, interventions, and programs in which students participated in a given school year or across time. Nor did the evaluation team have complete information on student participation in the program in 2012-13. These are important limitations, and the implications of these limitations are discussed in greater detail in Appendix E.

Table 3.24 presents the difference-in-difference (DiD) results, where the outcome was prior-achievement decile-grade-year standardized gain scores. The coefficients reported in the table represent estimated differences in standardized gain scores between participating and non-participating students.[70]

TTM users, particularly those who used the system more than higher usage thresholds, had statistically significant higher gain (e.g., decile-standardized gain scores). For example, Grade 6 students who used the system at least once had gain scores that were .067 standard deviations higher than non-users and, for those attempting 20 or more lessons, the gains were .215 standard deviations higher than non-users. Associations between TTM use and mathematics outcomes among low users were statistically significant in Grades 4, 7 and 8, with students who used TTM 1-4 times having weaker gains (e.g., lower decile-standardized gain scores) than non-users.

[69] Even these methods are no panacea. For instance, a crucial element missing from estimates of participation in TTM and STAAR-Mathematics performance was teacher quality. This will be discussed in more detail in the *Limitations* discussion that accompanies this section.

[70] Thus, gain score differences between participating and non-participating students were compared to students in the same grade and from the same prior-achievement decile in 2012-13. A standardized gain score model with one-period lagged STAAR-Mathematics score as a covariate was fit as well, and the results were strongly correlated with the results from Model 2. However, including a baseline measure in a gain score model is discouraged, because the residuals are correlated with the baseline term.

One takeaway here is that in lower grades (Grades 4 and 5), only high levels of TTM use were associated with significant and sizable gains, while gains were detected for other grades at lower thresholds or categories of TTM usage (as low as 5-9 attempts).[71]

Table 3.24. Estimated Effects of TTM Usage on STAAR-Mathematics Test Score Gains between 2012-13 and 2013-14, by Student Grade Level and Measure of Program Participation

Grade Level	System Usage Measure	B	SE	N
Grade 4	Used at least once	**0.029****	0.009	312,869
	1-4 attempts	***-0.047*****	0.010	
	5-9 attempts	-0.010	0.010	
	10-14 attempts	**0.025***	0.011	
	15-19 attempts	**0.044*****	0.012	
	20+ attempts	**0.144*****	0.011	
Grade 5	Used at least once	**0.026****	0.009	311,097
	1-4 attempts	-0.011	0.010	
	5-9 attempts	0.008	0.010	
	10-14 attempts	0.013	0.011	
	15-19 attempts	0.017	0.011	
	20+ attempts	**0.093*****	0.010	
Grade 6	Used at least once	**0.067*****	0.009	298,399
	1-4 attempts	-0.014	0.009	
	5-9 attempts	**0.058*****	**0.011**	
	10-14 attempts	**0.086*****	**0.013**	
	15-19 attempts	**0.109*****	**0.014**	
	20+ attempts	**0.215*****	**0.013**	
Grade 7	Used at least once	**0.083*****	**0.010**	284,479
	1-4 attempts	0.017	0.011	
	5-9 attempts	**0.044*****	**0.012**	
	10-14 attempts	**0.091*****	**0.014**	
	15-19 attempts	**0.121*****	**0.016**	

[71] Appendix B includes supplementary analyses exploring the extent to which the number of TTM lessons passed—relative to the number of lessons attempted—was associated with mathematics gains. Among all TTM users—both students who used the system above and below the five-lesson threshold, standardized gain scores increased as passing rates increased. This implies that students' success in passing TTM lessons was positively associated with students' performance on STAAR-Mathematics tests.

	20+ attempts	**0.220*** **	**0.014**	
Grade 8	Used at least once	**0.041*** **	**0.012**	231,674
	1-4 attempts	***-0.027****	**0.013**	
	5-9 attempts	**0.052*** **	**0.015**	
	10-14 attempts	**0.087*** **	**0.018**	
	15-19 attempts	**0.122*** **	**0.019**	
	20+ attempts	**0.155*** **	**0.019**	
Lagged STAAR-Mathematics Score?		No		
Campus Fixed Effects?		Yes		
Campus Random Effects		No		
Propensity Score Weighting?		No		
Outcome Measure		Prior achievement decile-grade-year standardized gain score		

Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2012-13-2013-14, Texas Education Agency.

Note: Reference category for system usage was zero attempts. Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in analyses. Number of attempts were prior to the first STAAR administration, contingent on grade level. Estimates were derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. Statistically significant positive coefficients are denoted in bold, and negative coefficients are denoted by bold and italicized font. *p<0.05, **p<0.01, ***p<0.00. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Guiding Research Question 7b: Summary of Findings

Descriptive results in this section show that students who did not use TTM met the STAAR passing standard in mathematics with roughly the same frequency as students who attempted 10 or more TTM lessons. Descriptive analyses did not adjust for other student, school, or district-level characteristics that may contribute to students' test performance. Thus, analytic models were designed to address this limitation, controlling for both stable and time-varying student- and school-level factors when considering relationships between TTM use and mathematics gains.

Across different model specifications, definitions of program participation, and mathematics outcomes, multivariate analyses demonstrate that TTM users who used the system more frequently—particularly those who attempted 20 or more lessons—had STAAR-Mathematics scaled scores and gains that were significantly higher than non-users, from both a statistical and practical perspective. However, low TTM using students (those who attempted between 1 and 4 lessons) experienced smaller gains in 2013-14.

Guiding Question 8b:

What is the relationship between the timing of system usage (i.e., the proximity to the STAAR test) and TTM usage levels (i.e., dosage as measured by lessons attempted, completed, and passed) and performance on the first administration of the STAAR-Mathematics exam?

Key Findings:

- **Continuous usage of the TTM system throughout the school year may be the most effective strategy for all grade levels, but particularly for middle school students.** Across all grades, there was a significant, positive relationship between usage continuity and STAAR-Mathematics gains, with an additional month of five or more lessons attempted associated with positive gains (ranging from 0.042 to 0.076 standard deviations). Gains were largest among middle school students, with the highest gains in Grade 8 (0.076 standard deviations). Among those students, an additional three months of five or more uses was associated with standardized gains of almost a quarter of a standard deviation and six additional months with almost a half a standard deviation gain.
- **In all grades, there was a positive relationship between mathematics gains and the number of lessons attempted in the month before the STAAR test.** Each additional TTM lesson attempted was associated with small, but statistically significant gains on the STAAR-Mathematics assessment.
- **A larger concentration of TTM use in the months leading up to the STAAR assessment—as opposed to use that was spread out more evenly across the school year—was significantly associated with very small gains in mathematics**. Middle grade students who used TTM more intensely in the months before STAAR had slightly larger, but statistically significant, mathematics gains compared to those who did not.

In this section, the relationship between the timing and proximity (to the STAAR-Mathematics exam) of TTM usage throughout the school year and student mathematics gain scores is explored. In other words, did a student who used the system only in September 2013, or a student who used the system in each month of the school year, experience different mathematics gain scores? Other measures of usage, employed in previous analyses, were based on total time spent using the system, or whether the student met the usage threshold recommended by TTM. As such, those measures do not capture the proximity of usage to the STAAR test, nor do they reflect the continuity of usage throughout the school year. Student performance may be more sensitive to an intervention—such as TTM—that occurs more closely to when the outcome is measured.

Using monthly TTM usage data, three measures (i.e., continuity of usage, proximity of usage to the STAAR assessment, and intensity of usage) were developed to investigate whether usage continuity throughout the school year and usage proximity to the STAAR test administration were associated with STAAR-Mathematics gains in 2013-14.

First, to capture **usage continuity**, the number of months in which a student recorded five or more lessons was calculated for each grade and month. The five lessons attempted threshold was used since it is the minimum number of attempts that TTM suggests signal student engagement with the system and, also, is the minimum category by which performance gains were detected in earlier analyses. The monthly flags were then summed to create a count of the total number of months in which the students' usage met or exceeded five lesson attempts.

Second, to assess **usage proximity to the test**, the total number of lesson attempts in each month *prior to the STAAR-Mathematics* test was used as a covariate in multivariate models to investigate differences in the relationship between dosage proximity and changes in student performance between 2012-13 and 2013-14.[72]

Third, for each month, as a measure of **intensity of usage**, the proportion of each student's total usage prior to the first administration of the STAAR-Mathematics exam that occurred in each month was calculated, and the proportion of usage that occurred *three months prior to the test administration* (e.g., for students in Grades 5 and 8, this period was January-March 2014) was summed to calculate the share of a student's total TTM usage that occurred temporally near the test administration.[73] In other words, this approach explored the joint relationship between usage proximity and intensity, and accounted for the fact that the first usage proximity measure (i.e., number of lessons attempted in the month of STAAR test administration) disregards total system usage.

Table 3.25 displays the results from the statistical models.[74] The findings suggest that continuous usage of the TTM system throughout the school year may be the most effective strategy for all grade levels, but particularly for middle school students. Across all grades, there was a significant, positive relationship

---

[72] For Grade 5 and 8 students, minutes were summed from August 2013 through March 2014, and for Grade 4, 6, and 7 students, minutes were summed from August 2013 through April 2014.

[73] This was used as a variable in the statistical models, and the association between this measure and changes in STAAR-Mathematics performance was estimated between students who met the grade-recommended usage threshold during the 2013-14 school year, and students who did not.

[74] Models were fit separately for each grade level and for each of the three indicators of TTM usage, including usage continuity (i.e., number of months in which a student attempted at least five lessons) and two measures of usage proximity to the STAAR test: the number of lessons attempted in March/April 2014 and the proportion of total use that occurred in the three months prior to STAAR test. The outcome of interest in these models was decile standardized gain scores, which represent standardized gains (scaled to have a mean of 0 and a standard deviation of 1) within deciles of prior student achievement. The technical matter for these statistical models can be found in Appendix B.

between usage continuity and STAAR-Mathematics gains, with an additional month of five or more lessons attempted positively associated with mathematics score gains.

Gains were largest among middle school students, with the largest mathematics score gains in Grade 8 (0.076 standard deviations). Among those students, an additional three months of five or more uses was associated with gains of almost a quarter of a standard deviation, and six additional months was associated with almost half a standard deviation gain (Figure 3.14 represents this effect graphically, with Grade 6 outcomes). By comparison, six additional months of five or more TTM lessons attempted was associated with a gain of a quarter of a standard deviation among Grade 5 students.

System usage intensity in the month immediately prior to the STAAR administration was also significantly and positively associated with student performance across all grades; however, the magnitude of the relationship was notably smaller than that between usage continuity and mathematics gains. There also did not appear to be a large difference in the magnitude of the differences of usage in April and gains across grade levels, with 10 additional lessons attempted in the month before STAAR associated with gains of roughly a tenth of a standard deviation across all grades.

Last, the proportion of usage that occurred in the three months before the STAAR test was examined, with results demonstrating statistically significant across most grades. And, even among these grades, the magnitude of the effect was very small (.001 standard deviations). These results imply that of the three measures of usage intensity and continuity the smallest gains were associated with having a substantial proportion of a student’s total TTM use concentrated in the months leading up to STAAR. Taken altogether, these results demonstrate that continuous usage above the recommended threshold was associated with most substantial gains—particularly at the middle school level—while the number of lessons attempted in March/April was associated with smaller, though still significant, gains across all grades.

Table 3.25. Estimates of the Relationship between TTM Usage Continuity and Proximity to the STAAR Test and Decile-Standardized Gain Scores, by Grade Level, 2013-14

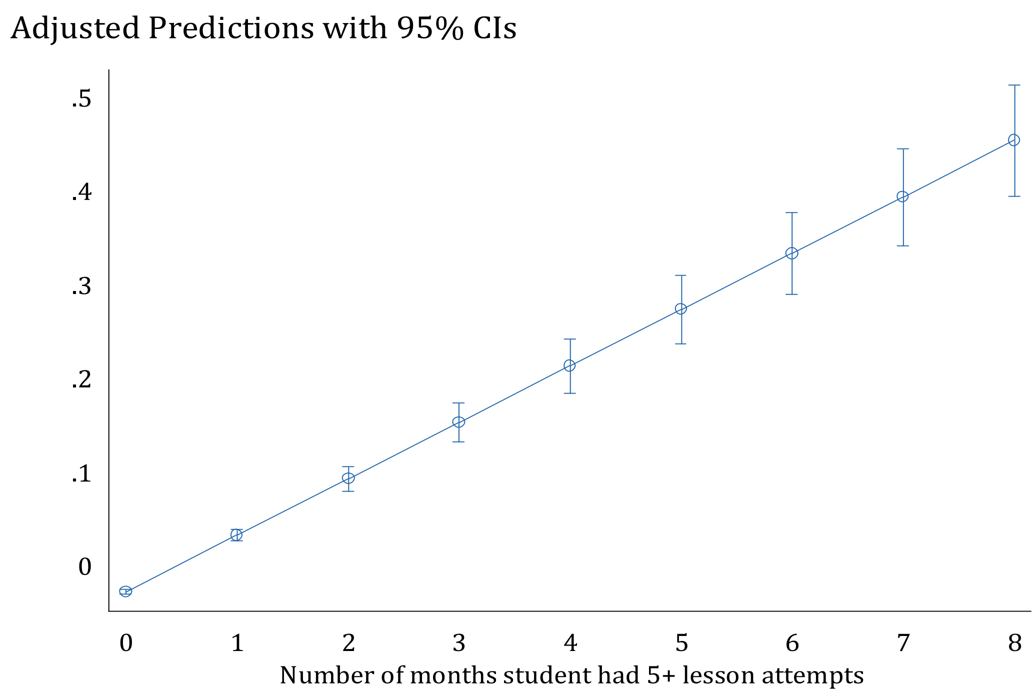
		B	SE	N
Grade 4	Number of months in which student recorded 5 or more lesson attempts	**0.049*****	0.003	314,308
	System usage attempts in April 2014	**0.010*****	0.001	
	Proportion of usage occurring in three months prior to STAAR test	0.000	0.000	
Grade 5	Number of months in which student recorded 5 or more lesson attempts	**0.042*****	0.003	312,746
	System usage attempts in April 2014	**0.008*****	0.001	
	Proportion of usage occurring in three months prior to STAAR test	**0.000***	0.000	
Grade 6	Number of months in which student recorded 5 or more lesson attempts	**0.060*****	0.004	300,145
	System usage attempts in April 2014	**0.012*****	0.001	
	Proportion of usage occurring in three months prior to STAAR test	**0.001*****	0.000	
Grade 7	Number of months in which student recorded 5 or more lesson attempts	**0.062*****	0.004	286,133
	System usage attempts in April 2014	**0.011*****	0.002	
	Proportion of usage occurring in three months prior to STAAR test	**0.001*****	0.000	
Grade 8	Number of months in which student recorded 5 or more lesson attempts	**0.076*****	0.006	232,855
	System usage attempts in April 2014	**0.013*****	0.002	
	Proportion of usage occurring in three months prior to STAAR test	**0.001*****	0.000	
Campus Fixed Effects?		Yes		
Campus Random Effects		No		
Propensity Score Weighting?		No		
Outcome Measure		Grade-year-decile standardized gain score		

Source: State of Texas Assessments of Academic Readiness and Public Education Information Management System data, 2011-12 – 2013-14, Texas Education Agency. Think Through Math session history table, 2013-14.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in the analyses. Estimates were derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariate. Statistically significant negative coefficients are denoted in bold and italic, and statistically significant positive coefficients are denoted in bold. * p<0.05; ** p<0.01; *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Figure 3.14 shows an example of the relationship between usage continuity across the school year and gain scores for Grade 6 students.[75] For each additional month in which a Grade 6 student used TTM for five or more lesson attempts, the decile standardized gain score rose .022 standard deviations. Thus, the estimated decile-standardized gain score for Grade 6 students who made at least five lesson attempts for seven months was .34, or over double the gains among students who used the system at this level of intensity for three months (.15 standard deviations).

Figure 3.14. Gain Scores by the Number of Months in which Grade 6 Students Attempted 5 or More TTM Lessons, 2013-14



Source: State of Texas Assessments of Academic Readiness and Public Education Information Management System data, 2011-12 – 2013-14, Texas Education Agency. Think Through Math session history table, 2013-14.

Note: Predicted average margins were calculated from results of linear regression results with all covariates held at their mean values.

Guiding Research Question 8b: Summary of Findings

To assess relationships between the consistency and timing (including proximity of usage to the STAAR assessment) of TTM usage throughout the school year and mathematics outcomes, three additional measures of TTM usage were created: 1) use *continuity* throughout the school year, 2) usage *proximity* to the STAAR test administration in April, and 3) *intensity of use* concentrated into the three months before the test was administered.

[75] Predictive margins were derived from the multivariate regression results.

Of the three measures, usage continuity had the strongest, statistically significant, positive associations with standardized gains on the STAAR-Mathematics assessment, particularly among middle grade students. These findings demonstrate that six months of TTM use above the five or more lessons threshold was associated with significant gains among students in Grades 5 and 8, and that better mathematics gains were associated with consistent system usage throughout the school year.

With regard to usage proximity to the STAAR test, positive associations emerged across all grades between mathematics gains and each additional lesson attempted in the month before STAAR. Statistically significant associations between usage concentrated in the three months before STAAR and standardized gains were found across grades, though the magnitude of these associations were quite small. The proportion of usage concentrated in the three months before STAAR was not significantly associated with mathematics gains in Grade 4.

Guiding Question 9b:

To what extent does performance differ between Texas SUCCESS participants (i.e., those using TTM) and non-participants among students at risk of being retained?

Key Findings:

- **Among Grade 5 students who had previously failed the STAAR-Mathematics assessment (in 2011-12 or 2012-13), lower thresholds of TTM use were associated with positive gains in mathematics.** Grade 5 students at risk of being retained in grade who attempted 5-9 lessons had gains that were higher (0.064 standard deviations) than their peers who were at risk of being retained and who did not use TTM.
- **Greater TTM use was significantly associated with weaker gains in mathematics among students in Grades 5 and 8 at risk of being retained in grade.** In Grade 5, students at risk of being retained in grade who attempted 20 or more lessons had weaker gains than students who had also previously failed the STAAR test, but did not use TTM. The same was true among at-risk students in Grade 8 who attempted between 15 and 19 lessons.

While prior analyses included all students in Grades 3-8, this research question drilled down further and only included students at risk of being retained in Grade 5 and Grade 8, the two SSI high stakes testing grades that require students to meet state standards on the STAAR assessment in order to be promoted to the next grade level. More specifically, the following analyses explore the relationship between TTM system usage for students at risk of being retained in Grades 5 and 8 and gains in mathematics scale scores.

As shown in descriptive analysis and earlier analytic models, student TTM usage, particularly at higher levels of intensity, was associated with larger increases in STAAR-Mathematics performance. This section explores the relationships between TTM usage—and usage intensity—and changes in student performance on STAAR-Mathematics in a different way: it focuses on the association between usage and STAAR-Mathematics assessment performance between students at risk of being retained and those not at risk of being retained.[76]

Table 3.26 displays the count of students in each grade level who were identified as being at risk of being retained between 2013-14 and 2014-15. In both grades, approximately one third of students in 2013-14 were categorized as being at risk of being retained in grade, and a slightly higher percentage of students in Grade 5 (41%) were at risk relative to students in Grade 8 (36%).

Table 3.26. Frequency Count and Percentage of Students at Risk of Being Retained in Grade between 2013-14 and 2014-15

Student Grade Level	Failed STAAR-Mathematics at Least Once in Prior Year	Count of Students	Percentage of Students
Grade 5	Never failed (Not At Risk)	193,254	59.26
Grade 5	Failed at least once (At Risk)	132,863	40.74
Grade 8	Never failed (Not At Risk)	183,329	64.20
Grade 8	Failed at least once (At Risk)	102,236	35.80

Source: State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

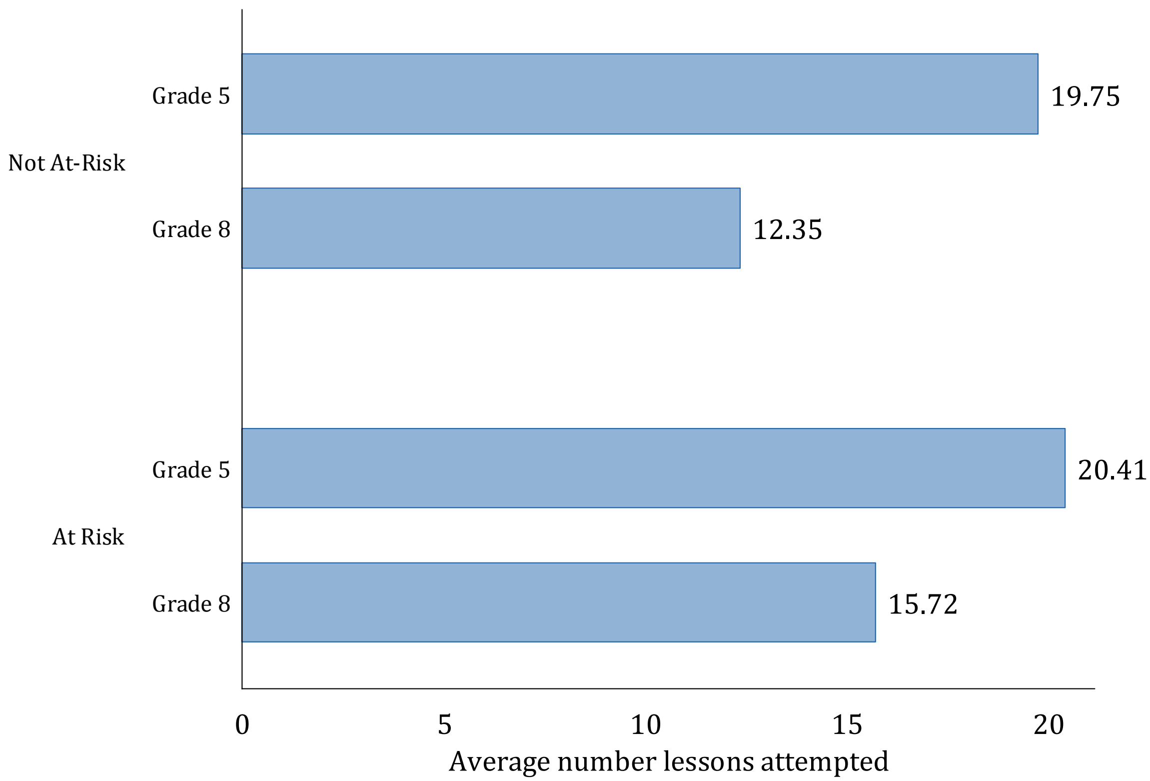
Note: Frequency counts and calculations only included students who met the following inclusion rules: were enrolled at a campus where at least one student was registered for TTM in 2013-14; had a valid STAAR-Mathematics test score in the prior year (2012-13). Only regular English and Spanish versions (i.e., not modified or alternate versions) were included. A student was defined as being at risk of being retained if he or she failed at least one STAAR-Mathematics assessment in 2011-12 or 2012-13.

Outcome Results

As Figure 3.15 illustrates, Grade 5 and 8 students at risk of being retained used the TTM system more intensively in 2013-14 than their peers who were not at risk of being retained in grade. This difference was smaller among Grade 5 students, where students at risk of being retained attempted roughly half of an additional TTM lesson, on average, compared to students who never failed the STAAR-Mathematics assessment in a prior school year. In Grade 8, students who had failed a STAAR-Mathematics assessment in a prior year attempted three or more TTM lessons compared to students who had never failed the STAAR test (15.72 compared to 12.35).

[76] A student was defined as being at risk of being retained if he or she failed at least one STAAR-Mathematics assessment in 2011-12 or 2012-13. The analysis was confined only to students who were in Grades 5 and 8 in 2013-14 (i.e., grades where grade promotion is conditional on passing STAAR exams), and who were enrolled at a school where at least one student was registered in the TTM system. In addition, only students with at least two valid STARR-Mathematics assessment scores (in 2012-13 and 2013-14) were included, eliminating students for whom the evaluation team had no information on prior test performance.

Figure 3.15. Average Number of TTM Attempts in 2013-14, by At-Risk Status and Grade Level



Source: State of Texas Assessments of Academic Readiness data, 2011-12 – 2013-14, Texas Education Agency. Think Through Math session history table, 2013-14.

Note: Calculations only included students who met the following inclusion rules: were enrolled at a campus where at least one student was registered for TTM in 2013-14; had a valid STAAR-Mathematics test score in the prior year (2012-13). Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in the at-risk indicator flag. A student was defined as being at risk of being retained if he or she failed at least one STAAR-Mathematics assessment in 2011-12 or 2012-13.

To further investigate the differences in TTM system usage and dosage on student performance between students at risk of being retained and students who were not at risk, the evaluation team used statistical models that explored the relationship between two separate measures of system dosage and students' decile-standardized STAAR-Mathematics gain scores between 2012-13 and 2013-14: 1) student used the system at least once (compared to no usage); and 2) the number of lesson attempts.[77] The results are presented in Table 3.27.

[77] More technical detail about the estimation procedure and functional form is provided in Appendix B.

The first row of estimates for each grade in the table shows the estimate comparing students at risk of being retained versus those who were not if they attempted at least one lesson. There was no statistically significant difference between the groups.

Grade 5 and 8 students who were at risk of being retained, who had more attempts, experienced smaller gains (-0.001 and -.0.002 deviations respectively) compared to students who were not at risk of being retained.

Among more frequent users, 20 or more lessons attempted was associated with significant but smaller gains (-0.084 standard deviations) relative to students who were not at risk of being retained. Findings in Grade 8 deviated somewhat from the descriptive trends, in that—while additional lessons attempted was associated with smaller descriptive declines—the only statistically significant association between TTM use and mathematics gain scores was the negative association between 15 and 19 lessons.

Table 3.27. Estimates of the Differences in the Relationship between Students at Risk of Being Retained, by TTM System Usage, 2013-14

Grade Level	Usage Measure	B	SE	N
Grade 5	Attempted at least one lesson	0.001	0.009	317,316
	Number of attempts	***-0.001******	0.000	
	1-4 attempts	0.024	0.014	
	5-9 attempts	**0.064*****	0.013	
	10-14 attempts	-0.009	0.014	
	15-19 attempts	0.030	0.016	
	20+ attempts	***-0.084******	0.012	
Grade 8	Attempted at least one lesson	0.022	0.013	235,039
	Number of attempts	***-0.002****	0.000	
	1-4 attempts	0.029	0.018	
	5-9 attempts	0.008	0.024	
	10-14 attempts	0.021	0.026	
	15-19 attempts	***-0.074****	0.032	
	20+ attempts	-0.053	0.028	

Source: State of Texas Assessments of Academic Readiness and Public Education Information Management System data, 2011-12 – 2013-14, Texas Education Agency. Think Through Math session history table, 2013-14.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in the analyses. Minutes were prior to the administration of the first assessment, which is contingent on grade level. Estimates were derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. TTM dosage threshold was 250 minutes in Grade 5 and 200 minutes in Grade 8. A student was defined as being at risk of being retained if he or she failed at least one STAAR-Mathematics assessment in 2011-12 or 2012-13. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Guiding Research Question 9b: Summary of Findings

Analyses in this section explored relationships between TTM usage and changes in student performance on STAAR-Mathematics assessments between Grade 5 and 8 students at risk of being retained and those not at risk of being retained. A student was defined as being at risk of being retained if he or she failed at least one STAAR-Mathematics assessment in 2011-12 or 2012-13.

TTM usage among students at risk of being retained in Grade 5—relative to students not at risk—was similar while, in Grade 8, students at risk of being retained completed over three more lessons than their peers not considered to be at risk. Multivariate analyses showed that Grade 5 students at risk of being retained, who attempted between 5-9 lessons experienced positive gains compared to students who were not at risk of being retained. Multivariate findings in Grade 8, however, deviated from the descriptive trends, in that the only statistically significant association between TTM use and mathematics gains was the negative association between 15 and 19 lessons TTM attempted.

Guiding Question 10b:

What is the relationship between program usage after failing the first administration of the STAAR-Mathematics assessment and the probability of passing subsequent administrations of the STAAR exam for math?

Key Descriptive Findings

- **Descriptive usage patterns varied by grade, with approximately 14% of Grade 8 students who failed the first administration attempting at least one TTM lesson in the period between the first and second administrations, compared to 29% of Grade 5 students.** On average, Grade 5 students who failed the first administration of the Grade 5 STAAR test completed over two times as many TTM lessons between the first and second administrations of the test as Grade 8 students (2.36 compared to .96 attempts).

Key Multivariate Findings

- **In Grade 5, high levels of TTM usage between STAAR-Mathematics test administrations was associated with higher passing rates on the second administration of the test.** Among Grade 5 students who failed the first STAAR-Mathematics assessment, students who attempted 15 or more lessons in the period between the first and second administrations were significantly more likely to pass the retest compared to students who did not use the TTM system. However, there was no statistically significant increase in the probability of passing STAAR upon retake among students with between 1 and 14 lessons attempted.
- **In Grade 8, TTM usage of 1 to 9 lessons between STAAR-Mathematics test administrations was associated with higher passing rates on the second administration of the test.** Among Grade 8 students who failed the first STAAR-Mathematics assessment, students who attempted between 1and 9 lessons during the period between the first and second administration were significantly more likely to pass the retest compared to students who did not use the system. There was no statistically significant increase in the probability of passing STAAR upon retake demonstrated among students with more than 10 lessons attempted.

Prior analyses related to STAAR-Mathematics assessment outcomes focused on students' results on the first administration of the STAAR-Mathematics exam. Analyses in this section explore, among students who did not pass the first administration, relationships between TTM use that occurred *between* the first and second administrations of the assessment and the propensity of students to pass the test on their second attempt. Among students who failed the first administration of Grades 5 and 8 STAAR-Mathematics, associations between TTM usage between the first and second assessment (April 3rd and May 13th, 2014) and the likelihood of passing the second administration was investigated by using a statistical model that controlled for differences in student characteristics and the amount of time a student used TTM between the first and second STAAR-Mathematics assessment.[78] Similar to the limitation expressed for Istation, It is important to recognize that students who fail the first administration of the STAAR assessment are provided with a wide array of intensive mathematics interventions which vary by school district. Therefore, in addition to TTM, a variety of factors and interventions may be contributing to student performance on the second administration of the STAAR-mathematics assessment.

Descriptive Analysis of Usage between First and Second STAAR Administration

Table 3.28 provides first and second administration dates for STAAR-Mathematics in 2013-14 for Grades 5 and 8. The last column presents dates for which TTM system usage was calculated. Lesson attempts during this period were calculated by summing total lessons attempted during this period.

Table 3.28. STAAR-Mathematics Test Administration Dates for Grades 5 and 8, 2013-14

	First Administration	Second Administration	Inter-Administration Usage Dates
Grade 5	April 2, 2014	May 14, 2014	April 3rd-May 13th
Grade 8	April 2, 2014	May 14, 2014	April 3rd-May 13th

Source: Texas Education Agency, 2014.

A total of 73,373 students who failed the first administration of the Grade 5 STAAR-Mathematics assessment (97%) and 60,320 who failed the first administration of the Grade 8 test (97%) in April 2014 had a valid score on the second administration in May 2014 (see Table 3.29).

[78] A logistical regression model was employed for this analysis because of the dichotomous nature of the dependent variable (i.e., whether or not the student passed the second administration of the STAAR-Mathematics assessment). With this framework, the differential probability of passing the second administration STAAR-Mathematics assessment between students who used TTM at different levels between assessments was estimated.

Table 3.29. Frequency Count of Students Who Failed the First Administration of STAAR-Mathematics and Received a Valid Test Score on the Second Administration, by Grade, 2013-14

Student Grade Level	Indicator of Whether Student Took 2nd Administration Test	Number of Students	Percentage of Students
Grade 5	No second administration test	2,414	3.18%
Grade 5	Took second administration test	73,373	96.81%
Grade 8	No second administration test	2,125	3.40%
Grade 8	Took second administration test	60,320	96.60%

Source: State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level included the second administration only, and regular English and Spanish versions only (i.e., not modified or alternate versions). Frequency counts only included students who failed the first administration of STAAR-Mathematics in April 2014.

Next, system usage patterns of students who failed the first administration are described. This analysis included only students in Grades 5 and 8 in 2013-14 who failed the first administration of STAAR-Mathematics, took the second administration of the test, and had a valid score in the TEA databases. Among students who failed the first administration of the STAAR-Mathematics assessment, TTM system usage patterns in the period between the first and second administrations were markedly different between Grade 5 and Grade 8 students. This finding is consistent with usage pattern differences throughout the year between Grade 5 and Grade 8 students.

Table 3.30 depicts counts and percentages of Grade 5 and Grade 8 students who attempted at least one lesson between first and second STAAR administrations. Three additional usage metrics are presented for these students: average number of lesson attempts during the year, average number of lessons attempted prior to the first STAAR administration, and average number of lessons attempted between first and second STAAR administrations. Roughly 14% of Grade 8 students who failed the first administration used TTM in the period between the first and second administrations, compared to 29% of Grade 5 students. The intensity of use (i.e., number of attempted TTM lessons) among Grade 5 students who failed the first administration was also higher, by nearly two and a half lessons, compared to Grade 8 students (2.31 attempts compared to .96 attempts) in the interim period.

Table 3.30. TTM Use between First and Second Administrations of STAAR-Mathematics, by Grade, 2013-14

Student Grade Level	Count of Students	Percentage of Students Who Attempted at Least One Lesson between the First and Second Administrations	Mean Number of Attempts During 2013-14	Mean Number of Attempts Prior to First Administration	Mean Number of Attempts between First and Second Administration
Grade 5	73,373	29.44%	10.16	0.48	2.31
Grade 8	60,320	14.03%	4.45	0.30	0.96

Source: Think Through Math session history table, 2013-14.

Note: Frequency counts and calculations only included students who failed the first administration of STAAR-Mathematics in April 2014 and had a validly scored test on the second administration. Only regular English and Spanish versions (i.e., not modified or alternate versions) were included. Period between first and second administrations was April 3, 2014 to May 13, 2014.

Table 3.31 provides the distribution of students by usage intensity category that were used in the multivariate analysis for this section. Most students (71% of Grade 5 students and 86% of Grade 8 students) who failed the first administration did not record lesson attempts in the interim period. Twenty-nine percent of Grade 5 students and 14% of Grade 8 students attempted one or more lessons between STAAR-Mathematics test administrations.

Table 3.31. Count and Percentage of Students by Interim Usage Category, by Grade Level, 2013-14

Student Grade Level	Usage Intensity Category	Number of Students	Percentage of Students
Grade 5	0 attempts	51,773	70.56
Grade 5	1-4 attempts	11,022	15.02
Grade 5	5-9 attempts	5,263	7.17
Grade 5	10-14 attempts	2,253	3.07
Grade 5	15-19 attempts	1,162	1.58
Grade 5	20+ attempts	1,900	2.59
Grade 8	0 attempts	51,856	85.97
Grade 8	1-4 attempts	4,721	7.83
Grade 8	5-9 attempts	1,866	3.09
Grade 8	10-14 attempts	868	1.44
Grade 8	15-19 attempts	416	0.69
Grade 8	20+ attempts	593	0.98

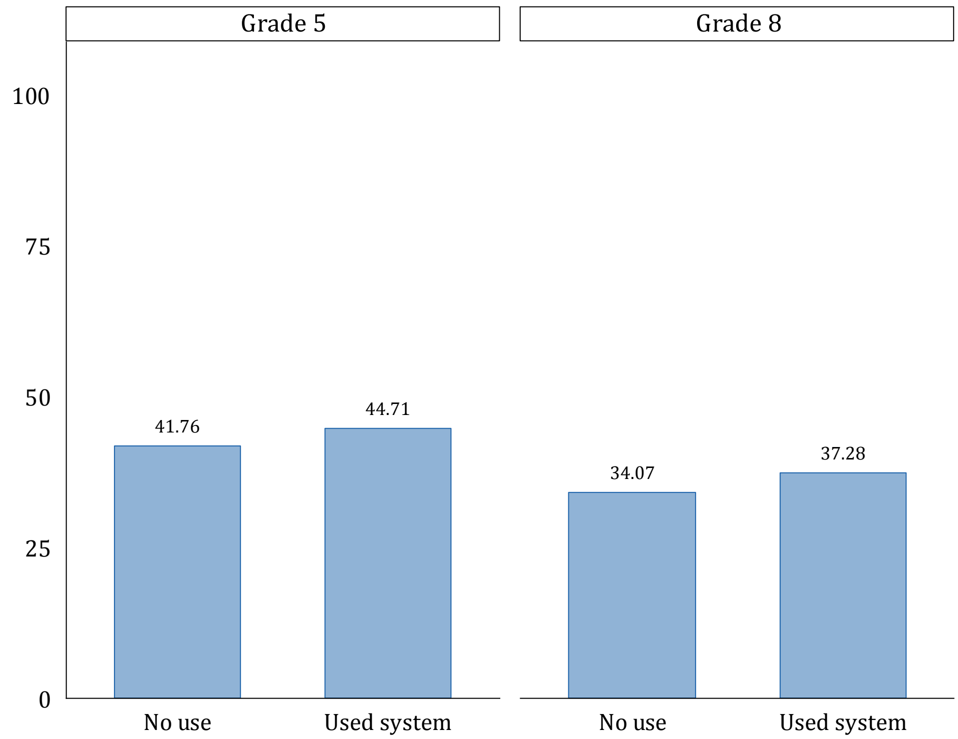
Source: Think Through Math session history table, 2013-14.

Note: Frequency counts and calculations only included students who failed the first administration of STAAR-Mathematics in April 2014 and had a validly scored test on the second administration. Period between first and second administrations was April 3, 2014 to May 13, 2014. A curriculum session is defined as a session with greater than zero minutes of use.

Outcome Results

Figures 3.16 and 3.17 provide descriptive information about the association between system use in the interim period and the percentage of students who met the passing standard on the second administration of the Grade 5 or 8 STAAR-Mathematics test. Figure 3.16 defines TTM usage narrowly, based only on whether the student attempted a lesson in the interim period between the first and second assessments, while Figure 3.17 classifies usage according to the number of lessons attempted during this period. For Grade 5 and Grade 8 students (Figure 3.16), the passing rate for students who used the system was about three percentage points higher than students who did not use the system (Grade 5: 45% compared to 42%; Grade 8: 37% compared to 34%).

Figure 3.16. Percentage of Students Who Met Phase-In 1 Level II (Satisfactory) Standard on Second Administration of STAAR-Mathematics, by Interim TTM Usage, Grades 5 and 8, 2013-14

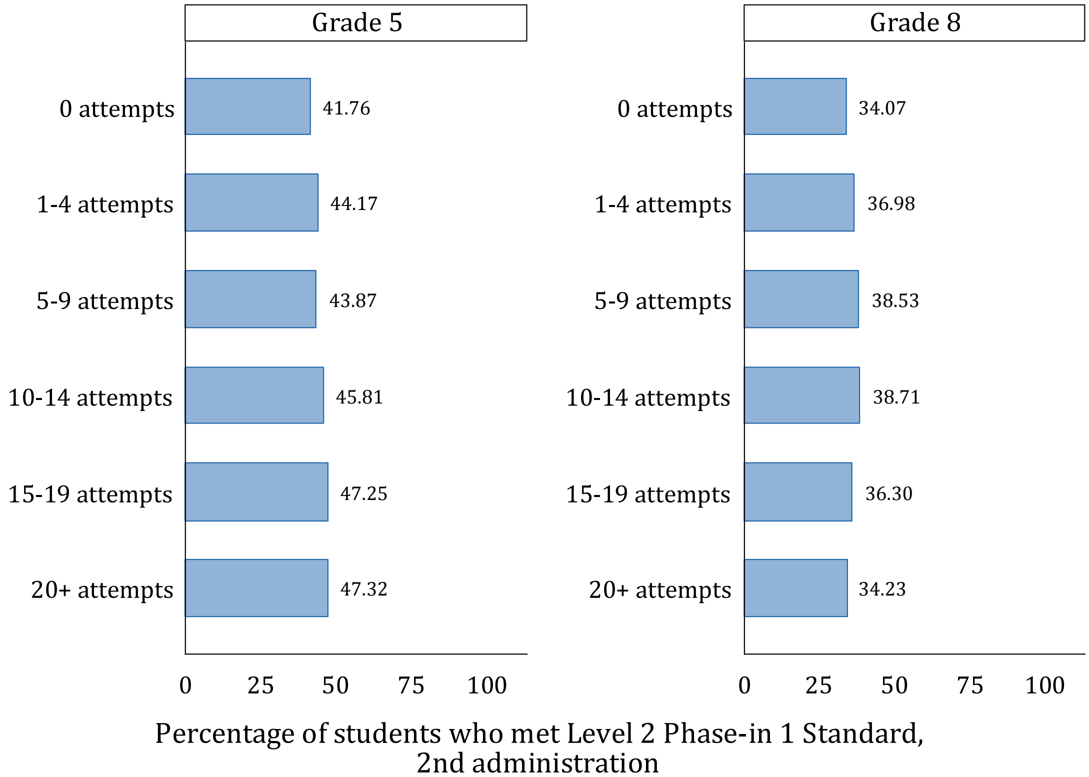


Source: Think Through Math session history table, 2013-14. State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: Calculations only included students who failed the first administration of STAAR-Mathematics (April 2014) and had a validly scored test on second administration. Only regular English and Spanish versions (i.e., not modified or alternate versions) were included. The period between first and second administrations was April 3, 2014 to May 13, 2014.

Figure 3.17 also indicates a positive relationship between usage intensity and the percentage of students who met the passing standard on the second administration among Grade 5 and Grade 8 students. In Grade 5, as the number of lessons attempted increased in the interim period between the first and second assessments, so did the percentage of students who met the passing standard on the second administration. Across all usage categories, students who attempted TTM lessons passed the second administration at higher rates than students who did not use the TTM system. This pattern was also apparent for Grade 8 students, though most notably through the 10-14 attempts threshold, after which the percent of students who met the phase-in 1 Level II (Satisfactory) standard declined. Those students with over 20 TTM lessons attempted passed the second administration at roughly the same rates as students with zero attempts in the interim period.

Figure 3.17. Percentage of Students Who Met the Phase-In 1 Level II (Satisfactory) Standard on Second Administration of STAAR-Mathematics, by Interim TTM Usage, Grades 5 and 8, 2013-14 School Year



Source: Think Through Math session history table, 2013-14. State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: Calculations only included students who failed the first administration of STAAR-Mathematics in April 2014 and had a validly scored test on the second administration. Only regular English and Spanish versions (i.e., not modified or alternate versions) were included. Period between first and second administrations was April 3, 2014 to May 13, 2014.

While these results begin to illustrate relationships between TTM usage and mathematics outcomes among students who retook the STAAR-Mathematics test in 2013-14, they are also purely descriptive. They do not account for differences in other characteristics between students who used the system and students who did not, and which may be correlated with changes in student performance, nor does it account for differences across campus types or environments. To account for these differences, the evaluation team investigated how TTM usage was associated with the student passing the second administration STAAR-Mathematics test. In the model, the evaluation team adjusted for student attributes, including their performance on the first STAAR-Mathematics administration. Models were fit separately for Grade 5 and Grade 8 to account for differences in the relationship between usage and usage intensity, and the likelihood of earning a satisfactory score, across grade levels. The results of this analysis are presented in Table 3.32.

Among Grade 5 students who failed the first STAAR-Mathematics assessment, students who attempted 15 or more lessons were significantly more likely to pass the second administration than students with no TTM usage. Notably, lower numbers of lesson attempts (between 1 and 14 lessons) were not significantly related to increased probabilities of passing the second administration.

By contrast, the likelihood of passing the second administration of the Grade 8 STAAR-Mathematics test was significantly higher among students who used the system at least once (0.111), particularly those with between 5 and 9 lessons attempted (0.177). In sum, the likelihood of passing the second administration of the STAAR was greater among Grade 5 students with at least 15 lessons attempted while, among Grade 8 students, 1 to 9 lessons attempted was associated with increased probability of passing the test.

Table 3.32. Estimated Relationships between TTM Usage between the First and Second STAAR-Mathematics Administration, Grades 5 and 8, 2013-14

Grade Level	Usage Measure	ME	SE	N
Grade 5	Used at least once	0.049	0.025	68,623
	Number of attempts	**0.005****	0.002	
	1-4 attempts	0.021	0.029	
	5-9 attempts	0.021	0.041	
	10-14 attempts	0.050	0.055	
	15-19 attempts	**0.226****	0.081	
	20+ attempts	**0.187****	0.064	
Grade 8	Used at least once	**0.111****	0.038	55,600
	Number of attempts	0.005	0.003	
	1-4 attempts	**0.091***	0.045	
	5-9 attempts	**0.177****	0.064	
	10-14 attempts	0.156	0.084	
	15-19 attempts	0.042	0.108	
	20+ attempts	0.047	0.134	

Source: Think Through Math session history table, 2013-14. State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: Calculations only included students who failed the first administration of STAAR-Mathematics in April 2014 and had a validly scored test on the second administration. Only regular English and Spanish versions (i.e., not modified or alternate versions) were included. Period between first and second administrations was April 3, 2014 to May 13, 2014. Estimates were derived from a logistic regression with cluster-robust standard errors at the campus level. The covariates in the full functional form are provided in Appendix 10. Marginal effects were derived by holding all values at their means. Marginal effects for covariates that were not statistically significant were not calculated and are not presented. Statistically significant positive coefficients are denoted by bold font. * p<0.05, ** p<0.01, *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Guiding Research Question 10b: Summary of Findings

To assess the relationship between program usage after failing the first administration of the STAAR-Mathematics assessment and the probability of passing subsequent administrations of the STAAR exam, evaluators examined usage and usage intensity during the period between failure of the first administration and the second administration of STAAR-Mathematics in SSI grades (Grades 5 and 8).

Descriptive usage patterns varied by grade, with approximately 14% of Grade 8 students who failed the first administration using TTM in the period between the first and second administrations, compared to 29% of Grade 5 students. Grade 5 students who failed the first administration of Grade 5 STAAR-Mathematics completed over two times as many TTM lessons between the first and second administrations of the test as Grade 8 students (2.31 compared to .96 attempts). Among students who used the system—relative to those who failed the first administration and did not use the system—the passing rate for both Grades 5 and 8 TTM users was roughly three percentage points higher than students who did not use the system after failing the first administration.

To control for other factors that may have impacted retest passing rates—outside of TTM use—the evaluation team fit a logistic regression model, regressing whether a student passed the second administration STAAR-Mathematics tests on TTM usage categories, while adjusting for other student attributes.

- Among Grade 5 students who failed the first STAAR-Mathematics assessment, students who attempted 15 or more lessons in the period between the first and second administrations were significantly more likely to pass the retest compared to students who did not use the TTM system. There was no statistically significant increase in the probability of passing STAAR upon retake among students with between 1 and 14 lessons attempted.
- Among Grade 8 students who failed the first STAAR-Mathematics assessment, students who attempted between 1 and 9 lessons during the period between the first and second administrations were significantly more likely to pass the retest compared to students who did not use the system. There was no statistically significant increase in the probability of passing STAAR upon retake demonstrated among students with more than 10 lessons attempted.

In sum, among Grade 5 students, the likelihood of passing the second administration of STAAR was only associated with higher numbers of uses—using the TTM system at least once was not significantly related to the probability of passing. Conversely, using the system at all—particularly attempting a moderate number of lessons—was associated with improved passing probabilities among Grade 8 students. The same was not true of Grade 8 students who attempted more than 10 lessons.

Section 4 – Additional FY 2015 Analyses

While this report addresses the vast majority of the research questions posed in this study of the Texas SUCCESS Initiative, three key research questions still remain and will be addressed in addendum reports to the Texas Education Agency as data become available. Data collection and analysis activities to address the following three questions will be conducted in spring/summer 2015 (with reports delivered to TEA by August 2015):

1) **To what extent do performance results for Grade 9 students who were retained in Grade 8 differ on the Grade 9 Algebra I and Grade 9 English I end-of-course (EOC) exams for students participating in the Texas SUCCESS program (i.e., using TTM and/or Istation) and non-participating students?**

   This research question involves a quantitative analysis of changes in student STAAR test performance after Texas SUCCESS implementation, between participating and non-participating students. Analyses will be disaggregated by changes in student assessment scores for students in Grade 8 who were repeaters in 2013-14 (i.e., they were not promoted between 2012-13 and 2013-14), but who were promoted in 2014-15 (i.e., they were promoted to 9th grade).

2) **To what extent do 2013-14 grade promotion/retention rates differ between students previously retained in grade for Grade 5 and Grade 8 students participating and not participating in the Texas SUCCESS program?**

   This research question involves a quantitative analysis of differences in odds of being retained between participating and non-participating students. Analyses will be disaggregated for two student groups:

   a) Student Group 1: Odds of first-time Grade 5 or 8 students in 2013-14 being retained in 2014-15.
   b) Student Group 2: Odds of first-time Grade 5 or 8 students in 2013-14, who were at risk of being retained in 2014-15, being retained in 2014-15.
   c) The relationship between program participation and grade retention/promotion rates will be addressed through the use of logistic regression models to calculate the log-odds of Grade 5 and Grade 8 students at risk of being retained in grade in 2013-14 actually being retained in grade in 2014-15. Differences in grade retention rates will be assessed for students using the Texas SUCCESS programs and non-users.

3) **What proportion of campuses substantively changed their Istation and TTM usage patterns between the 2013-14 and 2014-15 school years (i.e., moved from a low usage campus to a high usage campus or moved from a high usage campus to a low usage campus) and what were the primary reasons for the changes in usage?**

   This research question will be addressed through the analysis of 2014-15 campus-level Istation and TTM system usage data, comparing the results to 2013-14 usage metrics and determining if large shifts in usage occurred. Second, an online survey will be administered to campus staff at schools that experienced significant changes in usage patterns between 2013-14 and 2014-15 to determine reasons for modifications in usage patterns.

Appendix A – Istation Technical Section

Guiding Question 7a.1

Outcome Measure Operationalization and Calculation

Standardized gain scores

The outcome of interest for the evaluation was STAAR gain scores between 2012-13 (*T-1)* and 2013-14 (*T).* Gain scores were standardized to account for test differences between school years, grade-level assessments, and testing regimes, which, if not normalized, can be difficult to compare. The evaluation team approached standardization in two ways. First, the evaluation team calculated a simple gain score from standardized STAAR-Reading scores in 2012-13 and 2013-14. The standardization was done by grade level and school year, and is similar to the technique used in other evaluations that use gain scores as the outcome measure (Zimmer, Gill, Razquin, Booker, & Lockwood III, 2007; Springer, Pepper, & Ghosh-Dastidar, 2009; Heinrich et al., 2010).

For each school year, and grade level, the evaluation team calculated standardized gain scores. First, the team calculated this term by standardizing student test scores, by subject, grade, and year, and subtracted the standardized score at time *T* from standardized scores at time *T-1*. Positive scores indicate gains above the mean for students in the respective subject, grade, and year, while negative scores indicate that students' standardized test score was below the mean for students in the subject, grade, and year. The method for calculating this measure is as follows:

1. First calculate, for each subject, year, and grade, the average scaled score (represented below by SSC) at time *T* and subtract this from each student *i*'s actual score in subject *s,* grade *g,* at time *t*.
    a. $\mu_{istg} = SSC_{istg} - \overrightarrow{\overline{SSC_{stg}}}$
2. Next, divide μ_{stg} by the standard deviation (σ_{sgt}) of the scaled score to produce the subject-grade-year standardized test score:
    a. $STD_{istg} = \frac{\mu_{istg}}{\sigma_{sgt}}$
3. Standardized gain scores were calculated by simply subtracting each student's standardized test score at time *T-1* STD_{it-1} from their standardized score (STD_{it}) at time *T.*

Decile-standardized gain scores

Evaluators constructed an alternative measure of a standardized gain score based on the work of Hanushek, Kain, O'Brien, and Rivkin (2005),[79] and closely related to the approach proposed by Reback (2008).[80] The evaluation team calculated a decile-standardized gain score which standardizes changes in student performance, across years and grades, by their prior test performance decile. Decile-standardization has a number of properties which make it a desirable operationalization of student performance changes across time.

First, because student gain scores are standardized by their placement in the distribution of scores on the prior year test, the measurement is less sensitive to mean reversion and test ceiling effects (Hanushek et al., 2005; Reback, 2008). This is particularly important given evidence that student test scores are volatile across time, and that students at the bottom end of the test score distribution tend to exhibit larger gains between *T-1* and *T* than students at the upper portion of the distribution (Hanushek et al., 2005). Second, in the multivariate models fit in the second part of this section, standardizing gains by prior performance decile eliminates the need to control for prior performance in the statistical models.

To account for students placement in the prior-score distribution, which may determine how much students' scores change across years, and to account for possible mean reversion, the evaluation team implemented a gain score standardization method that standardizes test scores gains by subject, year, grade, and student's placement in the prior-year achievement distribution. The method is identical to the one proposed in Hanushek et al. (2005), and the notation and description of the calculation procedure below closely follows theirs.

1. First, for every subject, grade, and year (where years here include 2012-13 and 2013-14), divide the prior-year (2012-13 for students in 2013-14) subject and grade test score distribution into ten equal score intervals($c_m \; where \; m = 1, \ldots, 10)$
2. Next, calculate the mean and standard deviation of the *gains* for all students in each subject, grade, year, and *decile* interval with a test score at *T-1* (A_{isg-1}) and *T* (A_{isg}), using the notation from Hanushek et al. (2005):
    a. $\mu_{sgt}^{c_m} = \overrightarrow{A_{\imath sg} - A_{\imath sg-1}}$
    b. $\sigma_{sgt}^{c_m} = \sqrt{\sum((A_{isg} - A_{isg-1}) - \mu_{sgt}^{c_m})/n_{sgtc_m}}$
3. Then, the decile-standardized gain score for each student in interval c_m in each subject, year, and grade is:

[79] See Peterson and Chingos (2009) for an application of a similar technique in an evaluation of an educational intervention.

[80] See for instance Taylor, Gronberg, Jansen, and Karakaplan (2014) for a recent application of a comparable approach in Texas.

a. G_{isgt}=$[(A_{isg} - A_{isg-1}) - \mu_{sgt}^{c_m}]/\sigma_{sgt}^{c_m}$

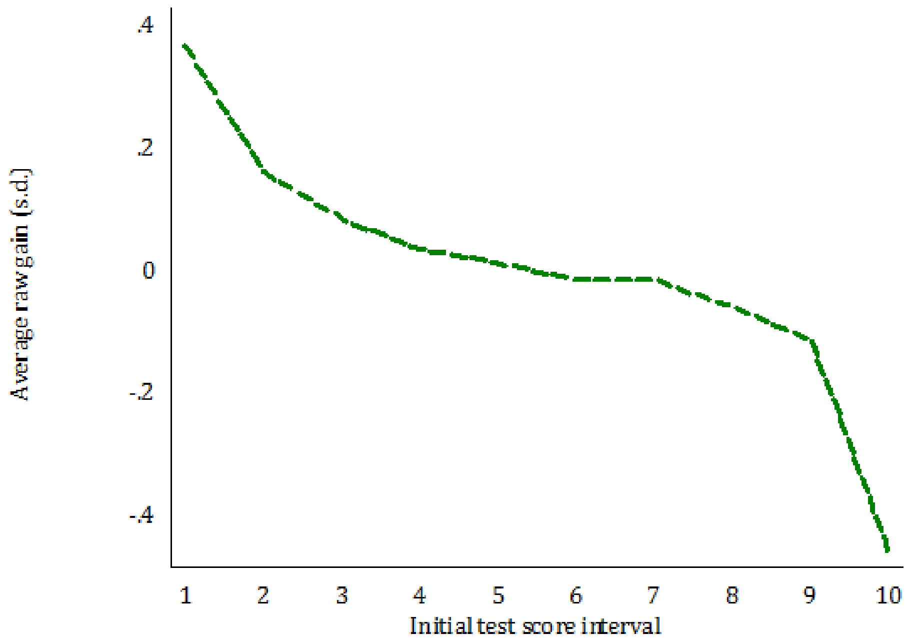
Empirical evidence justifying the use of decile-standardized gain scores

Performance gains on standardized tests are a function of a number of factors, including the quality of classroom instruction, unobserved and observed student-level attributes (e.g., motivation, parental involvement), transitory classroom or school shocks (such as a disruptive student or distractions due to school construction or renovations), and test measurement error (such as test ceiling or floor effects).[81] Furthermore, the amount of test score gains demonstrated by students is also conditioned, in part, on students' prior test performance level (Hanushek et al., 2005).

Figure A1 replicates the relationship between raw standardized gains (measured in standard deviations) and students' prior test score performance decile interval presented in (Hanushek et al., 2005). Raw standardized gains for students in the lowest intervals of the prior test score distribution (i.e., students in the bottom thee deciles) were much larger than those experienced by students in the top three deciles (i.e., 8, 9, and 10). The relationship depicted in Figure A1 introduces a disconcerting threat to the validity of inferences drawn from research designs that rely on the comparison of changes in test score across time, between populations of students with different underlying performance trajectories that are, at least in part, a function of their starting performance level. That is, using simple standardized gain scores may *misattribute mean-reverting gains to the program* since low-performing students disproportionately comprise the population of students who participated in the program, which occurs more frequently among students who use Istation Reading. That is, students in the participant group were more likely to have large gains precisely because they were low performing students. This bias was most severe in middle grades and in STAAR-Reading, where targeting was more prevalent based on low performance. This has important implications for the inferences drawn from the research design and statistical models: in essence, the gain would be fallaciously attributed to the program, which would make the program appear effective, when much of the gains that are being artificially attributed to the program are simply noise.

[81] See Kane and Staiger (2002) for an empirical demonstration of test score volatility, and its potential sources, across school years and test administrations.

Figure A1. Standardized Raw STAAR-Reading Gains between 2012-13 and 2013-14, by 2012-13 STAAR-Reading Performance Decile

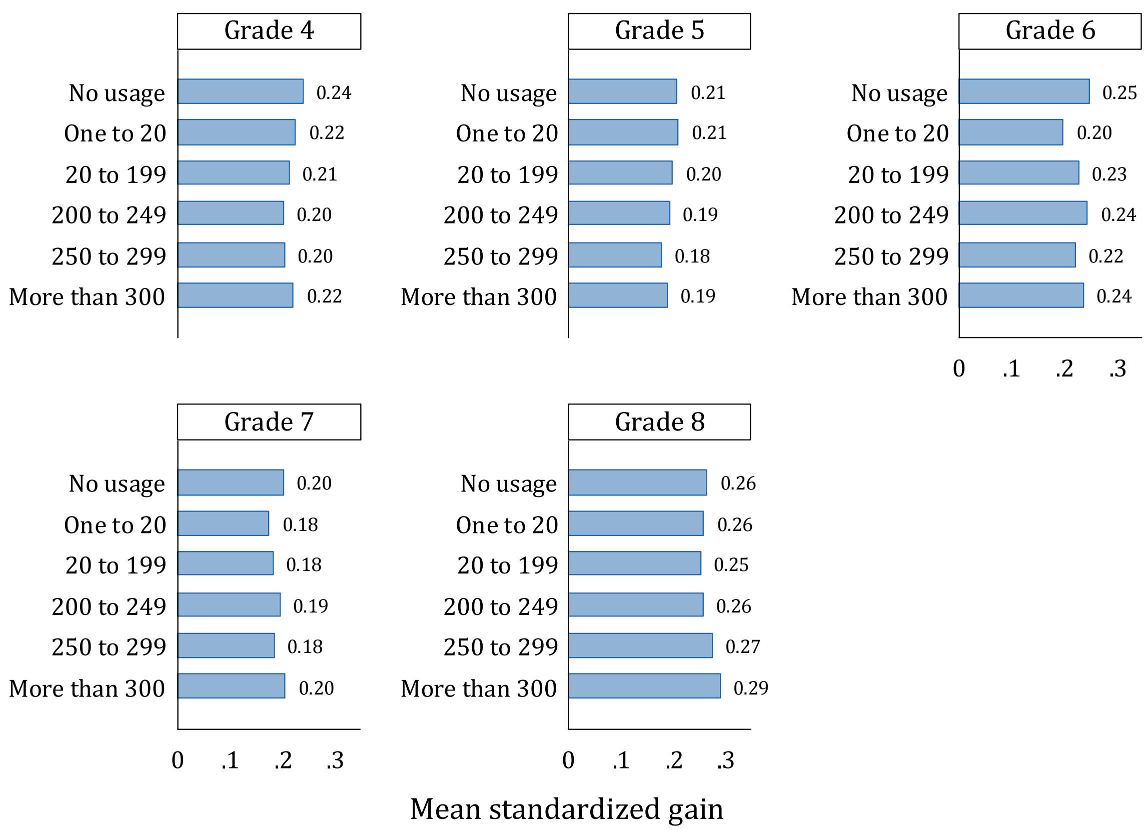


Source: State of Texas Assessments of Academic Readiness data, 2012-13 and 2013-14, Texas Education Agency.
Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish versions only (i.e., not modified or alternate versions). Only students with consecutive grade level assessments (e.g., three in 2012-13 and four in 2013-14) are included in the calculations. Initial test score interval is based on students' 2012-13 STAAR-Reading (*T-1)* scaled score.

Figures A2 and A3 provide additional evidence that supports the use of decile-standardized gain scores for the outcome measure. Figures A2 and A3 show the average standardized gain score between 2012-13 and 2013-14. The evaluation team calculated the average gain score by grade level and by the amount of time a student spent using Istation curriculum lessons in 2013-14 prior to the first STAAR administration. In addition, the evaluation team added an additional dimension to the figure to account for different performance levels on STAAR-Reading in 2012-13 between students across the system usage categories. Figure A2 includes students who were in the bottom quartile of performance on their prior year assessment (1st to 25thpercentile), and Figure A3 includes students in the second quartile (25th to 50th percentile).

There are two important patterns that are evident in Figures A2 and A3. First, the difference in the average grade-year standardized gain scores are much larger (they average around .25 of a standard deviation) among students who were in the bottom quartile of 2012-13 STAAR-Reading performance (Figure A2) compared to students in the second quartile (Figure A3), which ranged between -.04 and .07 standard deviations. Second, after adjusting with more precision for students' prior performance level by incorporating students' 2012-13 STAAR-Reading quartile, the differences between non-users and Istation users becomes more apparent. While differences in standardized gain scores between non-users and Istation users were either small or slightly higher among students who used the system at the highest dosage level measured in these figures for students in the bottom quartile of prior STAAR-Reading performance (Figure A2), non-users had larger gains than Istation users at each dosage level, with the exception of students in Grade 4 (Figure A3). Collectively, these figures illustrate the importance of accounting for students' baseline performance *level* when examining differences in test score gains between program participants and non-participants.

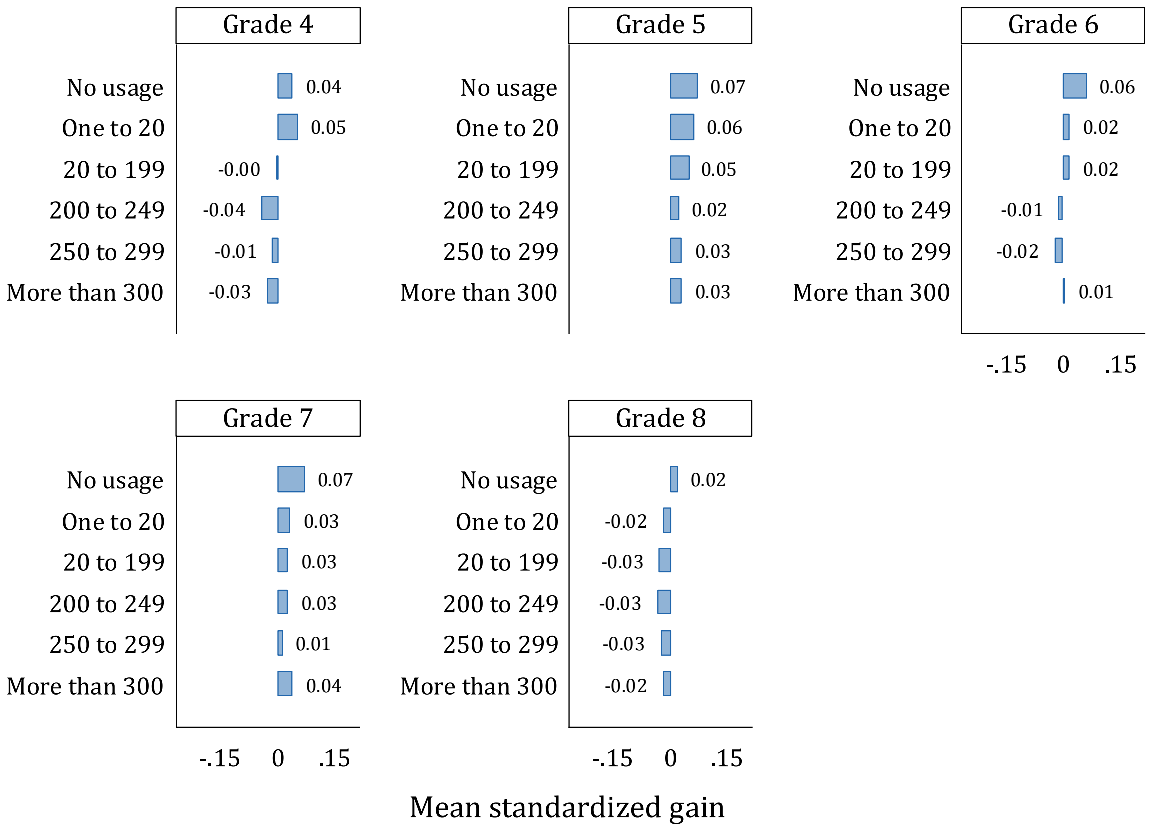
Figure A2. Mean Grade and Year Standardized Gain Score between 2012-13 and 2013-14 for Students in the Bottom Quartile of Scores on the 2012-13 STAAR-Reading Assessment, by Grade Level and Number of Istation Curriculum Minutes, 2013-14



Source: Istation session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions). Calculations only include students who took an ISIPER or ISIPAR assessment in September 2013.

Figure A3. Mean Grade and Year Standardized Gain Score between 2012-13 and 2013-14 for Students in the Second Quartile of Scores on the 2012-13 STAAR-Reading Assessment, by Grade Level and Number of Istation Curriculum Minutes, 2013-14



Source: Istation session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions). Calculations only students who took an ISIPER or ISIPAR assessment in September 2013.

As an additional specification check to support the use of decile-standardized gain scores for the outcome measure, the evaluation team constructed a "placebo" or "pseudo"-participant group comprised of students in 2013-14 who were non-users but who were observably similar to Istation users in 2013-14.[82] Most simply, if the pseudo-participant group demonstrated larger, statistically significant gains relative to other non-users between 2012-13 and 2013-14, after accounting for other observable differences between the two student groups, the estimates of the impact of program usage are likely to be biased. We performed this specification test using both the raw standardized gain score and the decile-

[82] See Bertrand, Duflo, and Mullainathan (2004) for a deeper, more technical exposition of this approach this as a sensitivity check for DiD designs, and see Figlio and Rouse (2006) for an application of this test in an evaluation of the effect of school accountability policies on changes in performance of low-performing schools.

standardized gain score. Using the simple raw standardized gain score, the pseudo-participant group had statistically significant, positive gains compared to non-participants, and these differences were most pronounced in upper elementary and middle grades, indicating that raw standardized gain scores may produce unreliable, biased estimates of the association between program participation and test score gains. However, transforming gains by decile-standardization generated estimates of the differences between program users and non-users that were not statistically different from zero for Grades 4-8. As a result, all statistical models that are presented in the narrative section of the evaluation report rely on decile-standardized gain scores as the outcome measure.

Inclusion rules for schools and students in the analytic sample

For the multivariate models of student outcomes, the evaluation team restricted the analytic sample based on a number of inclusion rules. They are:

1) Only campuses where at least one student was *registered* for Istation in 2013-14, to minimize incomparability across different educational settings.
2) Students who were promoted across consecutive grades between 2012-13 or 2013-14 to ensure students' with similar academic trajectories are included in the analytic sample.
3) Students who took grade-level STAAR-Reading assessments that matched their grade level in PEIMS. Thus, students in Grade 8 who took an above-grade standardized assessment—such an End of Course (EOC) exam, were omitted from the multivariate analysis to standardized the assessments that are being used to compare performance differences between Istation users and non-users.
4) Only students with a valid test score on the first STAAR administration, and who took the regular English and Spanish versions (i.e., not the modified or alternate versions).
5) Only students with at least two valid STAAR-Reading test scores in 2012-13 and 2013-14, since gain scores could not be calculated without both a prior and current test score.
6) Only students in Grades 4-8 in 2013-14.

Tables A1 and A2 provide summary statistics and frequency counts for students included in the analytic sample. Table A2 disaggregates this information by whether a student met the grade-level Istation usage threshold during the 2013-14 school year. The descriptive statistics, while not regression-adjusted for other differences between participating and non-participating students, illuminate some of the student-level mechanisms through which students are selected to participate in Istation.

Among Istation users in 2013-14, their 2011-12 and 2012-13 STAAR-Reading scores were lower than non-users, and these differences are larger when comparing users who meet the grade level recommended usage threshold to students who do not. Consequently, their standardized gain scores (not adjusted for their placement in the prior distribution of performance) were *larger* than non-users. This is an important finding to keep in mind throughout the outcomes sections, and particularly in the statistical models estimated within a DiD framework. In addition, this pattern highlights the importance of decile standardization of the gain score to account for students' location in the prior test score distribution, since this placement is correlated with changes in student performance across school years (Hanushek et al., 2005).

Table A1. Summary Statistics for Student Test Scores, Standardized and Unstandardized Gain Scores, and Prior Performance, by Istation Usage Measure and Grade Level, 2013-14

Grade Level	Usage Measure	Decile Standardized Gain Score	Standardized Gain Score	Unstandardized Gain Score	2013-14 STAAR-Reading Score	2012-13 STAAR-Reading Score	2011-12 STAAR-Reading Scaled Score	Number of Unique Students
Two Valid STAAR-Reading Test Scores								
Grade 4	Did not use	0.05	-0.02	74.82	1516.44	1440.06	NA	120,332
Grade 4	Used at least one minute	-0.02	-0.03	73.61	1494.18	1418.73	NA	241,617
Grade 5	Did not use	0.03	-0.01	44.02	1570.22	1525.05	1441.31	133,424
Grade 5	Used at least one minute	-0.02	-0.01	46.30	1544.49	1496.59	1415.07	227,744
Grade 6	Did not use	0.04	-0.01	38.90	1606.27	1567.39	1530.67	205,356
Grade 6	Used at least one minute	-0.05	0.00	40.06	1569.53	1529.15	1492.70	151,514
Grade 7	Did not use	0.03	0.00	39.29	1652.98	1613.40	1562.13	235,956
Grade 7	Used at least one minute	-0.04	0.03	52.24	1603.67	1551.01	1515.73	127,856
Grade 8	Did not use	0.02	0.00	49.09	1700.73	1651.98	1606.70	259,956
Grade 8	Used at least one minute	-0.05	0.03	49.89	1651.32	1602.43	1555.61	99,610

Source: State of Texas Assessments of Academic Readiness data, 2011-12 to 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions). Counts and summary statistics only include students who met the following sample inclusion rules: Only campuses where at least one student was *registered* for Istation in 2013-14; students who progressed consecutive grades consecutively between 2012-13 or 2013-14 or between 2011-12 and 2013-14; students who took grade-level STAAR-Reading assessments that matched their grade level in PEIMS. Only system usage minutes that occurred prior to the first test administration are used to categorize students based on their system usage. 2011-12 STAAR-Reading scaled scores are denoted by an "NA", since STAAR-Reading is only administered in Grades 3-8.

Table A2. Summary Statistics for Student Test Scores, Standardized and Unstandardized Gain Scores, and Prior Performance, by Whether Student Met the Grade Level Istation Usage Threshold and Grade Level, 2013-14

Grade Level	Usage Measure	Decile Standardized Gain Score	Standardized Gain Score	Unstandardized Gain Score	2013-14 STAAR-Reading Scaled Score	2012-13 STAAR-Reading Score	2011-12 STAAR-Reading Scaled Score	Number of Unique Students
Two Valid STAAR-Reading Test Scores								
Grade 4	Below threshold	0.02	-0.03	74.01	1509.79	1434.21	1257.34	255,469
Grade 4	At/above threshold	-0.05	-0.03	73.99	1481.51	1405.32	1245.40	106,480
Grade 5	Below threshold	0.01	-0.01	44.35	1562.58	1516.95	1433.74	272,895
Grade 5	At/above threshold	-0.04	0.00	48.94	1526.93	1476.06	1396.52	88,273
Grade 6	Below threshold	0.01	-0.01	38.72	1597.10	1558.31	1521.37	308,220
Grade 6	At/above threshold	-0.07	0.03	43.70	1549.02	1504.70	1470.21	48,650
Grade 7	Below threshold	0.01	0.00	41.96	1641.37	1599.09	1551.52	332,354
Grade 7	At/above threshold	-0.03	0.08	64.32	1573.65	1508.75	1483.76	31,458
Grade 8	Below threshold	0.00	0.00	48.85	1691.21	1642.91	1596.95	339,714
Grade 8	At/above threshold	-0.04	0.12	57.55	1615.15	1557.19	1515.31	19,852

Source: State of Texas Assessments of Academic Readiness data, 2011-12 to 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions). Counts and summary statistics only include students who met the following sample inclusion rules: Only campuses where at least one student was *registered* for Istation in 2013-14; students who progressed consecutive grades consecutively between 2012-13 or 2013-14 or between 2011-12 and 2013-14; students who took grade-level STAAR-Reading assessments that matched their grade level in PEIMS. 2011-12 STAAR-Reading scaled scores are denoted by an "NA", since STAAR-Reading is only administered in Grades 3-8. Grade level Istation usage thresholds are contingent on grade: the recommended level in Grades 4-5 is 250 minutes, and 200 minutes in Grades 6-8.

Supplementary analysis of the relationship between Istation usage and STAAR performance

Differences in Adjusted Levels of 2013-14 STAAR-Reading Performance

To investigate whether the inferences from the statistical models that are presented in the main sections of the report were consistent across different methods of measuring student performance, and across different assumptions about how students are selected to use Istation across schools in Texas, the evaluation team fit a number of multivariate regression models. The approach was to build sequentially more complex multivariate models that carry different assumptions about the process by which students choose, or are selected, to participate in the program.

Statistical models and covariates used to supplement the primary analysis presented in the body of the report for this research question are presented below. Each model is fit separately by grade level to capture the variability of the differences in outcomes between participating and non-participating students across grades, but also to account for important differences in how the system is used across grade levels.

- **Model 1A: Levels** of students' STAAR-Reading performance in 2013-14, taking into account their prior achievement, time-invariant (or, minimally variant) attributes such as ethnicity, economic disadvantaged status, and sex, and other pre-treatment (2012-13) covariates, including whether the student used Istation in 2012-13, with campus-level fixed effects to account for observable and unobservable differences across schools.
  - *Outcome variable:* 2013-14 STAAR-Reading scaled score
  - *Measures of program participation*
    - Number of curriculum hours
    - Used at least one minute compared to students who did not use the system (included registered students who did not record a curriculum session)
  - *Method for addressing within-school clustering:* Campus-level fixed effects
  - *Grade levels:* Students enrolled in Grades 4-8 in 2013-14

- **Model 1B:** Identical to Model 1A, with one exception: the evaluation team fit a multilevel model with random effects at the campus level, rather than estimating the campus-level effects directly with fixed effects, to account for non-independence within schools .

Table A3 presents the results for Models 1A and 1B, for each grade level, and for two measures of program participation: 1) whether the student used the system at least one minute, compared to those who did not use Istation; and 2) the total number of minutes spent on curriculum sessions during the respective school year. Only the coefficient on the measure of program participation is presented.

Table A3. Estimated Effects of Istation Usage on STAAR-Reading Test Scores in 2013-14, by Student Grade Level and Measure of Program Participation

		Model 1A			Model 1B		
		B	SE	N	B	SE	N
Grade 4	Curriculum hours	***-0.612****	0.054	335,089	***-0.450****	0.044	331,017
	Used at least one minute	***-6.819****	0.643		***-5.471****	0.539	
Grade 5	Curriculum hours	***-0.632****	0.066	332,264	***-0.396****	0.052	327,859
	Used at least one minute	***-5.779****	0.595		***-4.273****	0.486	
Grade 6	Curriculum hours	***-0.708****	0.075	327,800	***-0.563****	0.067	321,015
	Used at least one minute	***-7.942****	0.643		***-6.117****	0.553	
Grade 7	Curriculum hours	***-0.194***	0.064	335,966	***-0.133**	0.058	331,067
	Used at least one minute	***-4.475****	0.516		***-3.174****	0.456	
Grade 8	Curriculum hours	-0.016	0.088	330,631	0.051	0.077	326,321
	Used at least one minute	***-1.365**	0.626		-0.748	0.526	
Lagged STAAR-Reading Score?		Yes			Yes		
Campus Fixed Effects?		Yes			No		
Campus Random Effects		No			Yes		
Propensity Score Weighting?		No			No		
Outcome Measure		2013-14 STAAR-Reading			2013-14 STAAR-Reading		

Source: Istation session history tables and State of Texas Assessments of Academic Readiness data, 2012-13-2013-14, Texas Education Agency.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) are included in the analyses. Estimates derived from a linear regression. Statistically significant positive coefficients are denoted by bold font, and negative coefficients are denoted by bold and italicized font. * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Curriculum hours are calculated prior to the first administration of the state assessment, which varies by grade level. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Difference-in-Difference Estimation with Propensity Score Reweighting

To investigate the robustness of Model 2, propensity score reweighting was used to balance the non-participant group to achieve observable comparability with the participating group based on a propensity score generated from a logistic regression fit with a vector of student-, school-, and district-level covariates to estimate the conditional probability of using Istation at least one minute during the 2013-14 school year (Table A4). The conditional probability, then, is used to reweight the non-user observations so that the distribution of observed characteristics included in the model used to estimate the probability of using Istation match the distribution of participant students. With the exception of Grade 7, the results are consistent with the unweighted results.

Table A4: Propensity Score Reweighted Effects of Istation Usage on STAAR-Reading Test Score Gains between 2012-13 and 2013-14, by Student Grade Level and Measure of Program Participation

		B	SE	N
Grade 4	Used at least one minute (compared to not at all)	***-0.065****	0.008	331,017
Grade 5	Used at least one minute (compared to not at all)	***-0.047****	0.008	327,859
Grade 6	Used at least one minute (compared to not at all)	***-0.057****	0.007	321,015
Grade 7	Used at least one minute (compared to not at all)	-0.007	0.007	331,067
Grade 8	Used at least one minute (compared to not at all)	**0.026*** **	0.007	326,321
Lagged STAAR-Reading Score		No		
Campus Fixed Effects?		Yes		
Campus Random Effects		No		
Propensity Score Weighting?		Yes		
Outcome Measure		Prior achievement decile-standardized gain score		

Source: Istation session history tables and State of Texas Assessments of Academic Readiness data, 2012-13-2013-14, Texas Education Agency.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) are included in the analyses. Estimates derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. Observations were reweighted by a propensity score derived for a logistic regression model predicting the probability of using the system at least one minute (compared to not at all). Observations are restricted to those in the region of common support. Statistically significant positive coefficients are denoted by bold font, and negative coefficients are denoted by bold and italicized font. * p<0.05, ** p<0.01, *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Technical Material for Econometric Models

Model 1A

Model 1A is an unpooled (by grade level) OLS regression, where the outcome is students' first STAAR-Reading score in 2013-14. Campus-level heterogeneity was accounted for by the inclusion of campus-level fixed effects. The functional form of the model was:

$$\boldsymbol{SSC_{ijg} = \alpha SUCCESS_i + \beta X_{it-1} + \beta X_i + \pi_j + v_{ij}} \qquad \textbf{(1)}$$

Where:

- $\boldsymbol{SSC_{isg}}$ is the 2013-14 STAAR score for student *I*, attending school *j,* in grade *g*
- $\boldsymbol{\alpha SUCCESS_i}$ is an indicator of whether student *i* participated in a SUCCESS program in 2013-14[83]

[83] This term subsumes all variations of the measurement of SUCCESS participation, including total minutes and other binary flavors of participation.

- $\boldsymbol{\beta X_{it-1}}$ is a vector of time-varying student-level characteristics from the previous school year, and time-invariant characteristics, which included
  - 2012-13 (*T-1)* STAAR-Reading score
  - Lagged count of any type of disciplinary actions
  - Lagged attendance rate (2012-13)
  - Lagged indicator of whether student used Istation in the prior year (2012-13)
- $\boldsymbol{\beta X_i}$ is a vector of time-invariant student-level characteristics from 2013-14, which included:
  - Sex
  - Race
  - Current ELL student
  - Test language of 2013-14 STAAR test
  - Economic Disadvantaged status
  - Received any accommodation on the STAAR administration
  - Student was ever retained in grade
  - Student was considered at risk
  - Immigrant indicator
  - Special Education indicator
- $\boldsymbol{\pi_j}$ is a school fixed effect
- and $\boldsymbol{v_{ij}}$ is a random disturbance term

Model 1B

Model 1B is similar to Model 1A, with the exception that the assumption that the unobserved campus-level disturbances are correlated with the regressors in the statistical model is relaxed, which allows for the inclusion of more substantively meaningful fixed campus-level characteristics. The functional form of the model was:

$$\boldsymbol{SSC_{ijg} = \alpha SUCCESS_i + \beta X_{it-1} + \beta X_i + \delta X_i + u_i + \epsilon_i} \qquad \textbf{(1)}$$

Where:

- $\boldsymbol{SSC_{isg}}$ is the 2013-14 STAAR score for student *i*, attending school *j,* in grade *g*

- $\boldsymbol{\alpha SUCCESS_{it}}$ is an indicator of whether student *i* participated in a SUCCESS program in 2013-14[84]
- $\boldsymbol{\beta X_{it-1}}$ is a vector of time-varying student-level characteristics from the previous school year, and time-invariant characteristics, which included
  - 2012-13 (*T-1)* STAAR-Reading score
  - Lagged count of any type of disciplinary actions
  - Lagged attendance rate (2012-13)
  - Lagged indicator of whether student used Istation in the prior year (2012-13)
- $\boldsymbol{\beta X_{i}}$ is a vector of time-invariant student-level characteristics from 2013-14, which included:
  - Sex
  - Race
  - Current ELL student
  - Test language of 2013-14 STAAR test
  - Economic Disadvantaged status
  - Received any accommodation on the STAAR administration
  - Student was ever retained in grade
  - Student was considered at risk
  - Immigrant indicator
  - Special Education indicator
- $\boldsymbol{\delta X_{i}}$ is a vector of school and district-level characteristics, including
  - Education Service Center (ESC) region
  - District type derived from http://goo.gl/gSoiog
  - Title I status in 2013-14
  - Economic Disadvantaged status
  - Campus accountability rating in 2012-13
  - Campus type (e.g., elementary, secondary, or both)
  - Percentage of students who are minority (Hispanic and Black) in 2013-14
  - Percentage of students who are classified as ELL in 2013-14
  - Percentage of students who met the phase-in 1 Level 2 standard in 2012-13

[84] This term subsumes all variations of the measurement of SUCCESS participation, including total minutes and other binary flavors of participation.

  - Percentage of students who are classified as at-risk in 2013-14
- $\boldsymbol{u_i}$ is the between-group (school) disturbance term
- and $\boldsymbol{\epsilon_i}$ is within-group (students within schools) disturbance term

Model 2

Rather than simply looking at cross-sectional levels of student STAAR-Reading performance in 2013-14, the evaluation team fit a two-period DiD model using linear regression, with subject-grade-year and subject-grade-year-prior decile standardized gain scores as the response variable. This model does not account for fixed, unobserved student-level heterogeneity, nor does it account for dynamic selection into the program based on prior performance gains. The model was fit separately for Grades 4-8. The functional form of the statistical model was:

$$\boldsymbol{A_{ijt} - A_{ijt-1} = \alpha SUCCESS_{it} + \beta X_{it-1} + \; + \pi_j + \; v_{ijt}}$$

Where:

- $\boldsymbol{A_{ijt} - A_{ijt-1}}$ is the difference in standardized assessment scores (subject-grade-year-prior decile standardized scores) at time *t,* for student *i*, attending school *j*
- $\boldsymbol{\alpha SUCCESS_{it}}$ is an indicator of whether student *i* participated in a SUCCESS program at time *t*
- $\boldsymbol{\beta X_{it-1}}$ is a vector of time-varying student-level characteristics from the previous school year, and time-invariant characteristics, which included
  - Lagged count of any type of disciplinary actions
  - Lagged attendance rate (2012-13)
  - Lagged indicator of whether student used Istation in the prior year (2012-13)
- $\boldsymbol{\beta X_i}$ is a vector of time-invariant student-level characteristics from 2013-14, which included:
  - Sex
  - Race
  - Current ELL student
  - Test language of 2013-14 STAAR test
  - Economic Disadvantaged status
  - Received any accommodation on the STAAR administration
  - Student was ever retained in grade
  - Student was considered at risk
  - Immigrant indicator
  - Special Education indicator

- $\boldsymbol{\pi_j}$ is a school fixed effect
- and $\boldsymbol{v_{ij}}$ is a random disturbance term

Construction of the Propensity Score and Implementation of the Regression Reweighting Scheme

Following Nichols (2007), the evaluation team used a propensity score reweighting method, where a student's likelihood of being a member of the treatment group, however defined, in 2013-14 is conditioned on a number of pre-treatment school, student, and district covariatesX^C. The conditional probability $\hat{\lambda}$ of being in the treatment group derived from this model is then used to calculate a weight based on the odds$\hat{\lambda}/(1-\hat{\lambda})$.

Because evaluators had several different measures of program participation on which treatment and control groups were balanced based on the conditional probability of being a member of the treatment group, the evaluation team fit the propensity score model using several iterations. This method is described below.

Grades 4-8 in 2013-14 with *at least* two valid test scores

1. This sample was used in Guiding Question 7a and 7C, since only one valid prior test score was required. All valid cases in Grades 4-8 were included.
2. Fit propensity score model estimating the conditional probability of treatment using a logistic regression based on four separate measures of Istation use in 2013-14:
    a. Students who recorded at least once curriculum minute compared to students who did not record any curriculum minutes
    b. Students who met the recommended usage threshold for their grade compared to students who did not
    c. Students who met the recommended usage threshold for their grade compared to students who did not use the system
    d. Students who used the system 500 minutes or more compared to students who did not use the system
3. Vector of student, school, and district covariates included in the functional form included:
    a. 2012-13 (*T-1)* STAAR-Reading score
    b. Lagged count of any type of disciplinary actions
    c. Lagged attendance rate (2012-13)
    d. Lagged indicator of whether student used Istation in the prior year (2012-13)
    e. Sex
    f. Race

g. Grade level

h. Current ELL student

i. Test language of 2013-14 STAAR test

j. Economic Disadvantaged status

k. Received any accommodation on the STAAR administration

l. Student was ever retained in grade

m. Student was considered at risk

n. Immigrant indicator

o. Special Education indicator

p. Education Service Center (ESC) region

q. District type derived from http://goo.gl/gSoiog

r. Title I status in 2013-14

s. Economic Disadvantaged status

t. Campus accountability rating in 2012-13

u. Campus type (e.g., elementary, secondary, or both)

v. Percentage of students who are minority (Hispanic and Black) in 2013-14

w. Percentage of students who are classified as ELL in 2013-14

x. Percentage of students who met the phase-in I Level 2 standard in 2012-13

y. Percentage of students who are classified as at-risk in 2013-14

4. Calculate the conditional odds of being in each treatment group using the formula: $\hat{\lambda}/(1-\hat{\lambda})$
5. Assign a weight of 1 to *all* students in the treatment condition, and a weight equal to $\hat{\lambda}/(1-\hat{\lambda})$ for all students in the comparison group
6. Fit linear regressions for the response variable on the treatment indicator of interest (e.g., used the system at least one minute or used the system 500 minutes or more) applying the following restrictions, weights, covariates:

    a. Restrict analytic sample to only students in the region of common support based on the propensity score model estimated for the respective treatment measure

        i. This is defined, according to Leuven and Sianesi (2003), as cases where the propensity score of the control cases is within the range (minimum and maximum) of the propensity score of the treatment cases.

    b. Include all covariates included in the functional form for estimating the propensity score to achieve double-robustness

c. Apply probability weights using the weight calculated for the respective treatment measures

Guiding Question 7a.2

To address Guiding Question 7a.2, the base model described in Appendix 7a was amended to investigate whether the impact of participating in a SUCCESS program varied across student-level characteristics of substantive interest, including ethnicity, economic disadvantaged status, and ELL status. To investigate the heterogeneity of program participation across these student groups, a multiplicative term ($\boldsymbol{\lambda SUCCESS_{it}ETHNICITY_i}$) between the student's group membership indicator and the measurement of program participation is added to the model's functional form. This is simply the product of multiplying the two terms together, and allows both the identification of *mean* differences across student groups, but also, for continuous measures of student Istation usage, *slope differences*. Put another way: does the association between Istation usage and changes in STAAR-Reading performance differ between ELL students compared to non-ELL students, and does this differ based on the *intensity* of participation?

Technical Material for Econometric Models

More formally,

$$\boldsymbol{A_{ijt} - A_{ijt-1} = \alpha SUCCESS_{it} + \lambda SUCCESS_{it}ETHNICITY_i + \phi ETHNICITY_i + \beta X_{it-1} + \pi_j + v_{ijt}}$$

Where:

- $\boldsymbol{A_{ijt} - A_{ijt-1}}$ is the difference in standardized assessment scores (subject-grade-year-prior decile standardized scores) at time *t,* for student *i*, attending school *j*
- $\boldsymbol{\alpha SUCCESS_{it}}$ is an indicator of whether student *i* participated in a SUCCESS program at time *t*, and represents the mean program effect for students in the base reference category for the student group of interest (here, Hispanic students).
- $\boldsymbol{\lambda SUCCESS_{it}ETHNICITY_i}$ is the multiplicative term between the student group and the measure of program participation, capturing the mean *difference in the program effect* between students in the respective ethnic group who participated in the program relative to students from the base ethnic group who also participated.
- $\boldsymbol{\phi ETHNICITY_i}$ is the average effect of students in the ethnic group category *relative to students in the base category* (here Hispanic students) *who were not in the SUCCESS participant group*.
- $\boldsymbol{\beta X_{it-1}}$ is a vector of time-varying student-level characteristics from the previous school year, and time-invariant characteristics, which included
  - Lagged count of any type of disciplinary actions
  - Lagged attendance rate (2012-13)
  - Lagged indicator of whether student used Istation in the prior year (2012-13)

- $\boldsymbol{\beta X_i}$ is a vector of time-invariant student-level characteristics from 2013-14, which included:
  - Sex
  - Race
  - Current ELL student
  - Test language of 2013-14 STAAR test
  - Economic Disadvantaged status
  - Received any accommodation on the STAAR administration
  - Student was ever retained in grade
  - Student was considered at risk
  - Immigrant indicator
  - Special Education indicator
- $\boldsymbol{\pi_j}$ is a school fixed effect
- and $\boldsymbol{v_{ij}}$ is a random disturbance term

The base, or reference, category for student groups were:

- ELL status: Not ELL
- Economic disadvantaged status: Not economically disadvantaged
- Ethnicity: Hispanic

Models were fit separately by grade and by the student group measure of interest.

Guiding Question 7a.3

The evaluation team supplemented the multivariate models presented in the main section of the report with a quasi-experimental matching estimator using propensity score reweighting.[85] These models, labeled as Methods 4 and 5 in the main body of the report, compare students who used the system at different levels to an reweighted sample of students who *did not use the system at all but whose observable characteristics indicated that they had a similar likelihood of using the system at the corresponding usage level as students who actually used the system at that level*. This provides a more discrete estimate of the relationship between dosage level and STAAR-Reading gain scores, since high dosage students are not being compared to students with a range of exposure and usage of the system. Rather, they are being compared to a group of students who did not use the system, with the composition of the sample of non-users being re-weighted to ensure that, on average, they are observably similar to the students in the respective dosage-level user group.

For the propensity score reweighted model, a regression model with double-robust estimators which incorporate the same covariates used in the propensity score model into the functional form of the regression model for the outcome was fit (Lunceford & Davidian, 2004). Estimates of the effect of dosage on gain scores were robust to additional specifications with the comparison group reweighted by the conditional probability of using the system at each level of dosage (Table A5).[86] Dosage was negatively, and statistically significant associated with gain scores for Grade 4 and 5 students across all measures of usage intensity, and the relationship was positive among Grade 7 and 8 students. Despite their statistical significance, the effects remained small and never exceeded 15% of a standard deviation of the decile-standardized gain scores.

[85] The evaluation team proposed estimating both an unbalanced regression with covariate adjustment—which is the primary method presented and discussed in the main body of the report—as well as a propensity score reweighted model, with weights derived through propensity score matching. We proposed this for robustness, particularly given evidence published in Agodini and Dynarski (2004) and Wilde and Hollister (2007) that propensity score matching performed poorly in recovering impact estimates from experimental data that matched the experimental effects, and the loss of cases with extreme propensity scores that lie outside of the range for which matches were found. In addition, similar complementary approaches have been used in studies of similar educational interventions (Heinrich et al., 2010; Springer et al, 2009).

[86] More information about the derivation of the probability weights is in Appendix A.

Table A5. Propensity Score Reweighted Estimates of the Relationship between Intensity of Istation Usage and STAAR-Reading Decile-standardized Gain Scores, by Grade Level and Measure of Usage Intensity, 2013-14

		B	SE	N
Grade 4	At or above threshold (compared to students with no usage)	***-0.100****	0.013	201,582
	Students with 500 or more minutes compared to students with no minutes	***-0.097****	0.02	155,998
	At threshold (compared to below)	***-0.093****	0.014	201,582
Grade 5	At or above threshold (compared to students with no usage)	***-0.077***	0.013	192,987
	Students with 500 or more minutes compared to students with no minutes	***-0.106****	0.021	134,822
	At threshold (compared to below)	***-0.084****	0.014	192,987
Grade 6	At or above threshold (compared to students with no usage)	***-0.037***	0.014	217,759
	Students with 500 or more minutes compared to students with no minutes	0.013	0.021	190,119
	At threshold (compared to below)	***-0.040***	0.014	217,759
Grade 7	At or above threshold (compared to students with no usage)	**0.062*****	0.014	228,348
	Students with 500 or more minutes compared to students with no minutes	**0.070*****	0.024	211,233
	At threshold (compared to below)	**0.054*****	0.015	228,348
Grade 8	At or above threshold (compared to students with no usage)	**0.103*****	0.016	238,409
	Students with 500 or more minutes compared to students with no minutes	**0.121*****	0.027	226,650
	At threshold (compared to below)	**0.110*****	0.017	238,409
Campus Fixed Effects?	Yes			
Campus Random Effects	No			
Propensity Score Weighting?	Yes			
Outcome Measure	Decile-standardized gain score			

Source: State of Texas Assessments of Academic Readiness and Public Education Information Management System data, 2011-12 – 2013-14, Texas Education Agency. Istation session history table, 2013-14.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) are included the analyses. Hours are calculated prior to the administration of the first assessment, which is contingent on grade level. Estimates derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates, with probability sampling weights based on propensity scores fit for each dosage measure. Observations for each model were restricted to those in the region of common support based on the propensity score. Statistically significant negative coefficients are bolded and italicized, and statistically significant positive coefficients. * p<0.05, ** p<0.01, *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Guiding Question 9

To address Guiding Question 9, the base model described in Appendix A was amended to investigate whether the impact of participating in a SUCCESS program varied between Grade 5 and Grade 8 students who were considered at risk of being retained in grade compared to students who were not at risk of being retained in grade. This was done through the inclusion of an interaction term, between SUCCESS participation and the at-risk for grade retention indicator.

Technical Material for Econometric Models

More formally,

$$\boldsymbol{A_{ijt} - A_{ijt-1} = \alpha SUCCESS_{it} + \lambda SUCCESS_{it} ATRISK_i + \phi ATRISK_i + \beta X_{it-1} + + \pi_j + \upsilon_{ijt}}$$

Where:

- $\boldsymbol{A_{ijt} - A_{ijt-1}}$ is the difference in standardized assessment scores (either subject-grade-year standardized scores or subject-grade-year-prior decile standardized scores) at time *t,* for student *i*, attending school *j*
- $\boldsymbol{\alpha SUCCESS_{it}}$ is an indicator of whether student *i* participated in a SUCCESS program at time *t*, and represents the mean program effect for students in the base reference category for the student group of interest (here, Hispanic students).
- $\boldsymbol{\lambda SUCCESS_{it} ATRISK_i}$ is the multiplicative term between the at-risk flag and the measure of program participation, capturing the mean *difference in the program effect* between students who were at risk and who participated in the program relative to students who were not at risk and who participated in the program. This is the coefficient of substantive interest for identifying student group heterogeneity among students who participated in the program.
- $\boldsymbol{\phi ATRISK_i}$ is the average effect of students who were at risk *relative to students who were not at risk who were not in the SUCCESS participant group*.
- $\boldsymbol{\beta X_{it-1}}$ is a vector of time-varying student-level characteristics from the previous school year, and time-invariant characteristics, which included
  - Lagged count of any type of disciplinary actions
  - Lagged attendance rate (2012-13)
  - Lagged indicator of whether student used Istation in the prior year (2012-13)
- $\boldsymbol{\beta X_i}$ is a vector of time-invariant student-level characteristics from 2013-14, which included:
  - Sex
  - Race
  - Current ELL student

- Test language of 2013-14 STAAR test
- Economic Disadvantaged status
- Received any accommodation on the STAAR administration
- Student was ever retained in grade
- Student was considered at risk
- Immigrant indicator
- Special Education indicator

- $\boldsymbol{\pi_j}$ is a school fixed effect
- and $\boldsymbol{\upsilon_{ij}}$ is a random disturbance term

The base, or reference, category for student groups were:

- ELL status: Not ELL
- Economic disadvantaged status: Not economically disadvantaged
- Ethnicity: Hispanic

Models were fit separately by grade and program participation dosage measure.

Guiding Question 10

For Guiding Question 10, the outcome of interest is whether the student met the passing standard on the second administration of STAAR-Reading *after having failed the first administration.* The outcome is binary, so a multivariate logistic regression was fit, and a measure of SUCCESS usage *between the first and second administration* was calculated and included as a covariate in the model to determine the effect of system dosage on the probability passing the second administration of the respective test. Before models were fit, several business rules were established to determine which students were eligible to be included in the model. They were:

1. Student failed the first administration of the STAAR exam in 2013-14
    a. Only regular STAAR (e.g., no STAAR Modified or Alternate) tests were considered.
2. Student took the second administration STAAR exam in 2013-14
    a. Only regular STAAR (e.g., no STAAR Modified or Alternate) tests were considered.
3. Student was enrolled, and took, the Grade 5 or Grade 8 STAAR exams for the first and second assessment

Technical Material for Econometric Models

A logistic regression with student-, school-, and district-level covariates. Logistic regression estimates the log odds of the outcome (here, passing the second administration of STAAR) as a function of the included covariates. The evaluation team fit a logistic regression, with cluster-adjusted standard errors to account for within-school non-independence, with a logit link function ($\mathrm{n_{ijd}} = \log(\frac{\phi_{ijd}}{1-\phi_{ijd}})$), where $\mathrm{n_{ijd}}$ is the log-odds of student i in school j and district d repeating Grade 5 or Grade 8 in 2014-15. The functional form is formally expressed as

$$\boldsymbol{n_{ijd} = \beta_0 + \beta_1 SUCCESS_{ijd} + \beta_k X_{ijdt-1} + \beta_k Q_{jdt-1} + \beta_k Y_{dt-1} + e_{ijd} + \mu_{jd} + v_d} \quad \textbf{(3)}$$

Where:

- $\boldsymbol{\beta_1 SUCCESS_{ijd}}$is the measure of student *I's* participation in a SUCCESS program in 2013-14
- $\boldsymbol{\beta_k} X_{\boldsymbol{ijdt-1}}$ is a vector of student-level characteristics from the prior year and current year, including
  - Lagged count of any type of disciplinary actions
  - Lagged attendance rate (2012-13)
  - Lagged indicator of whether student used Istation in the prior year (2012-13)
  - STAAR-Reading scaled score from the first administration in March 2013-14
  - Sex
  - Race
  - Current ELL student
  - Test language of 2013-14 STAAR test
  - Economic Disadvantaged status
  - Received any accommodation on the STAAR administration
  - Student was ever retained in grade
  - Student was considered at risk
  - Immigrant indicator
  - Special Education indicator
- $\boldsymbol{\beta_k Q_{jdt-1}}$ is a vector of school-level characteristics
  - Title I status in 2013-14
  - Campus accountability rating in 2012-13
  - Campus type (e.g., elementary, secondary, or both)

- Percentage of students who are minority (Hispanic and Black) in 2013-14
- Percentage of students who are classified as ELL in 2013-14
- Percentage of students who met the phase-in 1 Level 2 standard in 2012-13
- Percentage of students who are classified as at-risk in 2013-14
- Percentage of students classified as economically disadvantaged in 2013-14

▪ $\boldsymbol{\beta_k} Y_{\boldsymbol{dt-1}} \boldsymbol{\pi_{jd}}$ is a vector of district-level characteristics

- Education Service Center (ESC) region
- District type derived from http://goo.gl/gSoiog

▪ $\boldsymbol{e_i}$ is a random error term for student *i* in school *j* in district *d*

Description and Preparation of Istation Files

Istation provided TEA—who, in turn, provided Gibson Consulting Group—four files representing different dimensions of the Istation program: a Usage Summary table,[87] an Assessment table, a Session History table, and a Session-Time-by-Product table. All files captured system usage information between August 2013 and June 2014 for students in Grades 3-8 during the 2013-14 school year. Of primary interest for the evaluation was Istation system usage, which was obtained primarily from the Session History table. The Session History table contains the most granular, lowest-level information about students' system usage. Each row reflects a single, unique Istation session launched by a student, including the date and time the session started, and the duration of the session.

The Session History table was then linked to the Session-Time-by-Product table to obtain more precise information about each unique session, including the types of activities that occurred during the session (e.g., "Assessment", "Other", or "Curriculum"). In addition, the Session-Time-by-Product table disaggregates usage duration by activity type, which allowed the research team to separate usage associated with curriculum—of which is of primary interest for the evaluation—from other types of activities. Roughly 100% (99.73) of records from the Session History table were linked to a corresponding session in the Session-Time-by-Product table.[88]

---

[87] Upon closer inspection of the Usage Summary table, the research team decided against using it for the evaluation. The data are not collected at the session-day level; rather, login and usage statistics are aggregated at the month level, by the type of activities that occurred within the respective month which does not allow the research team to report usage patterns and frequency with the precision required for the evaluation.

[88] Only records which were successfully matched between the two tables were retained. This resulted in a loss of less than one quarter of 1% of approximately 24 million records. We surmise that the small number of unmatched records was due to the records that were dropped during the initial cleaning phase that was conducted separately for each table.

The Assessment table contains Istation Indicators of Progress-Early Reading (ISIPER, for Grades K-3) and Istation Indicators of Progress-Advanced Reading (ISIPAR, for Grade 4-8) scores for each assessed domain subtest for each assessment taken by students during the 2013-14 school year. Only records containing students' *Overall* score, which measures their overall reading ability, and is derived from their performance on the composite subtests (Mathes, 2014), was retained. Next, for each student, and for each month, only students' *last* ISIP assessment that matched the student's grade level in 2013-14 was kept.[89] The reduced Assessment table was joined to their session-day usage data using a unique student identifier and the month in which a student took the assessment, linking students' last monthly ISIP assessment scores. Nearly 95%of all students' session-day level records were linked to at least one assessment that occurred during the respective month.

For each table, several business rules were developed and implemented to remove anomalous records from the Istation system usage and assessment file.

1. Assessment records
    a. Inclusion rules:
        i. Only the last assessment taken within each month was retained.
        ii. For the evaluation, only *Overall* scores were used, and the others were discarded.
        iii. Only assessments that aligned with students' grade level were retained. This means that assessments for Grade 3 students who took an ISIPAR were removed, and ISIPER assessments for students in Grades 4-8 were removed.
        iv. Assessments with an unreasonable date stamp (e.g., not in 2013 or 2014) were removed. This affected less than 1% of all records.
2. Session History records
    a. Inclusion rules:
        i. Sessions with an unreasonable date stamp (e.g., not in 2013 or 2014) were removed. This affected less than 300 out of 23 million records.
        ii. Sessions with anomalous durations (greater than 480 minutes) were removed.
        iii. Only sessions identified as Curriculum were retained. The type of session was obtained from the Session-Time-by-Product table.

[89] This decision is based on the recommendation of Istation personnel provided in a personal communication on Monday, June 2nd, 2014.

iv. Only records were in both the Session History and the Session-Time-by-Product tables were retained. This eliminated approximately one third of 1% of all records (64,141 out of 23,547,513 records).

Linking Istation Files to TEA's Unique Student Identifier

The Istation Reading tables that Gibson received contained a unique student database key that could not be joined to TEA administrative and assessment records. Beginning in the 2013-14 school year, Istation was required to record students' Unique ID identifier (or TSDS number) which would permit Istation system usage and assessment performance data to be linked to TEA accountability and assessment records, including student attendance, economic disadvantaged status, and STAAR test performance. Istation provided records for all students who were registered in the system with this identifier to TEA, along with other personally identifiable information contained within the system, including students' first and last name. These records were then linked back to TEA's unique state identifier, producing a match rate of 99.6%.[90]

These linked records were then delivered to Gibson, and were subsequently linked (using the Istation system identifier IUSER_OID) to the base Gibson base Istation usage file that contained students' session and assessment history for the 2013-14 school year. Of the 1,428,484 unique students with at least one login session or assessment attempt, 1,412,758 (98.6%) were successfully linked to the files containing the TEA unique state identifier. About 1% (15,726) of students were not found in the linked file TEA provided and were, consequently, removed from the analytic file.

The linked file that TEA provided to Gibson contained students who were *registered* in the Istation system, not necessarily students who *used* the system, including those who took an assessment or who took an assessment and recorded a curriculum session. This is an important distinction. Some schools, or even districts, batch uploaded entire rosters of students into the Istation system and many of these students did not record a single curriculum session, or take at least one assessment during the 2013-14 school year. The frequency counts for students falling into different usage categories are reported in the body of the evaluation report. (See Table A6.)

[90] Even though the match rate was 99.6%, this reflects some matches where there was a varying degree of confidence in the quality of the match between Istation records and TEA records. TEA staff created an indicator that quantified the quality of the match based on first and last names, as well as possible nicknames. Only records for which there was a high degree of confidence in the match were retained. This resulted in a loss of approximately 1.2% of students, or 26,634 individual students.

Table A6. Count of Unique Student Records in Istation Assessment and Usage and Registration Files, 2013-14 School Year, Grades 3-8

Description	Frequency
Total number of unique students in the combined Istation Assessment and Session History file	1,428,484
Total number of unique students in the combined Istation Assessment and Session History file that could be linked to TEA files	1,412,758
Total number of unique students in the Istation registration file based on students' IUSER_OID from Istation	2,119,315
Total number of unique students in the Istation registration file based on students' Unique ID (i.e., TSDS number)	2,050,170
Total number of unique students in the registration file after removing potentially inaccurate matches and Istation records that could not be linked to the Istation	2,028,527
Total number of unique students in the registration file that were linked to the PEIMS 101 Fall 2013 Snapshot reported in Table 2.1	1,993,693

Appendix B – TTM Technical Section

Research Question 7b.1

Outcome Measure Operationalization and Calculation

The evaluation team approached standardization in two ways. First, a simple gain score from standardized STAAR-Mathematics scores in 2012-13 and 2013-14 was calculated. The standardization was done by grade level and school year. The standardized scores have a mean of zero and a standard deviation of 1, in each grade and school year, so that a positive score indicates that a student's test score was above the mean in that grade and year, and a negative score indicates a student was below the mean in that grade and year. The evaluation team calculated a gain score by simply subtracting the 2012-13 standardized test score from the 2013-14 standardized test score.

Next, evaluators constructed an alternative measure of a standardized gain score based on the work of Hanushek et al. (2005),[91] and closely related to the approach proposed by Reback (2008), which is increasingly used in analyses of changes in test scores across time.[92] The evaluation team calculated a decile-standardized gain score that standardizes changes in student performance, across years and grades, by their prior test performance decile.[93] Interpretation is similar to the one provided for standardized gain scores in the preceding paragraph, with one exception: now, standardized gains are relative to students who were in the same prior test performance decile.

This standardization has a number of properties that make it a desirable operationalization of student performance changes across time. First, because student gain scores are standardized by their placement in the distribution of scores on the prior year test, the measurement is less sensitive to mean reversion and test ceiling effects (Hanushek et al., 2005; Reback, 2008). This is particularly important given evidence that student test scores are volatile across time, and that students at the bottom end of the test score distribution tend to exhibit larger gains between T-1 and T than students at the upper portion of the distribution (Hanushek et al., 2005). Second, in the multivariate models fit in the second part of this section, standardizing gains by prior performance decile obviates the need to control for prior performance in the statistical models. Below, and in the multivariate sections, results using both measures are presented to ensure the stability of the results across measurement choice.

[91] See Peterson and Chingos (2009) for an application of a similar technique in an evaluation of an educational intervention.

[92] See for instance Taylor et al. (2014) for a recent application of a comparable approach in Texas.

[93] See Appendix A for more details about this calculation.

Standardized Gain Scores

For each school year, and grade level, the evaluation team calculated standardized gain scores. First, calculated this term by first standardizing student test scores, by subject, grade, and year, and subtracted the standardized score at time t from standardized scores at time t-1. Positive scores indicate gains above the mean for students in the respective subject, grade, and year, while negative scores indicate that students' standardized test score was below the mean for students in the subject, grade, and year (Zimmer et al., 2007; Springer et al., 2009; Heinrich et al., 2010).

The method for calculating this measure is as follows:

1. First calculate, for each subject, year, and grade, the average scaled score (represented below by SSC) at time *T* and subtract this from each student *i*'s actual score in subject *s,* grade *g,* at time *t*.
   a. $\mu_{istg} = SSC_{istg} - \overrightarrow{SSC_{stg}}$
2. Next, divide μ_{stg} by the standard deviation (σ_{sgt}) of the scaled score to produce the subject-grade-year standardized test score:
   a. $STD_{istg} = \frac{\mu_{istg}}{\sigma_{sgt}}$
3. Standardized gain scores were calculated by simply subtracting each student's standardized test score at time *T-1* STD_{it-1} from their standardized score (STD_{it}) at time *T.*

Decile-Standardized Gain Scores

Next, evaluators constructed an alternative measure of a standardized gain score based on the work of Hanushek et al. (2005),[94] and closely to the approach proposed by Reback (2008).[95] The evaluation team calculated a decile-standardized gain score which standardizes changes in student performance, across years and grades, by their prior test performance decile.[96] Decile-standardization has a number of properties which make it a desirable operationalization of student performance changes across time.

First, because student gain scores are standardized by their placement in the distribution of scores on the prior year test, the measurement is less sensitive to mean reversion and test ceiling effects (Hanushek et al., 2005; Reback, 2008). This is particularly important given evidence that student test scores are volatile across time, and that students at the bottom end of the test score distribution tend to exhibit larger gains between *T-1* and *T* than students at the upper portion of the distribution (Hanushek et al., 2005). Second, in the multivariate models fit in the second part of this section, standardizing gains by prior performance decile eliminates the need to control for prior performance in the statistical models.

---

[94] See Peterson and Chingos (2009) for an application of a similar technique in an evaluation of an educational intervention.

[95] See for instance Taylor et al. (2014) for a recent application of a comparable approach in Texas.

[96] See Appendix A for more details about this calculation.

To account for students placement in the prior-score distribution, which may determine how much students' scores change across years, and to account for possible mean reversion, the evaluation team implemented a gain score standardization method that standardizes test scores gains by subject, year, grade, and student's placement in the prior-year achievement distribution. The method is identical to the one proposed in Hanushek et al. (2005), and the notation and description of the calculation procedure below closely follows theirs.

1. First, for every subject, grade, and year (where years here include 2012-13 and 2013-14), divide the prior-year (2011-12 for students in 2012-13, and 2012-13 for students in 2013-14) subject and grade test score distribution into ten equal score intervals($c_m \; where \; m = 1, \dots, 10)$
2. Next, for each year (again, 2012-13 and 2013-14), calculate the mean and standard deviation of the *gains* for all students in each subject, grade, year, and *decile* interval with a test score at *T-1* (A_{isg-1}) and *T* (A_{isg}), using the notation from Hanushek et al. (2005):
   a. $\mu_{sgt}^{c_m} = \overrightarrow{A_{\imath sg} - A_{\imath sg-1}}$
   b. $\sigma_{sgt}^{c_m} = \sqrt{\Sigma((A_{isg} - A_{isg-1}) - \mu_{sgt}^{c_m})/n_{sgtc_m}}$
3. Then, the decile-standardized gain score for each student in interval c_m in each subject, year, and grade is:
   a. G_{isgt}=$[(A_{isg} - A_{isg-1}) - \mu_{sgt}^{c_m}]/\sigma_{sgt}^{c_m}$

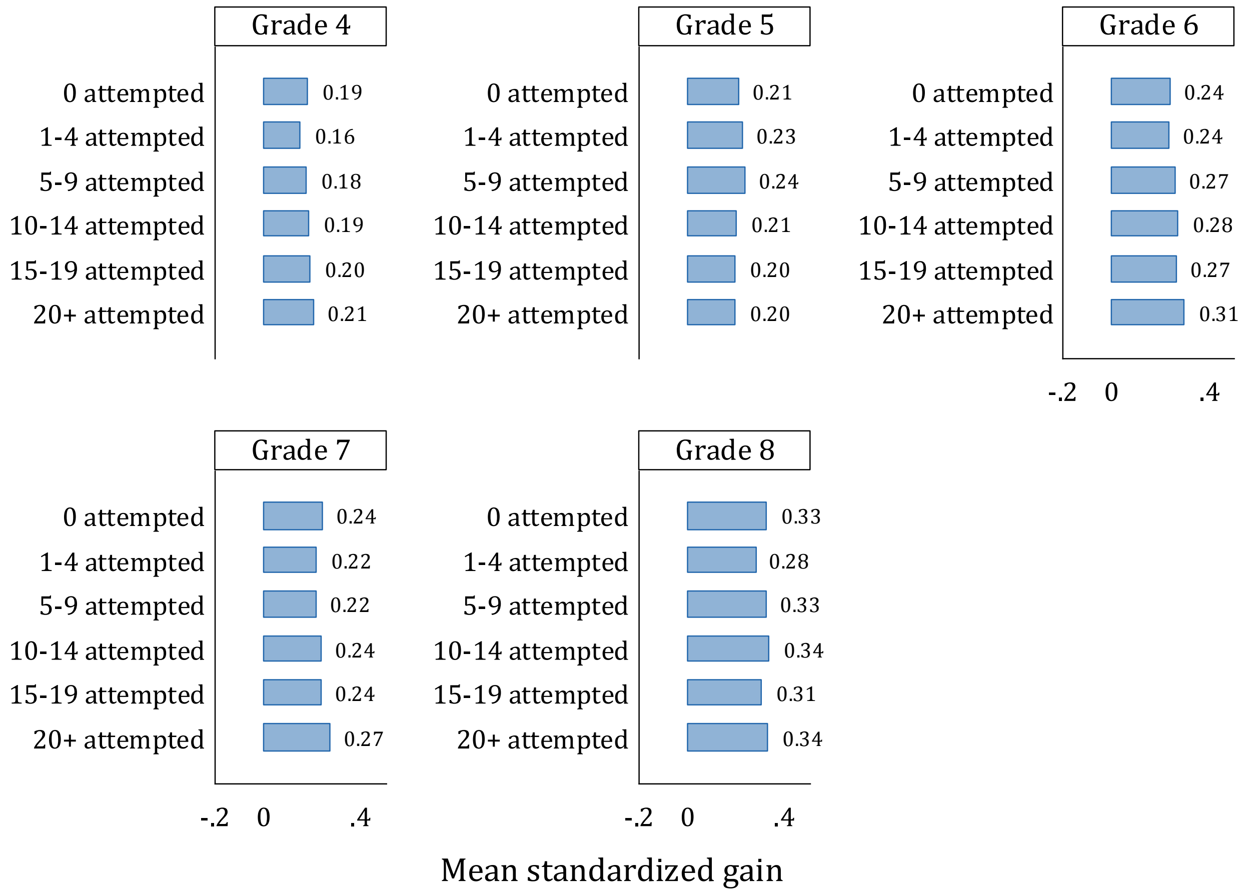
Figures B1, B2, and B3 show the average standardized and decile-standardized gain scores between 2012-13 and 2013-14. The evaluation team calculated the average gain score by grade level and TTM lesson usage in 2013-14. In Figures B1 and B2, which includes the standardized gain score not related to prior performance levels, the evaluation team added an additional dimension to the figure to account for difference performance levels on STAAR-Mathematics in 2012-13 between students across TTM usage categories. Figure B1 includes students who were in the bottom quartile of performance on their prior year assessment (1st to 25th percentile), and Figure B2 includes students in the second quartile (25th to 50th percentile). Finally, for the figures that follow, the evaluation team unpacked the TTM threshold of 10 or more lessons to include more categories with a bin width of five in order to explore how student performance gains might be associated with higher thresholds of TTM usage (e.g., fifteen to nineteen lessons or twenty or more lessons attempted).[97]

Descriptive results show a number of patterns. First, the importance of standardizing gain scores by students' prior performance is clear. Between Figures B1 and B2, the difference in the average grade-year standardized gain scores was much larger among students the bottom quartile of 2012-13 STAAR-Mathematics performance (Figure B1) compared to students in the second quartile (Figure B2),

[97] The analysis focused on lessons attempted because threshold metrics provided by TTM were based on lessons attempted as opposed to passed. Lessons attempted also are a more accurate measure of effort put forth by the student than lessons passed.

particularly in Grades 4 through 6. Second, except for students who used the system twenty or more times, second quartile students in Grade 6 experienced negative gains, but all gains for those students were smaller than gains for bottom quartile of Grade 6 students. Third, Figure B2 shows that, among second quartile students, gains for students attempting more than twenty lessons were strongest, particularly in Grade 7.

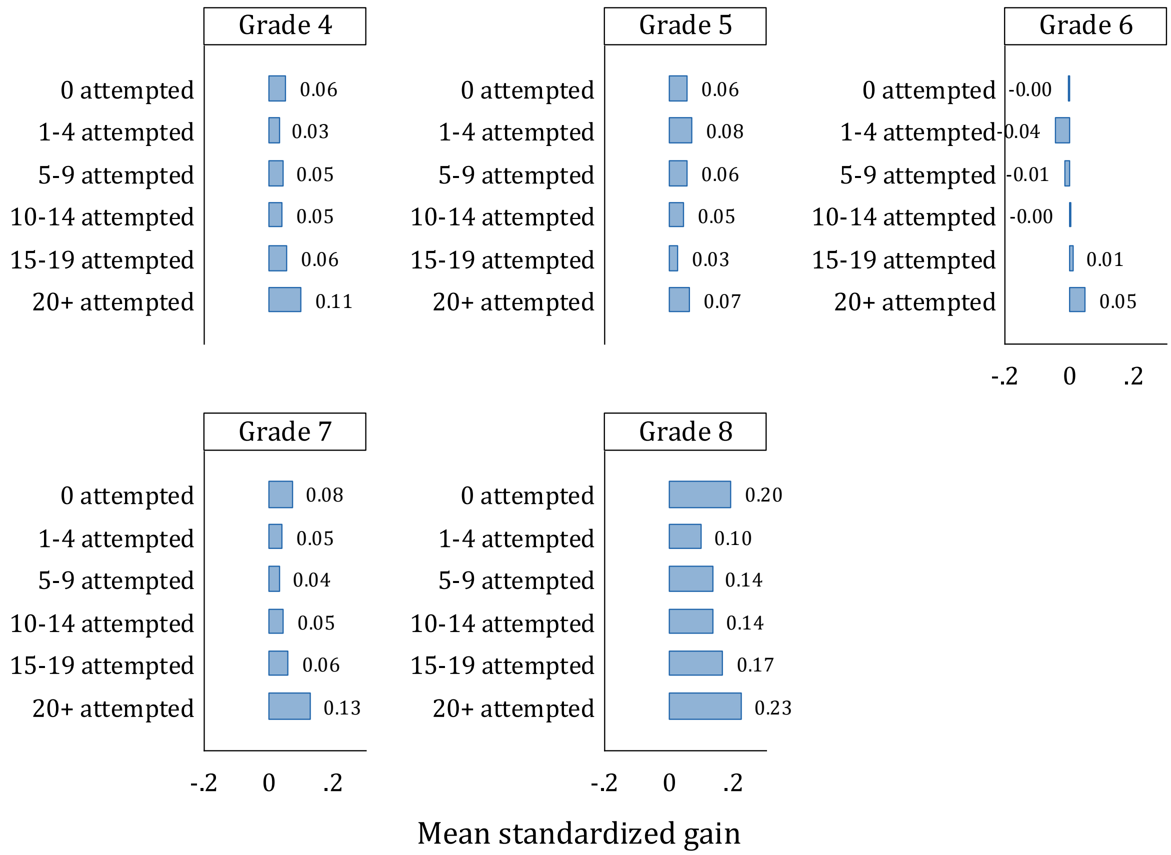
Figure B1. Mean Grade and Year Standardized Gain Score between 2012-13 and 2013-14 for Students in the Bottom Quartile of Scores on the 2012-13 STAAR-Mathematics Assessment, by Grade Level and Number of TTM Lessons Attempted, 2013-14



Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions).

Figure B2. Mean Grade and Year Standardized Gain Score between 2012-13 and 2013-14 for Students in the Second Quartile of Scores on the 2012-13 STAAR-Mathematics Assessment, by Grade Level and Number of TTM Lessons Attempted, 2013-14

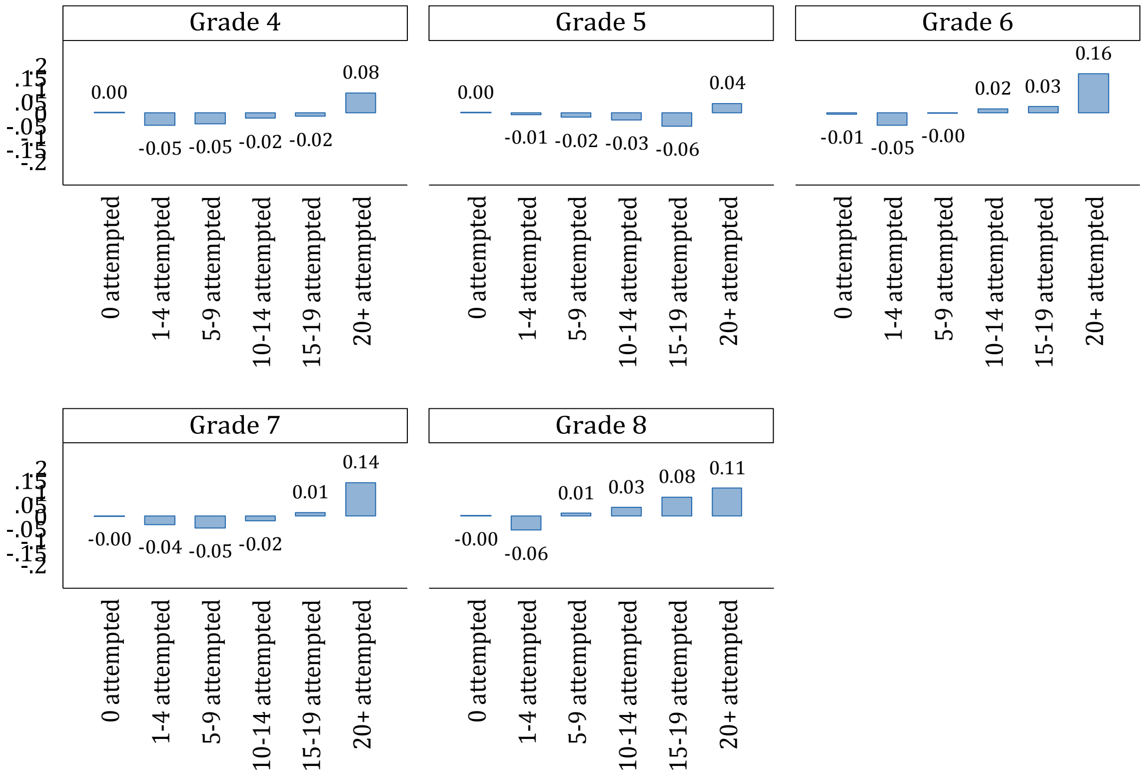


Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions).

In Figure B3, prior performance was adjusted with more precision than prior figures because it accounts for students' prior performance level (prior achievement decile). With this adjusted performance metric, differences between non-users and TTM users become more apparent. That is, students who used TTM more had higher gains relative to their prior performance decile. This is particularly true for students who used the system for more than twenty lessons. In terms of less intensive usage, Grades 6 and 8 showed strongest gains with less exposure to lessons. In Grade 6, slight gains were shown with 10 or more attempted lessons and in Grade 8, gains started at five or more lessons attempted.

Figure B3. Mean Decile-Standardized Gain Scores between 2012-13 and 2013-14, by Grade Level and Number of TTM Lessons Attempted, 2013-14



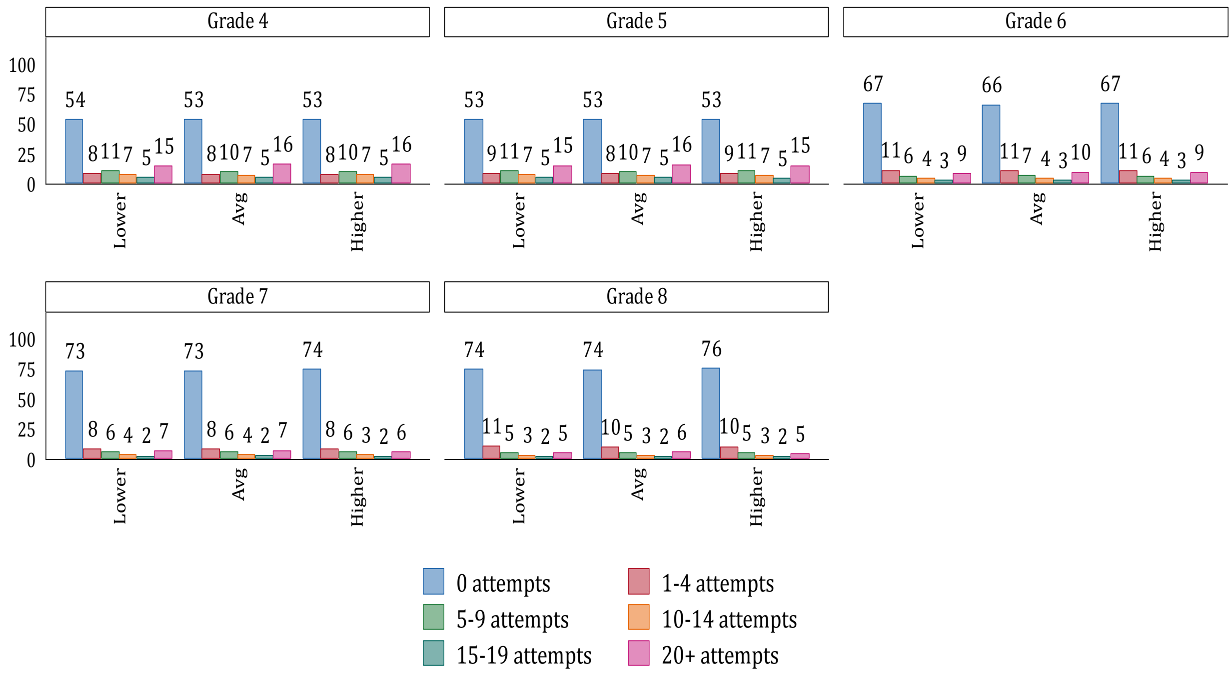
Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions).

The next descriptive analysis examined how student STAAR-Mathematics performance was associated with TTM usage in a different way: the evaluation team calculated whether the student was above, below, or on par with their prior STAAR-Mathematics achievement decile (called 'performance tiers') and shows those students by their TTM usage. The idea here was to see if students who were higher or lower than average were exposed to a higher or lower number TTM lessons, and thus, to explore descriptively whether prior performance may have influenced students' TTM usage.

Figure B4 shows the percentage of students who attempted TTM lessons by student performance tier from the prior year. Performance tiers are defined as students whose performance in 2013-14 was one standard deviation or more below their mean prior achievement decile (*Lower*), one standard deviation or more higher than their mean prior achievement decile (*Higher*), or within one standard deviation above or below their mean prior achievement decile (*Avg*). Generally, these results do not suggest that students were identified for TTM usage differently based on their prior STAAR-Mathematics results.

Figure B4. Percentage of Students Who Attempted TTM Lessons, by Grade and Number of Lessons Attempted, 2013-14



Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions). Performance tiers are defined as one standard deviation or more lower than their mean prior achievement decile (Lower), one standard deviation or more higher than their mean prior achievement decile (Higher), or students within one standard deviation above or below their mean prior achievement decile (Avg).

Inclusion rules for schools and students in the analytic sample

For the multivariate models of student outcomes, the evaluation team restricted the analytic sample based on a number of rules. Those were:

1. Only campuses where at least one student had attempted TTM in 2013-14, to minimize incomparability across different settings.
2. Students who progressed consecutive grades consecutively between 2012-13 or 2013-14, depending on the multivariate model.
3. Students who took grade-level STAAR-Mathematics assessments that matched their grade level in PEIMS. Thus, students in Grade 8 who took an above-grade standardized assessment—such an End of Course (EOC) exam, were omitted from the multivariate analysis to standardized the assessments that are being used to compare performance differences between Istation users and non-users.

4. Only students with a valid test score on the first STAAR administration, and who took the regular English and Spanish versions (i.e., not the modified or alternate versions).
5. Only students with at least two valid STAAR-Mathematics test scores in 2012-13 and 2013-14, since gain scores could not be calculated without both a prior and current test score.
6. Only students in Grades 4-8 in 2013-14.

Analytic Sample Descriptions

Table B1 adds to the descriptive statistics provided in this chapter about student TTM usage and STAAR-Mathematics scores. The frequency counts and means for 2011-12 through 2013-14 STAAR-Mathematics tests are presented, along with standardized and unstandardized gain scores on STAAR-Mathematics. These calculations were disaggregated by students' system usage, and by whether students had two or three valid test scores between 2011-12 and 2013-14. It is relevant to the analyses that follow to note that gain scores and prior achievement were not substantially nor higher for students who were TTM users than students who did not use the TTM system. In addition to the descriptive analyses provided earlier (particularly Figure B4), this table demonstrates that targeting students for TTM was not strongly tied to their prior standardized test performance.

Table B1. Summary Statistics for Student Test Scores, Standardized and Unstandardized Gain Scores, and Prior Performance, by TTM Usage Measure and Grade Level, 2013-14

Grade Level	Usage Measure	Decile Standardized Gain Score	Standardized Gain Score	Unstandardized Gain Score	2013-14 STAAR-Mathematics Scaled Score	2012-13 STAAR-Mathematics Score	2011-12 STAAR-Mathematics Scaled Score	Number of Unique Students
Analytic Sample 1: Two Valid STAAR-Mathematics Test Scores								
Grade 4	Did not use	-0.01	-0.02	84.66	1,550.01	1,465.30	1,292.62	154,538.00
Grade 4	Used at least once	0.00	-0.02	84.08	1,555.74	1,471.77	1,296.39	155,927.00
Grade 5	Did not use	-0.01	-0.02	65.28	1,603.44	1,538.69	1,464.58	156,365.00
Grade 5	Used at least once	-0.01	-0.01	66.67	1,601.45	1,535.09	1,461.27	156,541.00
Grade 6	Did not use	-0.03	-0.02	29.82	1,623.68	1,595.93	1,539.65	189,551.00
Grade 6	Used at least once	0.03	0.03	37.62	1,622.98	1,588.42	1,532.80	109,604.00
Grade 7	Did not use	-0.02	0.06	16.43	1,623.57	1,617.86	1,591.37	200,911.00
Grade 7	Used at least once	0.02	0.09	23.14	1,613.58	1,600.40	1,577.63	83,531.00
Grade 8	Did not use	-0.01	0.15	61.59	1,670.18	1,627.10	1,626.67	172,055.00
Grade 8	Used at least once	0.02	0.18	65.92	1,657.91	1,612.29	1,609.45	67,054.00

Source: State of Texas Assessments of Academic Readiness data, 2011-12 to 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions). Counts and summary statistics only include students who met the following sample inclusion rules: Only campuses where at least one student was *registered* for TTM in 2013-14; students who progressed consecutive grades consecutively between 2012-13 and 2013-14; students who took grade-level STAAR-Mathematics assessments that matched their grade level in PEIMS.

Supplementary analysis of the relationship between TTM usage and STAAR performance

Differences in Adjusted Levels of 2013-14 TTM Performance

The evaluation team classified and described the statistical models and covariates used to address this research question below.[98] Each model was fit separately by grade level to capture the variability of the differences in outcomes between participating and non-participating students across grades, but also to account for important differences in how the system was used across grade levels.

- **Model 1A: Levels** of students' STAAR-Mathematics performance in 2013-14, conditioned on prior achievement, time-invariant (or, minimally variant) attributes such as ethnicity, economic disadvantage status, and sex, and other pre-treatment (2012-13) covariates, including whether the student used TTM in 2012-13, with campus-level fixed effects to account for observable and unobservable differences across schools.
    - *Outcome variable:* 2013-14 STAAR-Mathematics score
    - *Measures of program participation*
        - Attempted at least one TTM lesson compared to students who did not use TTM
    - *Method for addressing within-school clustering:* Campus-level fixed effects
    - *Grade levels:* Students enrolled in Grades 4-8 in 2013-14
- **Model 1B:** Identical to Model 1A, with one exception: the evaluation team fit a multilevel model with random effects at the campus level, rather than estimating the campus-level effects directly with fixed effects, to account for non-independence within schools.
- **Model 2**: To assess differences in student test scores, the evaluation team estimated a DiD framework where within-student changes in standardized gain scores were compared between students who participated in the program in 2013-14 to those who did not use the program. Model 2 included decile-grade-year standardized gain scores.
    - *Outcome variables*: decile standardized gain scores, between 2012-13 and 2013-14
    - *Measures of program participation:* Attempted at least one TTM lesson compared to students who did not use TTM
    - *Method for addressing within-school clustering*: Campus-level fixed effects
    - *Grade levels:* Students enrolled in Grades 4-8 in 2013-14 and Grades 3-7 in 2012-13
    - *School years:* 2012-13 and 2013-14

[98] The formal description of these models can be found in Appendix B.

Model Results for Relationships between TTM Use and 2014 Mathematics STAAR Performance

Table B2 presents the results for Models 1A, and 1B, and Table B3 presents the results for Model 2, for each grade level, and program participation defined as the student using the system for at least one lesson and number of lessons attempted. The coefficient on the number of lessons attempted categories provides a measure of program usage intensity or dosage. Only coefficients on measures of program participation are presented in these tables.

Recall that the evaluation team fit a number of different statistical models with varying underlying assumptions about how students are selected to use TTM in order to assess the robustness of the statistical models. Basic information about the controls and outcome measure used in each model is placed at the bottom of each table. To account for differences between participating and non-participating students, the evaluation team adjust for an array of time-varying, student-level pre-treatment covariates, as well as stable student-level covariates (such as ethnicity and gender). And, when campus-level differences were modeled with random effects, the evaluation team adjust for campus-level attributes that may confound the association between participation and student outcomes. These include, but are not limited, to Title I status, region, and district urbanicity.[99]

The TTM usage metric provided in Table B2 is the categorical dosage or intensity metric for number of attempts, with the reference or base for usage categories being students who did not use TTM (zero attempts). Echoing the descriptive results presented previously, these results show that there are thresholds at which there are significant, positive increases in standardized test performance. In Models 1A and 1B (Table B2), which use STAAR-Mathematics in 2013-14 as the outcome measure, TTM users who used the system more frequently—particularly those that attempted 20 or more lessons in the system—had STAAR-Mathematics scaled scores that were statistically significant higher than non-users. The score increases associated with TTM use ranged from 1.9 to 20 scaled score points for Model 1A, and an increase of 1.3 to 19.5 scaled score points for Model 1B, compared with students who did not use TTM in 2013-14. In considering the range of use among students who attempted 20 or more lessons, it is important to note that this category does contain a wider range of usage (up to nearly 500 attempts). However, aside from TTM use, this group was not visibly different than students in other categories of TTM use, based on observable characteristics. That is, the size of the group was not much larger than other categories, and average performance and gains on STAAR-Mathematics tests were not materially nor systematically different than students in the other categories. While there may have been other characteristics, not captured by the model, that explain how students in the 20+ category were different from other students in ways that may have impacted their mathematics achievement, it does appear that there was a statistically measurable increase in scores for students who reach this threshold of TTM use.

Other associations between TTM use and mathematics outcomes followed different patterns, particularly at low use categories. For example, in all grades except Grade 7 (Model 1A), students who used TTM one to four times had small decreases in test scores as compared with students who did not use TTM; however, students who used it more than 15 times experienced significant increases in performance

[99] A full list of the covariates included in each model is provided in Appendix B.

(about 3.9 scale score points for 15-19 attempts and 15.4 scale score points for 20 or more attempts). The largest increases associated with higher thresholds of TTM usage were in Grade 6. In some grades, like Grade 5, statistically significant, positive increases were seen with as little as five or more or 10 or more attempts (again mirroring descriptive analysis findings). Importantly, while significance tests for the usage metric below compares TTM categories of use against a base of no TTM lessons attempted, testing coefficients among other use categories shows that thresholds of 10 or above were all statistically significantly different from one another.[100]

Table B2. Estimated Effects of TTM Usage on STAAR-Mathematics Test Scores in 2013-14, by Student Grade Level and Measure of Program Participation

		Model 1A			Model 1B		
Grade Level	**System Usage Measure**	**B**	**SE**	**N**	**B**	**SE**	**N**
Grade 4	Used at least once	2.720***	0.781	305,921	2.131**	0.711	309,328
	1-4 attempts	-5.746***	0.948		-6.297***	0.907	
	5-9 attempts	-1.660	0.910		-2.566**	0.851	
	10-14 attempts	1.966*	0.974		0.851	0.913	
	15-19 attempts	3.982***	1.062		2.765**	0.995	
	20+ attempts	15.688***	0.984		14.225***	0.904	
Grade 5	Used at least once	2.013**	0.700	308,078	1.428*	0.642	306,986
	1-4 attempts	-1.700*	0.828		-2.232**	0.794	
	5-9 attempts	-0.310	0.830		-1.056	0.783	
	10-14 attempts	0.353	0.918		-0.500	0.863	
	15-19 attempts	0.705	0.975		-0.206	0.919	
	20+ attempts	9.849***	0.924		8.814***	0.854	
Grade 6	Used at least once	4.658***	0.843	294,157	4.550***	0.805	292,952
	1-4 attempts	-3.224***	0.886		-3.460***	0.865	
	5-9 attempts	2.898**	1.080		2.560*	1.047	
	10-14 attempts	6.460***	1.228		6.114***	1.194	
	15-19 attempts	8.397***	1.327		7.780***	1.294	
	20+ attempts	19.959***	1.295		19.361***	1.239	
Grade 7	Used at least once	3.743***	0.651	279,451	3.737***	0.625	280,576
	1-4 attempts	-0.285	0.766		-0.248	0.751	
	5-9 attempts	0.618	0.843		0.484	0.828	
	10-14 attempts	3.410***	0.923		3.234***	0.899	
	15-19 attempts	5.997***	1.096		5.845***	1.073	
	20+ attempts	13.177***	1.005		12.928***	0.978	
Grade 8	Used at least once	0.681	0.811	227,056	0.565	0.769	228,851
	1-4 attempts	-3.438***	0.908		-3.480***	0.880	

[100] That is, using a Wald composite hypothesis test, the null hypothesis that coefficients across these categories are the same could be rejected (at a p-value of 0.05).

		Model 1A			Model 1B		
Grade Level	**System Usage Measure**	**B**	**SE**	**N**	**B**	**SE**	**N**
	5-9 attempts	1.523	1.068		1.259	1.036	
	10-14 attempts	3.169*	1.241		2.827*	1.210	
	15-19 attempts	5.700***	1.386		5.391***	1.354	
	20+ attempts	7.535***	1.356		7.127***	1.311	
Lagged STAAR-Mathematics Score?		Yes			Yes		
Campus Fixed Effects?		Yes			No		
Campus Random Effects		No			Yes		
Propensity Score Weighting?		No			No		
Outcome Measure		2013-14 STAAR-Mathematics			2013-14 STAAR-Mathematics		

Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2012-13-2013-14, Texas Education Agency.

Note: Reference category for system usage was zero attempts. Only regular English and Spanish versions (i.e., not modified or alternate versions) were included in analyses. Number of attempts were prior to the first STAAR administration, contingent on grade level. Estimates were derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. Statistically significant positive coefficients are denoted in bold, and negative coefficients are denoted by bold and italicized font. *p<0.05, **p<0.01, ***p<0.00. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

To investigate the robustness of Model 2, propensity score reweighting (i.e., matching) was applied to balance the non-participant group to achieve comparability with the participating group (Table B3). The results were consistent with the unweighted results provided in Table 3.24. Across all grades, students who used TTM had higher decile-standardized gain scores compared to students who did not use the system.

It is important to note some intrinsic features of the usage and intensity metrics used for TTM in these models. TTM usage was manifested as the number of lessons a student attempted in a defined period of time. This measure directly quantified students' exposure to the content and assessments that comprise lessons within the system with significant precision.

Table B3. Propensity Score Reweighted Effects of TTM Usage on STAAR-Mathematics Test Score Gains between 2012-13 and 2013-14, by Grade

		Model 3		
Grade Level	**System Usage Measure**	**B**	**SE**	**N**
Grade 4	Used at least once	0.030***	0.006	310,751
Grade 5	Used at least once	0.021**	0.006	308,616
Grade 6	Used at least once	0.042***	0.006	294,677
Grade 7	Used at least once	0.040***	0.006	282,209
Grade 8	Used at least once	-0.017*	0.007	230,021
Lagged STAAR-Mathematics Score?			No	
Campus Fixed Effects?			Yes	
Campus Random Effects			No	
Propensity Score Weighting?			Yes	
Outcome Measure		Prior achievement decile-grade-year standardized gain score		

Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2012-13-2013-14, Texas Education Agency.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) are included in the analyses. Number of attempts are prior to the administration of the first assessment, which is contingent on grade level. Estimates derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. Statistically significant positive coefficients are denoted by bold font, and negative coefficients are denoted by bold and italicized font. * p<0.05, ** p<0.01, *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Technical Material for Econometric Models

Model 1A

Model 1A is an unpooled (by grade level) OLS regression, where the outcome is students' first STAAR-Mathematics or score in 2013-14. Campus-level heterogeneity was accounted for by the inclusion of campus-level fixed effects. The functional form of the model was:

$$\boldsymbol{SSC_{ijg} = \alpha SUCCESS_i + \beta X_{it-1} + \beta X_i + \pi_j + \upsilon_{ij}} \qquad \textbf{(1)}$$

Where:

- $\boldsymbol{SSC_{isg}}$ is the 2013-14 STAAR score for student *i*, attending school *j,* in grade *g*
- $\boldsymbol{\alpha SUCCESS_i}$ is an indicator of whether student *i* participated in a SUCCESS program in 2013-14[101]
- $\boldsymbol{\beta X_{it-1}}$ is a vector of time-varying student-level characteristics from the previous school year, and time-invariant characteristics, which included

[101] This term subsumes all variations of the measurement of SUCCESS participation, including total minutes and other binary flavors of participation.

- 2012-13 (*T-1)* STAAR-Mathematics score
- Lagged count of any type of disciplinary actions
- Lagged attendance rate (2012-13)
- Lagged indicator of whether student used TTM in the prior year (2012-13)

- $\boldsymbol{\beta X_i}$ is a vector of time-invariant student-level characteristics from 2013-14, which included:
  - Sex
  - Race
  - Current ELL student
  - Test language of 2013-14 STAAR test
  - Economic Disadvantaged status
  - Received any accommodation on the STAAR administration
  - Student was ever retained in grade
  - Student was considered at risk
  - Immigrant indicator
  - Special Education indicator
- $\boldsymbol{\pi_j}$ is a school fixed effect
- and $\boldsymbol{v_{ij}}$ is a random disturbance term

Model 1B

Model 1B is similar to Model 1A, with the exception that the assumption that the unobserved campus-level disturbances are correlated with the regressors in the statistical model is relaxed, which allows for the inclusion of more substantively meaningful fixed campus-level characteristics. The functional form of the model was:

$$\boldsymbol{SSC_{ijg} = \alpha SUCCESS_i + \beta X_{it-1} + \beta X_i + \delta X_i + u_i + \epsilon_i} \qquad \textbf{(1)}$$

Where:

- $\boldsymbol{SSC_{isg}}$ is the 2013-14 STAAR score for student *i*, attending school *j,* in grade *g*
- $\boldsymbol{\alpha SUCCESS_{it}}$ is an indicator of whether student *i* participated in a SUCCESS program in 2013-14[102]

[102] This term subsumes all variations of the measurement of SUCCESS participation, including total minutes and other binary flavors of participation.

- βX_{it-1} is a vector of time-varying student-level characteristics from the previous school year, and time-invariant characteristics, which included
  - 2012-13 (*T-1)* STAAR-Mathematics score
  - Lagged count of any type of disciplinary actions
  - Lagged attendance rate (2012-13)
  - Lagged indicator of whether student used TTM in the prior year (2012-13)
- βX_i is a vector of time-invariant student-level characteristics from 2013-14, which included:
  - Sex
  - Race
  - Current ELL student
  - Test language of 2013-14 STAAR test
  - Economic Disadvantaged status
  - Received any accommodation on the STAAR administration
  - Student was ever retained in grade
  - Student was considered at risk
  - Immigrant indicator
  - Special Education indicator
- δX_i is a vector of school and district-level characteristics, including
  - Education Service Center (ESC) region
  - District type derived from http://goo.gl/gSoiog
  - Title I status in 2013-14
  - Economic Disadvantaged status
  - Campus accountability rating in 2012-13
  - Campus type (e.g., elementary, secondary, or both)
  - Percentage of students who are minority (Hispanic and Black) in 2013-14
  - Percentage of students who are classified as ELL in 2013-14
  - Percentage of students who met the phase-in 1 Level 2 standard in 2012-13
  - Percentage of students who are classified as at-risk in 2013-14
- u_i is the between-group (school) disturbance term
- and ϵ_i is within-group (students within schools) disturbance term

Model 2

Rather than simply looking at cross-sectional levels of student STAAR-Mathematics performance in 2013-14, the evaluation team fit a two-period DiD model using linear regression, with subject-grade-year and subject-grade-year-prior decile standardized gain scores as the response variable. This model does not account for fixed, unobserved student-level heterogeneity, nor does it account for dynamic selection into the program based on prior performance gains. The model was fit separate for Grades 4-8. The functional form of the statistical model was:

$$\boldsymbol{A_{ijt} - A_{ijt-1} = \alpha SUCCESS_{it} + \beta X_{it-1} + + \pi_j + \upsilon_{ijt}}$$

Where:

- $\boldsymbol{A_{ijt} - A_{ijt-1}}$ is the difference in standardized assessment scores (either subject-grade-year standardized scores or subject-grade-year-prior decile standardized scores) at time *t,* for student *i*, attending school *j*
- $\boldsymbol{\alpha SUCCESS_{it}}$ is an indicator of whether student *i* participated in a SUCCESS program at time *t*
- $\boldsymbol{\beta X_{it-1}}$ is a vector of time-varying student-level characteristics from the previous school year, and time-invariant characteristics, which included
  - Lagged count of any type of disciplinary actions
  - Lagged attendance rate (2012-13)
  - Lagged indicator of whether student used TTM in the prior year (2012-13)
- $\boldsymbol{\beta X_i}$ is a vector of time-invariant student-level characteristics from 2013-14, which included:
  - Sex
  - Race
  - Current ELL student
  - Test language of 2013-14 STAAR test
  - Economic Disadvantaged status
  - Received any accommodation on the STAAR administration
  - Student was ever retained in grade
  - Student was considered at risk
  - Immigrant indicator
  - Special Education indicator
- $\boldsymbol{\pi_j}$ is a school fixed effect
- and $\boldsymbol{\upsilon_{ij}}$ is a random disturbance term

Construction of the Propensity Score and Implementation of the Regression Reweighting Scheme

Following Nichols (2007), the evaluation team used a propensity score reweighting method, where a student's likelihood of being a member of the treatment group, however defined, in 2013-14 is conditioned on a number of pre-treatment school, student, and district covariates X^C. The conditional probability $\hat{\lambda}$ of being in the treatment group derived from this model is then used to calculate a weight based on the odds $\hat{\lambda}/(1-\hat{\lambda})$.

Because evaluators had several different measures of program participation on which treatment and control groups were balanced based on the conditional probability of being a member of the treatment group, the evaluation team fit the propensity score model using several iterations. This method is described below.

Grades 4-8 in 2013-14 with *at least* two valid test scores

7. This sample was used in Research Question 7, since only one valid prior test score was required. All valid cases in Grades 4-8 were included.
8. Fit propensity score model estimating the conditional probability of treatment using a logistic regression based on three measures of TTM use in 2013-14:
    a. Students who attempted at least one TTM lesson compared to students who did not
    b. Students who met the thresholds of interest for lesson attempts (e.g., five or more, 10 or more) for their grade compared to students who did not
    c. Students lesson passing rate (ratio of lessons attempted to lessons passed)
9. Vector of student, school, and district covariates included in the functional form included:
    a. 2012-13 (*T-1)* STAAR-Mathematics score
    b. Lagged count of any type of disciplinary actions
    c. Lagged attendance rate (2012-13)
    d. Lagged indicator of whether student used TTM in the prior year (2012-13)
    e. Sex
    f. Race
    g. Grade level
    h. Current ELL student
    i. Test language of 2013-14 STAAR test
    j. Economic Disadvantaged status
    k. Received any accommodation on the STAAR administration

l. Student was ever retained in grade

m. Student was considered at risk

n. Immigrant indicator

o. Special Education indicator

p. Education Service Center (ESC) region

q. District type derived from http://goo.gl/gSoiog

r. Title I status in 2013-14

s. Economic Disadvantaged status

t. Campus accountability rating in 2012-13

u. Campus type (e.g., elementary, secondary, or both)

v. Percentage of students who are minority (Hispanic and Black) in 2013-14

w. Percentage of students who are classified as ELL in 2013-14

x. Percentage of students who met the phase-in 1 Level 2 standard in 2012-13

y. Percentage of students who are classified as at-risk in 2013-14

10. Calculate the conditional odds of being in each treatment group using the formula: $\hat{\lambda}/(1-\hat{\lambda})$
11. Assign a weight of 1 to *all* students in the treatment condition, and a weight equal to $\hat{\lambda}/(1-\hat{\lambda})$ for all students in the comparison group
12. Fit linear regressions for the response variable on the treatment indicator of interest applying the following restrictions, weights, covariates:

a. Restrict analytic sample to only students in the region of common support based on the propensity score model estimated for the respective treatment measure

i. This is defined, according to Leuven and Sianesi (2003), as cases where the propensity score of the control cases is within the range (minimum and maximum) of the propensity score of the treatment cases.

b. Include all covariates included in the functional form for estimating the propensity score to achieve double-robustness

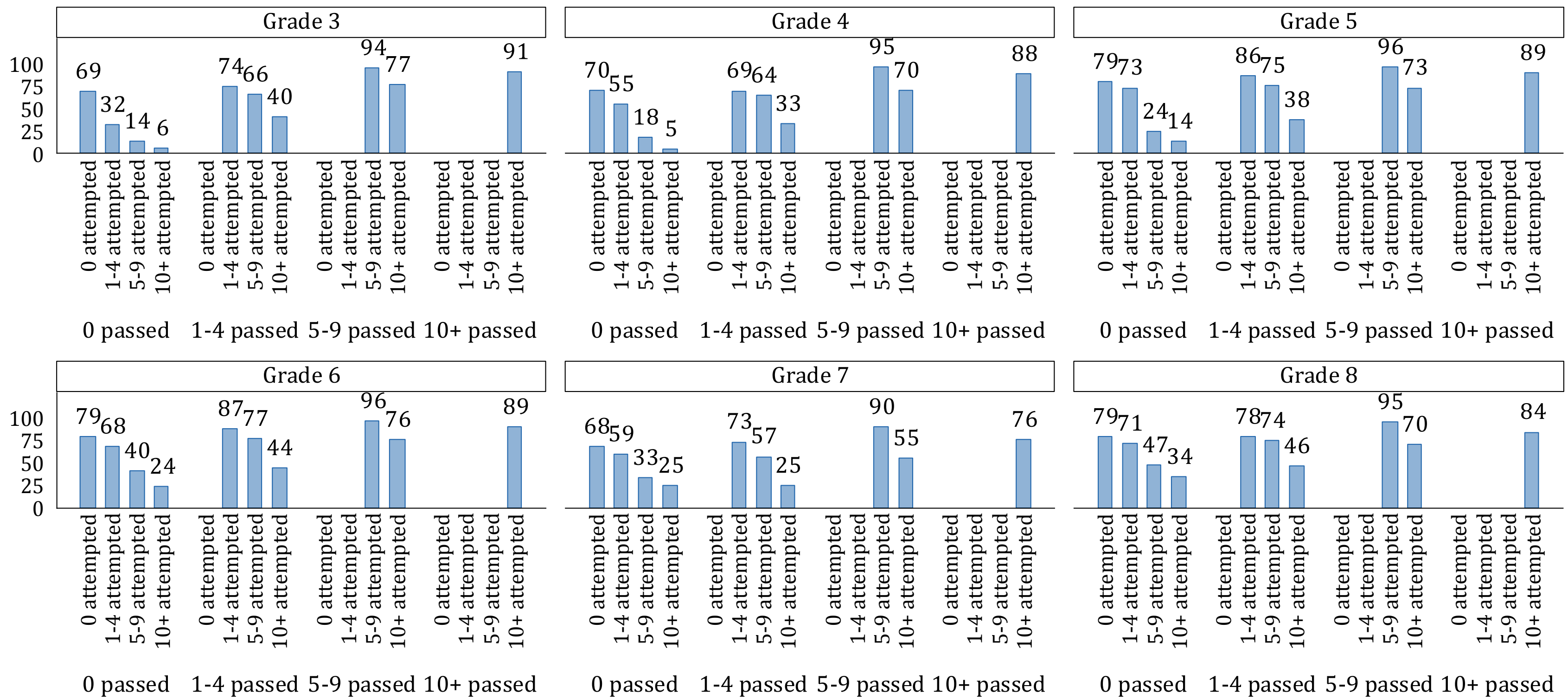
c. Apply probability weights using the weight calculated for the respective treatment measures

Supplementary Analysis using Student Passing Rate

Figure 3.13 depicted the percentage of students, at each grade level, who met the STAAR passing standard in mathematics on the first administration of STAAR in the spring of the 2013-14 school year. Figure B5 shows the same cross-tabulation but with an additional dimension added to the TTM usage metric; the number of TTM lessons *passed*. Students with zero lessons attempted are shown as having passed zero lessons, so grades presented in this group have the same passing rates as students in the zero attempts category in Figure 3.13.

Among students who used the system, Figure B5 demonstrates students who passed more TTM lessons were more likely to meet state standards on the STAAR-Mathematics assessment. For example, among Grade 3 students who attempted 10 or more lessons, only 6% of students who passed none of their attempted lessons met the STAAR passing standard while 91% of students who passed 10 or more lessons met the standard; in other words, there was an 85 percentage point gap in passing rates between students who attempted but did not pass 10 or more lessons and students who attempted and passed 10 or more TTM lessons. This gap decreased across grades, with the smallest percentage point difference in passing rates demonstrated among Grade 8 students, 38% of those who attempted 10 or more lessons and passed zero lessons meeting the STAAR standard versus 84% of students who attempted and passed 10 or more TTM lessons (a 46 percentage point difference). Another interesting pattern here is that students who registered five to nine attempts and the same number of passed lessons had, overall, the highest rates of meeting the STAAR standard. That is, across all grades, 90% or more of students who attempted and passed 5-9 lessons met the STAAR standard.

Figure B5. Percentage of Students Meeting the STAAR-Mathematics Passing Standard, by Grade and TTM Lessons Attempted and Passed (2013-14)



Source: Think Through Math assessment and session history tables and State of Texas Assessments of Academic Readiness data, 2013-14, Texas Education Agency.

Note: State of Texas Assessments of Academic Readiness performance level includes the first administration only, and regular English and Spanish Versions only (i.e., not modified or alternate versions).

Table B4 provides a different operationalization of student TTM usage and intensity. In these analyses, the evaluation team used Model 2 with TTM lesson passing rates, rather than lessons attempted, to explore the association between TTM lesson passing rate and STAAR decile-standardized gains in mathematics. Interpretation of this table is different than previous models. First, the passing rate is a ratio of number of lessons passed to the number of lessons attempted, meaning that having a high passing rate was a function of the number of TTM lessons attempted. This means that a student who attempted one lesson and passed had the same passing rate as a student who attempted and passed 100 lessons. Therefore, the passing rate is interacted with whether the student met the five or more lessons attempted threshold. The rationale here is that this shows that students who only attempted a few lessons and yet passed many of them were different from students who attempted and passed many lessons. The reference category for the passing rate coefficients is non-users (who automatically have a passing rate of zero).

Overall, students with higher passing rates on TTM lessons also had higher STAAR decile-standardized gains. In Grade 4, for example, students who attempted five or more lessons demonstrate a .783 standard deviation increase in gains as their passing rate increased (in comparison with non-users), and students with less than five lessons attempted still had an increase in STAAR gains when their passing rate increased. Associations between passing rates and decile standardized gains were higher among students whose usage was above the threshold than those with usage below the threshold. In Grade 5, for example, students who attempted five or more lessons demonstrated a .784 standard deviation increase in gains as their passing rate increased (in comparison with non-users) while students whose usage was below the threshold demonstrated an increase of .279 standard deviations (both as compared to non-users). Across all grades, associations between decile-standardized gains and TTM use were weaker among students whose TTM use was above the threshold, but who did not pass any of the lessons they attempted.

Table B4. Estimated Effects of TTM Passing Rate and Usage on STAAR-Mathematics Test Score Gains between 2011-12 and 2013-14, by Student Grade Level and Measure of Program Participation

		Model 2 (Passing Rate)		
Grade Level	**System Usage Measure**	**B**	**SE**	**N**
Grade 4	Passing Rate (at/above threshold)	**0.783*** **	0.031	156,813
	Passing Rate (below threshold)	**0.376*** **	0.027	
	Above threshold (no lessons passed)	***-0.318****	0.016	
Grade 5	Passing Rate (at/above threshold)	**0.784*** **	0.029	155,073
	Passing Rate (below threshold)	**0.279*** **	0.024	
	Above threshold (no lessons passed)	***-0.403****	0.016	
Grade 6	Passing Rate (at/above threshold)	**0.620*** **	0.032	108,809
	Passing Rate (below threshold)	**0.390*** **	0.023	
	Above threshold (no lessons passed)	***-0.252****	0.017	
Grade 7	Passing Rate (at/above threshold)	**0.590*** **	0.040	83,176
	Passing Rate (below threshold)	**0.209*** **	0.029	
	Above threshold (no lessons passed)	***-0.231****	0.019	
Grade 8	Passing Rate (at/above threshold)	**0.675*** **	0.045	65,277
	Passing Rate (below threshold)	**0.296*** **	0.030	
	Above threshold (no lessons passed)	***-0.225****	0.023	
Lagged STAAR-Mathematics Score?			No	
Campus Fixed Effects?			Yes	
Campus Random Effects			No	
Propensity Score Weighting?			No	
Outcome Measure		Prior achievement decile-grade-year standardized gain score		

Source: Think Through Math session history tables and State of Texas Assessments of Academic Readiness data, 2012-13-2013-14, Texas Education Agency.

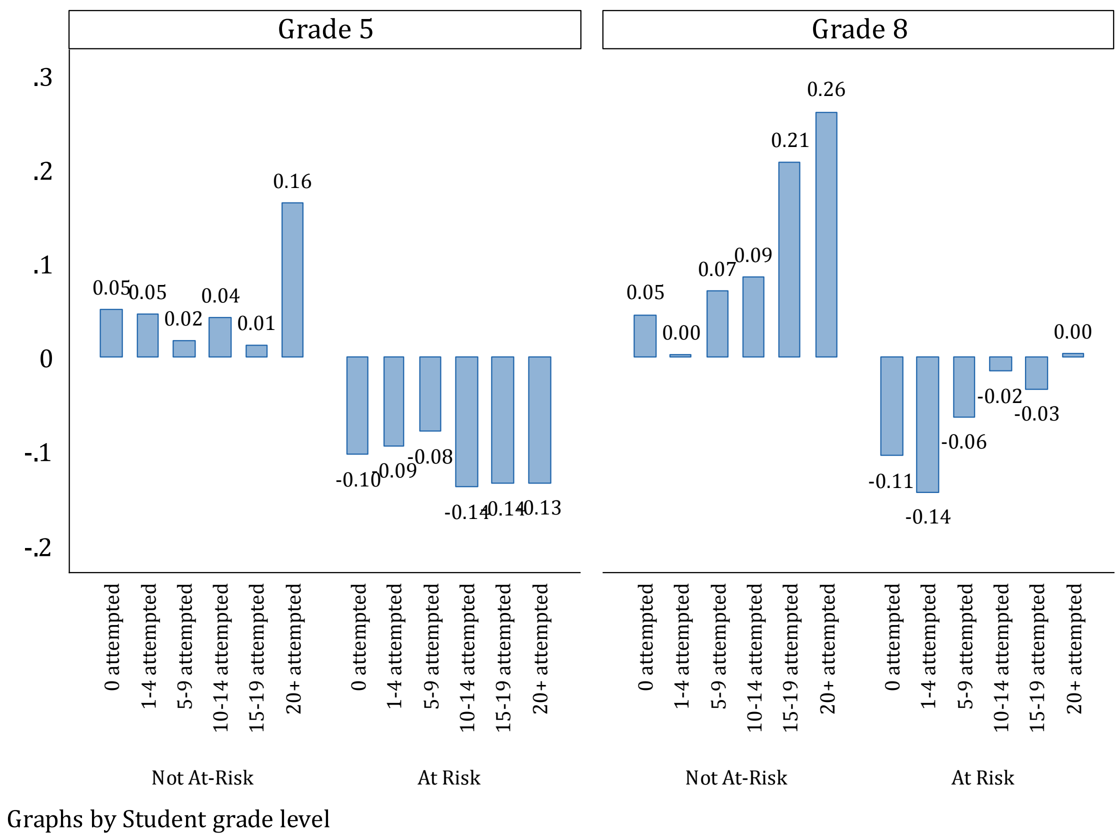
Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) are included in the analyses. Number of attempts are prior to the administration of the first assessment, which is contingent on grade level. Estimates derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. Statistically significant positive coefficients are denoted by bold font, and negative coefficients are denoted by bold and italicized font. * p<0.05, ** p<0.01, *** p<0.001. B refers to the Beta coefficient for the statistical model and SE refers to the standard error.

Research Question 9

The figure below displays the descriptive relationship between STAAR-Mathematics performance (i.e., decile-grade-year standardized score gains on the STAAR-Mathematics assessment) and TTM system usage in 2013-14. TTM usage level is divided into six categories: No attempts, between one and four attempts, between five and nine attempts, between 10 and 14 attempts, between 15 and 20 attempts minutes, and 20 or more attempts. Across all TTM usage categories, students in Grades 5 and 8 at risk of being retained demonstrated declines in decile-standardized STAAR-Mathematics assessment scores while students who were not classified as at risk of being retained generally demonstrated gains.

Considering just students classified as at risk, the relationship between mathematics outcomes and TTM usage differed by grade. That is, in Grade 5, declines in STAAR-Mathematics assessment scores were larger among students who attempted 10 or more lessons while, in Grade 8, declines were smaller among students with such use. Conversely, more intensive TTM usage—particularly 20 or more lesson attempts—was associated with better mathematics outcomes among both students at risk of being retained and students not at risk of being retained in Grade 8 while the same was true only for Grade 5 students who were not at risk.

Figure B6. Mean Decile-Standardized Gain Scores between 2012-13 and 2013-14 on STAAR-Mathematics, by Grade Level and TTM Usage



Source: State of Texas Assessments of Academic Readiness data, 2011-12 – 2013-14, Texas Education Agency. TTM session history table, 2013-14.

Note: Calculations only include students who met the following inclusion rules: were enrolled at a campus where at least one student was registered for TTM in 2013-14; had a valid STAAR-Mathematics test score in the prior year (2012-13). Only regular English and Spanish versions (i.e., not modified or alternate versions) are included in the at-risk indicator flag. Minutes are prior to the administration of the first assessment, which is contingent on grade level. A student was defined as being at risk of being retained if they failed at least one STAAR-Mathematics assessment in 2011-12 or 2012-13.

Econometric Model Specification

To address Research Question 9, the base model described in Appendix 7 was amended to investigate whether the impact of participating in a SUCCESS program varied between students who were considered at risk of being retained in grade compared to students who were not at risk of being retained in grade. This was done through the inclusion of an interaction term, between SUCCESS participation and the at-risk for grade retention indicator. More formally,

$$A_{ijt} - A_{ijt-1} = \alpha SUCCESS_{it} + \lambda SUCCESS_{it}ATRISK_i + \phi ATRISK_i + \beta X_{it-1} + + \pi_j + \upsilon_{ijt}$$

Where:

- $A_{ijt} - A_{ijt-1}$ is the difference in standardized assessment scores (either subject-grade-year standardized scores or subject-grade-year-prior decile standardized scores) at time *t,* for student *i*, attending school *j*
- $\alpha SUCCESS_{it}$ is an indicator of whether student *i* participated in a SUCCESS program at time *t*, and represents the mean program effect for students in the base reference category for the student group of interest (here, Hispanic students).
- $\lambda SUCCESS_{it}ATRISK_i$ is the multiplicative term between the at-risk flag and the measure of program participation, capturing the mean *difference in the program effect* between students who were at risk and who participated in the program relative to students who were not at risk and who participated in the program. This is the coefficient of substantive interest for identifying student group heterogeneity among students who participated in the program.
- $\phi ATRISK_i$ is the average effect of students who were at risk *relative to students who were not at risk who were not in the SUCCESS participant group*.
- βX_{it-1} is a vector of time-varying student-level characteristics from the previous school year, and time-invariant characteristics, which included
  - Lagged count of any type of disciplinary actions
  - Lagged attendance rate (2012-13)
  - Lagged indicator of whether student used TTM in the prior year (2012-13)
- βX_i is a vector of time-invariant student-level characteristics from 2013-14, which included:
  - Sex
  - Race
  - Current ELL student
  - Test language of 2013-14 STAAR test
  - Economic Disadvantaged status
  - Received any accommodation on the STAAR administration
  - Student was ever retained in grade
  - Student was considered at-risk
  - Immigrant indicator
  - Special Education indicator
- π_j is a school fixed effect

- and $\boldsymbol{v_{ij}}$ is a random disturbance term

The base, or reference, category for student groups were:

- ELL status: Not ELL
- Economic disadvantaged status: Not economically disadvantaged
- Ethnicity: Hispanic

Models were fit separately by grade and program participation dosage measure.

Research Question 10

For Research Question 10, the outcome of interest is whether the student met the passing standard on the second administration of STAAR-Mathematics *after having failed the first administration.* The outcome is binary, so a multivariate logistic regression was fit, and a measure of SUCCESS usage *between the first and second administration* was calculated and included as a covariate in the model to determine the effect of system dosage on the probability passing the second administration of the respective test. Before models were fit, several business rules were established to determine which students were eligible to be included in the model. They were:

4. Student failed the first administration of the STAAR exam in 2013-14
    a. STAAR-Mathematics failure were considered separately for the subject-specific models.
    b. Only regular STAAR (e.g., no STAAR Modified or Alternate) tests were considered.
5. Student took the second administration STAAR exam in 2013-14
    a. Only regular STAAR (e.g., no STAAR Modified or Alternate) tests were considered.
6. Student was enrolled, and took, the Grade 5 or Grade 8 STAAR exams for the first and second assessment

A logistic regression with student-, school-, and district-level covariates. Logistic regression estimates the log odds of the outcome (here, passing the second administration of STAAR) as a function of the included covariates. The evaluation team fit a logistic regression, with cluster-adjusted standard errors to account for within-school non-independence, with a logit link function ($\mathrm{n_{ijd}} = \log(\frac{\phi_{\mathrm{ijd}}}{1-\phi_{\mathrm{ijd}}})$), where $\mathrm{n_{ijd}}$ is the log-odds of student i in school j and district d repeating Grade 5 or Grade 8 in 2014-15. The functional form is formally expressed as

$$\boldsymbol{n_{ijd} = \beta_0 + \beta_1 SUCCESS_{ijd} + \beta_k X_{ijdt-1} + \beta_k Q_{jdt-1} + \beta_k Y_{dt-1} + e_{ijd} + \mu_{jd} + v_d} \qquad \textbf{(3)}$$

Where:

- $\boldsymbol{\beta_1 SUCCESS_{ijd}}$is the measure of student *I's* participation in a SUCCESS program in 2013-14

- $\boldsymbol{\beta_k} X_{\boldsymbol{ijdt-1}}$ is a vector of student-level characteristics from the prior year and current year, including
  - Lagged count of any type of disciplinary actions
  - Lagged attendance rate (2012-13)
  - Lagged indicator of whether student used TTM in the prior year (2012-13)
  - STAAR-Mathematics scaled score from the first administration in March 2013-14
  - Sex
  - Race
  - Current ELL student
  - Test language of 2013-14 STAAR test
  - Economic Disadvantaged status
  - Received any accommodation on the STAAR administration
  - Student was ever retained in grade
  - Student was considered at risk
  - Immigrant indicator
  - Special Education indicator
- $\boldsymbol{\beta_k Q_{jdt-1}}$ is a vector of school-level characteristics
  - Title I status in 2013-14
  - Campus accountability rating in 2012-13
  - Campus type (e.g., elementary, secondary, or both)
  - Percentage of students who are minority (Hispanic and Black) in 2013-14
  - Percentage of students who are classified as ELL in 2013-14
  - Percentage of students who met the phase-in 1 Level 2 standard in 2012-13
  - Percentage of students who are classified as at-risk in 2013-14
  - Percentage of students classified as economically disadvantaged in 2013-14
- $\boldsymbol{\beta_k} Y_{\boldsymbol{dt-1}} \boldsymbol{\pi_{jd}}$ is a vector of district-level characteristics
  - Education Service Center (ESC) region
  - District type derived from http://goo.gl/gSoiog
- $\boldsymbol{e_i}$ is a random error term for student *i* in school *j* in district *d*

Description and Preparation of Think Through Mathematics Files

TTM provided TEA—who, in turn, provided Gibson Consulting Group—two files representing different dimensions of the TTM program: a Lesson table, and an Enrollment table. Of primary interest for the evaluation is lesson session usage, which the evaluation team obtained primarily from the Lesson table. The Lesson table contains the most granular, lowest-level information about students' system usage. Each row reflects a single, unique TTM lesson launched, or attempted, by a student, including the date and time the session started and the whether the lesson was passed.

The Lesson table was then linked to the Enrollment table to obtain more precise information about each unique session, and to calculate the number of students who enrolled in TTM but never participated in lessons. Roughly 94% of records from the Lesson table were linked to a corresponding user in the Enrollment table.[103]

The TTM tables that Gibson received contain a unique student database key that could not be joined to TEA administrative and assessment records. Beginning in the 2013-14 school year, TTM was required to record students' state identifier which would permit TTM system usage and assessment performance data to be linked to TEA administrative and assessment records, including student attendance, economic disadvantaged status, and STAAR test performance. TTM provided records for all students who were registered in the system with this identifier, along with other personally identifiable information contained within the system, including students' first and last name. These records were then linked back to TEA's unique state identifier, producing a match rate of about 72%.[104]

These linked records were then provided to Gibson, and were subsequently linked to the base Gibson file that contained students' session information for the 2013-14 school year. The base Gibson files includes 23,795,362 enrollment and lesson records representing 8,169,964 unique students. Overall, 216,609 unique students were removed from the data because they were in the lessons table only but had no enrollment information, they had enrolled but never taken a lesson session, or they had enrolled and taken TTM lesson sessions but they could not be matched to PEIMS data or were not in Grades 3-8 in 2013-14. The final analytic dataset contains 15,848,902 lesson records for students matched with TEA administrative data. These records represent 2,298,111 non-duplicated students (836,228 of which are TTM users) in Grades 3 through 8 for 2013-14 across 3,118 campuses.

[103] Only records which were successfully matched between the two tables were retained. A small proportion of unmatched records from the Lessons file that did not have Enrollment table matches were due to slightly different time frames used by the vendor to query the data.

[104] TTM IDs were not validated to match TEA IDs when they were entered, which contributed to the low match rate when TTM submitted the data to be joined with TEA data. Also, the TEA enrollment platform allows for both individual teachers to enroll students or bulk upload. Because the Unique IDs were not known by everyone, there were more room for mistakes or missing values.

Appendix C – Campus and District Staff Interview Research Methods

Campus Staff Interviews

The sampling design for reading interventionists identified a random sample of the highest, medium, and lowest campuses in terms of how much students used Istation (i.e., intensity of usage) and how many students (within and across grades) at a campus used Istation (i.e., coverage). The sampling unit is at the campus-type-level. The sampling frame included all enrolled students for any campuses that have at least one student who used Istation during the 2013-14 school year. The intensity (e.g., average minutes used) and coverage (e.g., proportion of students) of Istation usage was calculated for each grade level at each campus (at the campus-grade level).[105] For the sampling of campus-level mathematics interventionists, the intensity (e.g., number of lessons per student) and coverage (e.g., proportion of students who complete five or more lessons, which is TTM's metric for a student engagement) of TTM usage was calculated for each grade level at each campus (at the campus-grade level). For Istation and TTM utilization, these metrics were combined with a measure of the proportion of grades at a campus that utilizes each system, and transformed into a standardized scale, which identified campus-level system usage relative to similar campus types (e.g., elementary or middle).

For both Istation and TTM, random samples of 50 campuses were drawn from the top 10%, the bottom 10%, and the middle 20% of the distribution based on campus scores according to this standardized scale. Samples for elementary and middle school campus levels were drawn independently (with campuses that span across both of these levels randomly assigned to either the elementary or middle school category).

Table C1. Sampling Approach for Spring 2014 Campus-Level Interviews

School Type	Lowest 10%	Middle 20%	Highest 10%
Elementary	50 schools	50 schools	50 schools
Middle	50 schools	50 schools	50 schools

For both Istation and TTM samples, six strata or selection categories were created: high, medium, and low Istation usage groupings for both the elementary and middle school campus levels. Fifty campuses were randomly selected into each of the six stratum, producing a total of 300 campuses included in the sample of Istation campuses. Campuses were oversampled significantly in order to have a sufficient number of backups, given the narrow window in which the evaluation team had to conduct these interviews. The evaluation team only attempted to contact 15 campuses in each stratum, for a total of 90 campuses.

The evaluation team conducted telephone interviews with campus staff at 78 elementary and 77 middle school campuses over the April 22, 2014 to June 8, 2014 period. In all, 83 interviews were conducted with

[105] The sampling approach was slightly different for TTM because the intensity and coverage variables were different in the two systems. Otherwise, all else remained comparable between the Istation and TTM sampling methods.

staff charged with implementation of TTM and other mathematics interventions and 72 were charged with implementing Istation and other English Language Arts (ELA) interventions. A total of 155 interviews were completed.

Interviews were conducted via telephone and responses were entered directly into an online data collection system. Interviews were also tape recorded so that data collection staff could refer to the recordings for additional information, to be certain respondents were accurately quoted, and to clarify where necessary.

In all, staff at 111 different districts participated in the interviews. Of those, 56 districts are represented in the Istation data (with 11 districts represented more than once), and 61 districts are represented in the TTM interview data (with 10 represented more than once). Table C2 shows the number of completed interviews with campus-level reading and mathematics intervention staff stratified by elementary and middle schools and by system usage level (i.e., high, moderate, and low).

Table C2. Completed Istation and TTM Interviews with Campus-Level Staff

Implementation Level	Completed Istation (Reading) Interviews		Completed TTM (Mathematics) Interviews	
	Elementary School	**Middle School**	**Elementary School**	**Middle School**
High	13	15	12	19
Moderate	13	13	12	14
Low	12	6	16	10
TOTALS	**38**	**34**	**40**	**43**
	72		**83**	

Source: Spring 2014 Interviews with Campus-level Staff.

District Interviews

Between June 1, 2014 and June 25, 2014, the evaluation team conducted telephone interviews with district staff at 29 districts. In all, 30 interviews were conducted with staff responsible for district-wide coordination of mathematics and reading interventions, including TTM and Istation. Districts were included in the sample for potential interviews if mathematics or reading interventionists at one or more campuses within their district were interviewed for this study.

Table C3 shows the planned number of interviews and the final number of district personnel interviews in each category. All districts with a completed campus level interview were contacted to participate in the district-level interview. Although the district sample was a convenience sample, the evaluation staff worked to balance the number of districts that were interviewed by initial campus implementation level (high, moderate, low) and campus level (elementary, middle school) to the extent possible. Table C3 below shows the planned and actual number of completed district interviews by these variables.

Table C3. Final District Interview Totals by Implementation and School Level

Elementary School	Middle School
4	4
6	2
16	***14***
30 Completed Interviews	

Note: The 30 interviews reflect 29 different school districts

Appendix D – Supplementary Istation and TTM System Usage Tables

This appendix contains supplementary tables related to descriptive reading and mathematics findings.

Reading

Table D1. Istation System Usage Disaggregated by Title I Status and Grade Level, 2013-14

Grade Level	Title I Status	Number of Students	Percent of Students Using System at Recommended Threshold Levels	Percent of Students Registered to Use Istation Who Attempted a Curriculum Session	Mean Number of Total Minutes per Student Spent on Curriculum Sessions
Grade 3	Not Title 1	78,429	29.54%	66.97%	232
Grade 3	Title 1	309,992	40.99%	72.60%	360
Grade 4	Not Title 1	80,418	21.17%	61.72%	165
Grade 4	Title 1	301,641	33.68%	70.76%	274
Grade 5	Not Title 1	87,189	17.25%	57.73%	135
Grade 5	Title 1	294,034	31.97%	69.78%	252
Grade 6	Not Title 1	125,989	9.19%	35.18%	61
Grade 6	Title 1	248,986	16.68%	51.38%	113
Grade 7	Not Title 1	144,966	6.20%	28.92%	45
Grade 7	Title 1	239,081	10.81%	46.70%	77
Grade 8	Not Title 1	144,747	4.54%	25.18%	33
Grade 8	Title 1	233,459	7.89%	40.67%	55

Source: Istation session-by-product table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. The recommended minimum Istation threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum Istation threshold usage level for Grades 6-8 is 200 minutes.

Table D2. Istation Usage Disaggregated by Campus Accountability Rating and Grade, 2013-14

Grade Level	Accountability Rating	Number of Students	Percent of Students Using System at Recommended Threshold Levels	Percent of Students Registered to Use Istation Who Attempted a Curriculum Session	Mean Number of Total Minutes per Student Spent on Curriculum Sessions
Grade 3	Improvement Required	34,061	45.11%	74.07%	395
Grade 3	Met Standard	350,832	38.11%	71.35%	330
Grade 4	Improvement Required	33,718	37.57%	72.57%	309
Grade 4	Met Standard	345,954	30.34%	68.46%	246
Grade 5	Improvement Required	26,363	41.24%	76.45%	321
Grade 5	Met Standard	352,176	27.64%	66.35%	218
Grade 6	Improvement Required	24,604	20.74%	55.77%	137
Grade 6	Met Standard	344,949	13.76%	45.48%	93
Grade 7	Improvement Required	22,777	13.47%	54.08%	101
Grade 7	Met Standard	357,307	8.73%	39.22%	63
Grade 8	Improvement Required	21,938	8.97%	46.59%	65
Grade 8	Met Standard	352,689	6.48%	34.10%	45

Source: Istation session-by-product table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. The recommended minimum Istation threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum Istation threshold usage level for Grades 6-8 is 200 minutes.

Table D3. Istation Usage Disaggregated by Priority or Focus Schools and Grade, 2013-14

Grade Level	Status	Number of Students	Percent of Students Using System at Recommended Threshold Levels	Percent of Students Registered to Use Istation Who Attempted a Curriculum Session	Mean Number of Total Minutes per Student Spent on Curriculum Sessions
Grade 3	Focus	30,596	46.11%	77.74%	412
Grade 3	Priority	10,834	47.82%	73.44%	418
Grade 4	Focus	30,310	39.40%	77.26%	329
Grade 4	Priority	10,018	41.09%	73.61%	323
Grade 5	Focus	26,258	38.66%	76.04%	315
Grade 5	Priority	8,643	41.27%	79.23%	322
Grade 6	Focus	37,758	19.54%	61.33%	130
Grade 6	Priority	10,214	14.90%	48.88%	116
Grade 7	Focus	33,137	15.00%	62.13%	100
Grade 7	Priority	9,441	9.54%	42.30%	65
Grade 8	Focus	31,637	11.03%	56.29%	71
Grade 8	Priority	9,339	4.27%	36.18%	40

Source: Istation session-by-product table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. The recommended minimum Istation threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum Istation threshold usage level for Grades 6-8 is 200 minutes.

Table D4. Istation System Usage Disaggregated by Urbanicity, 2013-14

Urbanicity Classification	Number of Students	Percent of Students Using System at Recommended Threshold Levels	Percent of Students Registered to Use Istation Who Attempted a Curriculum Session	Mean Number of Total Minutes per Student Spent on Curriculum Sessions
Charter School	89,283	8.13	22.35	66
Independent Town	111,889	26.50	55.39	206
Major Suburban	755,501	19.41	50.87	155
Major Urban	427,447	25.97	72.33	215
Non-Metropolitan Fast Growing	13,372	13.18	32.22	98
Non-Metropolitan Stable Growth	127,075	22.34	51.33	171
Other Central City	375,533	22.15	57.76	172
Other Central City Suburban	322,566	20.99	50.62	163
Rural	72,712	21.83	40.88	174
Total	**2,295,378**	**21.41**	**54.67**	**170**

Source: Istation session-by-product table, Public Education Information Management System data, and author's calculations, 2014.

Note: This table includes students who did not register or use the Istation system to reflect statewide Grade 3-8 estimates of usage. The data presented in the totals column may vary slightly from table to table because data used to disaggregate results were not always available for all students. The recommended minimum Istation threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum Istation threshold usage level for Grades 6-8 is 200 minutes.

Table D5. Istation System Usage Disaggregated by Gender and Grade Level, 2013-14

Grade Level	Gender	Number of Students	Percent of Students Using System at Recommended Threshold Levels	Mean Number of Total Minutes per Student Spent on Curriculum Sessions
Grade 3	Female	189,627	38.35%	325
Grade 3	Male	200,350	38.80%	341
Grade 4	Female	187,142	30.24%	241
Grade 4	Male	196,397	31.67%	259
Grade 5	Female	187,073	27.70%	216
Grade 5	Male	195,819	29.30%	232
Grade 6	Female	183,275	13.46%	90
Grade 6	Male	193,283	14.76%	100
Grade 7	Female	187,487	8.39%	60
Grade 7	Male	197,978	9.67%	69
Grade 8	Female	185,358	5.98%	42
Grade 8	Male	194,322	7.17%	50

Source: Istation session-by-product table and author's calculations, 2014.

Note: The recommended minimum threshold usage level for Grades 3-5 is 250 minutes, and the recommended minimum threshold usage level for Grades 6-8 is 200 minutes.

Table D6. Percent of Campus Staff "Very Satisfied" with Various Aspects of the Istation Program

	Overall (n=68)	High Usage (n=26)	Moderate Usage (n=25)	Low Usage (n=17)	Elementary School (n=36)	Middle School (n=32)
The user-friendliness of Istation	70%	73%	72%	59%	64%	75%
The appropriateness of content for your students	64%	77%	64%	47%	67%	63%
The delivery of the content in an engaging manner	60%	65%	64%	47%	63%	58%
The level of support, professional development and/or technical assistance from vendors	36%	38%	48%	20%	43%	30%
The ESC 20 online help function or technical phone support	61%	60%	69%	50%	67%	60%

Source: Spring 2014 Interviews with campus-level staff.

Table D7. Percentage of School Staff Reporting Use of Istation Curriculum, by Grade Level

Grade Level	Percent of Istation Users (N= 68)
K—2	66%
3	92%
4	95%
5	95%
6	65%
7	79%
8	79%

Source: Spring 2014 Interviews with campus-level staff.

Table D8. Strategies and Resources Used to Address SSI Requirements by Usage and School-Level

	Overall (n=69)	High Usage (n=26)	Moderate Usage (n=25)	Low Usage (n=18)	Elementary School (n=36)	Middle School (n=33)
In class strategies	71%	62%	88%	61%	56%	88%
Assessments	14%	8%	16%	22%	14%	15%
Out of school strategies	41%	23%	56%	44%	47%	33%
Other online programs	35%	38%	36%	22%	36%	30%

Source: Spring 2014 Interviews with campus-level staff.

Mathematics

Table D9. TTM System Usage Disaggregated by Title I Status and Grade Level, 2013-14

Grade Level	Title I Status	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM More TTM
Grade 3	Not Title 1	37,527	85.78%	61.29%	59.76%	32.12%
Grade 3	Title 1	130,021	86.79%	63.70%	46.73%	21.50%
Grade 4	Not Title 1	39,430	81.64%	58.75%	59.55%	36.42%
Grade 4	Title 1	135,152	82.97%	61.08%	49.26%	27.14%
Grade 5	Not Title 1	38,644	77.24%	53.97%	57.99%	35.06%
Grade 5	Title 1	135,889	82.79%	60.58%	53.59%	28.35%
Grade 6	Not Title 1	40,404	63.31%	44.54%	45.22%	28.96%
Grade 6	Title 1	83,551	70.09%	50.44%	43.56%	23.80%
Grade 7	Not Title 1	38,329	67.51%	46.75%	47.07%	28.75%
Grade 7	Title 1	62,600	70.30%	47.86%	41.71%	21.66%
Grade 8	Not Title 1	37,506	62.28%	42.55%	43.35%	26.43%
Grade 8	Title 1	55,494	59.19%	39.46%	34.64%	17.51%

Source: Think Through Math lesson table and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table. The "5 or more" and "10 or more" categories are not mutually exclusive (i.e., student included in 10 or more percentages are also included in 5 or more percentages).

Table D10. TTM System Usage Disaggregated by Urbanicity, 2013-14

Urbanicity Classification	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
Charter School	13,202	82.26%	60.47%	53.58%	31.65%
Independent Town	44,918	78.25%	59.00%	49.33%	25.80%
Major Suburban	272,417	77.56%	56.26%	51.17%	29.44%
Major Urban	174,971	72.29%	49.08%	41.41%	21.03%
Non-Metropolitan Fast Growing	2,257	81.61%	56.05%	53.04%	22.95%
Non-Metropolitan Stable	42,751	80.87%	59.50%	50.89%	26.99%
Other Central City	137,089	76.62%	55.24%	47.55%	25.84%
Other Central City Suburban	130,880	79.21%	57.24%	49.67%	25.89%
Rural	17,362	81.67%	60.69%	54.73%	31.14%
Total	**836,228**	**76.93%**	**55.21%**	**48.30%**	**26.26%**

Source: Think Through Math lesson table and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table. The "5 or more" and "10 or more" categories are not mutually exclusive (i.e., student included in 10 or more percentages are also included in 5 or more percentages).

Table D11. TTM System Usage Disaggregated by Priority or Focus Schools and Grade Level, 2013-14

Grade Level	Status	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
Grade 3	Focus	13,713	88.86%	66.78%	41.98%	18.78%
Grade 3	Priority	4,301	86.14%	63.96%	35.57%	13.76%
Grade 4	Focus	14,139	82.80%	60.86%	43.31%	22.24%
Grade 4	Priority	4,283	81.09%	61.73%	39.62%	20.06%
Grade 5	Focus	13,287	84.84%	64.56%	50.49%	26.33%
Grade 5	Priority	4,056	86.49%	67.83%	51.23%	26.33%
Grade 6	Focus	12,606	66.09%	44.70%	35.08%	16.85%
Grade 6	Priority	3,585	67.45%	52.50%	38.02%	19.16%
Grade 7	Focus	10,435	69.49%	46.05%	36.19%	17.19%
Grade 7	Priority	2,678	78.34%	57.77%	43.61%	21.02%
Grade 8	Focus	8,400	61.38%	40.08%	31.94%	14.26%
Grade 8	Priority	1,866	53.27%	34.41%	27.81%	13.99%

Source: Think Through Math lesson table and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table. The "5 or more" and "10 or more" categories are not mutually exclusive (i.e., student included in 10 or more percentages are also included in 5 or more percentages).

Table D12. TTM Usage Disaggregated by Prior Year (2012-13) STAAR-Mathematics Performance, 2013-14

Grade Level	Prior Year STAAR-Mathematics Performance	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
Grade 4	Bottom quartile	40,820	80.74%	58.47%	31.17%	12.38%
Grade 4	Top quartile	38,702	84.93%	65.14%	72.48%	50.06%
Grade 5	Bottom quartile	44,243	84.21%	62.82%	42.93%	17.43%
Grade 5	Top quartile	37,637	79.68%	58.60%	67.37%	45.13%
Grade 6	Bottom quartile	30,465	70.04%	50.93%	34.74%	16.51%
Grade 6	Top quartile	27,208	69.20%	51.17%	57.79%	40.77%
Grade 7	Bottom quartile	29,437	75.42%	54.64%	40.16%	19.75%
Grade 7	Top quartile	19,470	66.62%	47.19%	56.11%	37.91%
Grade 8	Bottom quartile	24,938	64.45%	45.79%	34.29%	15.64%
Grade 8	Top quartile	16,112	62.75%	43.01%	50.65%	34.62%

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table. The "5 or more" and "10 or more" categories are not mutually exclusive (i.e., student included in 10 or more percentages are also included in 5 or more percentages).

Table D13. 2013-14 TTM Usage Disaggregated by Race/Ethnicity

Prior Year	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
American Indian	2,962	78.09%	55.60%	48.75%	26.00%
Asian/Pacific Islander	33,834	82.60%	65.99%	69.36%	50.17%
Black	108,296	76.90%	54.54%	43.82%	22.67%
Hispanic	440,800	76.87%	55.22%	44.81%	22.88%
White	234,956	76.24%	53.96%	53.56%	30.52%
Two or more	15,380	77.02%	54.82%	52.85%	30.72%
Total	**836,228**	**76.93%**	**55.21%**	**48.30%**	**26.26%**

Source: Think Through Math lesson table and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table. The "5 or more" and "10 or more" categories are not mutually exclusive (i.e., student included in 10 or more percentages are also included in 5 or more percentages).

Table D14. TTM System Usage Disaggregated by Gender and Grade Level, 2013-14

Grade Level	Gender	Number of Students	Percent of Students Attempting 5 or More TTM Lessons	Percent of Students Attempting 10 or More TTM Lessons	Percent of Students Passing 5 or More TTM Lessons	Percent of Students Passing 10 or More TTM Lessons
Grade 3	Female	81,693	85.55%	59.93%	47.06%	20.97%
Grade 3	Male	86,389	87.54%	66.23%	52.19%	26.70%
Grade 4	Female	85,552	81.06%	57.11%	48.76%	26.74%
Grade 4	Male	89,496	84.22%	63.89%	54.39%	31.72%
Grade 5	Female	85,705	80.45%	56.64%	53.08%	28.14%
Grade 5	Male	89,171	82.63%	61.48%	56.01%	31.51%
Grade 6	Female	60,265	66.78%	47.12%	44.36%	25.37%
Grade 6	Male	63,865	68.94%	49.87%	43.91%	25.64%
Grade 7	Female	48,908	68.69%	46.79%	44.19%	24.66%
Grade 7	Male	52,115	69.79%	48.09%	43.37%	24.09%
Grade 8	Female	45,779	60.44%	40.64%	38.91%	21.67%
Grade 8	Male	47,290	60.41%	40.76%	37.42%	20.55%

Source: Think Through Math lesson table, Public Education Information Management System data, and author's calculations, 2014.

Note: The data presented in the totals column may vary slightly from table. The "5 or more" and "10 or more" categories are not mutually exclusive (i.e., student included in 10 or more percentages are also included in 5 or more percentages).

Table D15. Percentage of School Staff Reporting Use of TTM, by Grade Level

Grade Level	Percent of TTM Users (N= 83)
3	82%
4	85%
5	85%
6	84%
7	88%
8	86%

Source: Spring 2014 Interviews with campus-level staff.

Table D16. Major Themes Related to Changes in TTM Usage, Overall and By Usage Patterns, 2013-14

	Overall (n= 62)	High Usage (n= 30)	Moderate Usage (n= 18)	Low Usage (n= 14)
Used TTM more	38%	37%	33%	36%
Used TTM less	10%	0%	11%	29%
Used TTM differently	26%	20%	39%	21%

Source: Texas SUCCESS TTM Interviews, 2014.

Table D17. TTM Use Settings, By School Level (2013-14)

	Overall (n= 80)	Elementary School (n=39)	Middle School (n=41)
Regular Classroom (w/ available computers)	15%	15%	15%
Computer Lab	40%	33%	46%
Combination of Settings (blended approach)	45%	51%	39%

Source: Spring 2014 Interviews with campus-level staff.

Table D18. Monitoring TTM Use, By Usage, and By School Level (2013-14)

	Overall (n=80)	High Usage (n=30)	Moderate Usage (n=29)	Low Usage (n=21)	Elementary School (n=39)	Middle School (n=41)
Monitored TTM Usage (overall)	79%	90%	76%	67%	79%	79%
Monitored TTM Usage (by administrators)	9%	10%	10%	5%	13%	9%
Monitored TTM Usage (by teachers)	16%	17%	17%	14%	21%	16%
Monitored TTM Usage (by teachers and administrators)	26%	30%	17%	0%	21%	26%
Monitored TTM Usage (student progress on system)	36%	47%	21%	43%	28%	43%

Source: Spring 2014 Interviews with campus-level staff.

Note: Percentages do not total to 100% because there are different Ns (between whether or not they monitored usage and, if they did, how they monitored).

Table D19. Major Themes Related to Changes in TTM Usage, Overall and By Usage Patterns, 2013-14

	Overall (n= 62)	High Usage (n= 30)	Moderate Usage (n= 18)	Low Usage (n= 14)
Used TTM consistently throughout the year	63%	83%	46%	46%
Implemented TTM with fidelity	54%	80%	39%	33%

Source: Texas SUCCESS TTM Interviews, 2014.

Appendix E– Study Limitations

Students were not randomly assigned to participate in either Istation Reading or Think Through Math (TTM). In the absence of random assignment, observed, estimated differences in the outcome of interest between participants and non-participants can be attributable to a number of factors. Some of these differences are attributable to characteristics that are amenable to measurement and can be adjusted for with multivariate regression models, including ethnicity, academic aptitude captured by student test scores, or whether the student is an English Language Learner (ELL). Others are not, which threatens the internal validity—that is, the confidence in the *estimate* of the effect of program participation on the outcome representing the *true* effect of program participation—of inferences drawn from the statistical models estimating the effect of program participation. Below, we list a number of limitations that should be kept in mind when reviewing the results from the evaluation.

1. Unmeasured teacher quality

The research team had no information about the teachers to which students were assigned during the period included in this evaluation. This is a potentially serious source of omitted bias, since system usage and usage intensity may be correlated with teacher attributes, including quality. If students assigned to teachers from the bottom portion of the teacher quality distribution are more likely to use either SUCCESS program, this non-random sorting could mistakenly attribute lower student test score gains among students who use the system to program participation, rather than to their teachers.

2. Missing information about the types of supplemental instruction or interventions students received

Schools and districts implement a plethora of interventions and supplementary services to improve their students' academic outcomes. The research team did not have any systematic information on the other types of supplementary instruction or services participants and non-participants received. This is important, particularly since the assumption underpinning the research design and multivariate analyses is that the difference in outcomes between participant students and non-participants represents the difference between students who use a SUCCESS program compared to students under the "business as usual" condition, or those students who received the typical assortment of program supports and interventions that are available to students who were not SUCCESS participants. This assumption may not hold if, for instance, students who are assigned to use a SUCCESS program are also given a number of other interventions that may neutralize, or complement, the effect of either SUCCESS intervention on student performance. Thus, the estimate of the effect of program participation may be comprised of a number of other interventions that are unmeasured in the evaluation.

3. Unmeasured differences between participating and non-participating students

Despite best efforts, including comparing within-student changes in performance between participating and non-participating students while controlling for other fixed and varying student-level characteristics, supplementing this design with propensity score reweighting based on observable characteristics, and confining the analytic sample to campuses with registered students, no guarantee can be made that participants and non-participants are identical with the exception of their exposure to the SUCCESS program. This is a fundamental, and unavoidable, challenge confronting any attempts to draw inferences about the effect of a social phenomenon (such as an academic intervention) using observational data where students were not randomized to receive, or not receive, treatment. If these unmeasured, or omitted, factors are correlated with program participation or the outcome, the estimates of the effect of program intervention are biased. See Gelman and Hill (2007) and Angrist and Pischke (2009) for accessible discussions of this source of bias.

4. Error in the measure of student participation in Texas SUCCESS during the 2012-13 and 2013-14 school years

Program participation and usage data were obtained from both Istation and TTM for the 2012-13 school year. However, school district staff were not required to use Unique IDs for students who were uploaded to each vendor's registration system until the 2013-14 school.[106] Consequently, the match rate between TEA administrative records and the registration and usage information from each vendor was weaker in 2012-13 compared to 2013-14, and it varied systematically between vendors, and across grade levels. Thus, students who participated in 2012-13 but who did not have a Unique ID in the Istation and TTM systems would not be identified as having participated in 2012-13. This measurement error will produce attenuation bias in the estimates of the effect of 2012-13 Istation participation on the outcome. It is important to note that the teacher enrollement platform used for TTM was the primary reason for the poor student match rate for that system.

5. Imprecision in Istation dosage measure

Exposure to, and utilization of, TTM is manifested in the number of lessons a student attempted and passed in a defined period of time. This measure directly quantifies students' exposure to the content and assessments that comprise lessons within the system with a great deal of precision. The dosage metric for Istation, however, is murkier and less precise since it was impossible to determine *what occurs* and *how a student performs* within or across curriculum sessions. For instance, some students, even after adjusting for prior academic performance and other observable characteristics, may move more slowly through the curriculum, which conflates system *usage* or *dosage* with a number of other student-level characteristics that may, also, be correlated with student test performance, including their familiarity and comfort with computers and online programs, their general level of engagement or disengagement, classroom distractions, or inattentive or busy teachers who are not able provide assistance quickly to help students who struggle. All of these intrinsic and extrinsic factors may contribute to increased time spent in the system and may be confounded with student test performance.

[106] The unique ID was not available until after the 2012-13 PEIMS submission 1 in February/March 2013. The unique ID was only required in 2013-14 when it became available to school district staff.

Appendix F – Campus Staff Interview Protocol

Texas SUCCESS Program

Campus Mathematics and ELA Lead Interventionist Telephone Interview

Hi, my name is _____ and I am part of the research team at Gibson Consulting Group. We are conducting a study for the Texas Education Agency looking at the use of the two online programs that are part of the Texas SUCCESS Initiative, Istation Reading (Istation) and Think Through Math (TTM).

I am trying to get in touch with the person on your campus who is responsible for coordinating academic interventions for struggling students at your campus, and for overseeing the use of (Istation Reading/Think Through Math). Am I speaking with the correct person?

[If no then as who would be the correct person and get their number.]

Part of this study includes interviews with staff who are responsible for reading/mathematics interventions, and the implementation of (Istation/TTM) on their campuses. We would like to find out how you use (Istation/TTM) on your campus, how it's working and if it's helping kids improve in reading and mathematics on your campus.

Your participation is voluntary and you can stop the interview at any point. Your responses are confidential to the extent permitted by law, and no individuals will be identified by name in the reporting of study findings. Only aggregate results will be shared.

The interview will take approximately 30-40 minutes to complete. Do you have time to do that now?

[If no, then schedule a time and get their contact info (email) to send a reminder. Send an email to them immediately with the date/time/number you will call and all of your contact information.]

Thanks for agreeing to participate in this study, I know this is a very busy time for you. Like other telephone interviews, I will be entering what you tell me in the computer as we talk. I would also like to record this if you don't mind so if I miss anything you say I can go back to it. Is that OK?

[If no then just tell the person that you may have to ask them to speak more slowly or to repeat so you get it all.]

Although we are recording and taking your answers individually, your name and individual school will not be shared as part of the study. We will be describing in the report, aggregate data only, so for example, 50% of schools used Istation/TTM this way. Does that make sense? Ok let's start.

Interviewer Note:

If interviewing reading intervention lead, use Istation/Reading verbiage in questions and if interviewing mathematics intervention lead use TTM/Mathematics verbiage in questions where applicable.

Introduction

1. Describe your role at the campus.

   Probe: What is your title?

   Probe: Are you a reading interventionist, mathematics interventionist, or both? If you are not an interventionist, what is your role?

   Probe: Does your role involve coordinating the implementation of TTM/Istation?

[If there is a separate person on the campus for the other program, get their contact info here]

2. In addition to TTM/Istation, does your campus currently use any other math/reading interventions or other math/reading online learning programs? If so, which programs?
3. To your knowledge, how many years has TTM/Istation been used at this campus?
   a. When did students start using the TTM/Istation program at your campus during the 2013-14 school year (i.e., month/year)?
   b. Are students still using the TTM/Istation program or has usage stopped for the year?
   c. Were students using TTM/Istation consistently throughout the 2013-14 school year (that is, TTM/Istation was made available to students throughout the school year)? If not, what was the usage pattern like?

Implementation

4. In which grades are students using TTM/Istation?

   3, 4, 5, 6, 7, 8 [Ask only the grades included in this campus level]

   [Note: If they indicate that they are using Istation with Grades K-2, ask if they use the full program or just the ISIP for early reading assessment purposes.]

5. In what ways are students targeted for participation in TTM/Istation (Interviewer Note: Read each response option below out loud):
   a) By grade level at your campus? If so, for which grades does it differ and why? (ask about Grades 3, 5 and 8 specifically)
   b) By groups or subpopulations? If so, and why?
   c) In response to poor prior performance on STAAR exams (from previous years)?
   d) Based on student performance on interim assessments (e.g., BOYs, MOYs, etc.)

e) At the classroom level?

6. Please describe the settings in which TTM/Istation are delivered?

   Probe for: Where delivered (e.g., in regular classroom, in computer lab)?

   Probe for: Instructional setting (e.g., part of regular class, as a pull out for intervention or remediation; as part of your afterschool offerings; as homework)

7. (If TTM/Istation was in place prior to 2013-14 – Intro Q3) Has the manner in which TTM/Istation is implemented at your campus changed since you first starting using the program in 2012-13 or earlier? If so, please explain how the program is being implemented differently in 2013-14.

8. At your campus, what is targeted number of hours on TTM/Istation for students? Does this vary for different student groups or grades?

9. Is a process in place at your campus to monitor usage of the TTM/Istation program? If so, please explain the monitoring system.

10. Please describe the professional development and/or technical assistance you have received for TTM/Istation.

    a. How was this professional development or technical assistance delivered? (e.g., by ESC staff, technical assistance vendor)

    b. Is the professional development or technical assistance you received sufficient?

    c. What other types of professional development or technical assistance would you like to receive?

11. In your opinion, has the TTM/Istation program been implemented or used with fidelity at your campus? Please explain.

12. What have you found to be important in the effective implementation of TTM/Istation at your campus?

13. Have you encountered any barriers to implementing TTM/Istation at your campus? If so, please describe any barriers you encountered.

Resources

14. Besides students' teachers, are other school staff assisting students with TTM/Istation at your school (e.g., educational aides, interventionists)? If so, how effective are they at helping students?

15. Do you have enough of the following to effectively implement the TTM/Istation intervention at your campus Istation (Interviewer Note: Read each response option below out loud)?

    a) Instructional staff

    b) Educational aides

    c) Desktop and/or portable computers

   d) Internet connectivity

16. Does your school allow students to check out portable computers so they can use TTM/Istation at home?

Coordination or Availability of other SSI Services

We understand that Texas SUCCESS online programs may not be the only way you are supporting struggling mathematics and reading students at your campus. Now I am going to ask you about other programs or services you might use with your grade struggling students to meet the grade promotion requirement Student Success Initiative (SSI).

Interviewer Note: If the respondent asks about what the SSI requirements are, please use the following:

As part of the Student Success Initiative, students in Grades 3, 5, and 8 are required to meet state standards on the STAAR-Reading and mathematics exams, or through the action of a grade placement committee, in order to be promoted to the next grade level.

17. In what ways (if any) do you feel the various SSI-related interventions provided to struggling students at your campus may be impacting academic performance in mathematics and reading?
18. Is TTM/Istation sufficient for helping struggling students on your campus or do you use other math/reading support services or programs to help struggling students at your school?

    (Note to Interviewer: Responses may include the following – Small group/direct teach sessions with interventionists, Accelerated Instruction, Other online learning programs such as Success Maker, 21st Century Community Learning Centers afterschool programs)

19. Do you coordinate services and programs funded through other sources with the TTM/Istation interventions? If so, how?

    Probe: Specifically, what resources and interventions does your campus use to address the SSI grade promotion requirements?

20. (if yes to Q1 or Q2) What are the challenges with coordinating interventions for students?
21. For students participating in TTM/Istation, how do you determine which other interventions are appropriate for them?

Satisfaction with and Perceived Impact of Istation/TTM

22. In general, how satisfied are you with the TTM/Istation?
    a) Is it user friendly?
    b) Does it have appropriate content for your students?
    c) Does it deliver the content in an engaging manner?
    d) Level of support (PD/technical assistance) from vendors?

e) Online help function or technical phone support?

23. How (if at all) would you modify the TTM/Istation program to meet the needs of your students?
24. In what ways (if any) do you feel student participation in TTM/Istation may be impacting their academic performance?

    Probe for: students in Grades 3, 5 and 8?

    Probe for: students previously retained in grade?
25. In what ways (if any) do you feel student participation in TTM/Istation may be impacting their confidence in reading/math?

    Probe for: students in Grades 3, 5 and 8?

    Probe for: students previously retained in grade?
26. Do you feel that there are any student groups or grade levels that may be benefitting more from participation in TTM/Istation than others?

Those are all of the questions I have for you. Is there anything else that I have not asked you about that you think I need to know about the implementation of Istation/TTM at your campus?

Thank you so much for your time. You have my contact information if you need to add anything or ask me about this study.

Appendix G – District Staff Interview Protocol

Texas SUCCESS Program

District Staff Telephone Interview Regarding the Coordination of ELA and Mathematics Interventions across the District

Hi, my name is _____ and I am part of a research team at Gibson Consulting Group conducting a study for the Texas Education Agency (TEA) related to the use of Istation Reading (Istation) and Think Through Math (TTM) that are part of the Texas SUCCESS Initiative.

I am trying to get in touch with the person at your district who is responsible for coordinating academic interventions for struggling students across campuses, and for overseeing the use of Istation and TTM in your district. Am I speaking with the correct person?

[If no then as who would be the correct person and get their number.]

A portion of the study we are conducting for TEA includes interviews with district staff are responsible for coordinating reading/mathematics interventions across the district, and the implementation of (Istation/TTM) at campuses within the district. We would like to learn more about why your district decided to use TTM/Istation, how it is being implemented across campuses in your district, and your level of satisfaction with the program. We are also interested in how other academic interventions are being used to complement the use of TTM/Istation and how the program(s) may help kids improve in reading and mathematics in your district.

Your participation is voluntary and you can stop the interview at any point. Your responses are confidential to the extent permitted by law, and no individuals will be identified by name in the reporting of study findings. Only aggregate results will be shared.

The interview will take approximately 30-40 minutes to complete. Do you have time to do that now?

[If no, then schedule a time and get their contact info (email) to send a reminder. Send an email to them immediately with the date/time/number you will call and all of your contact information.]

Thanks for agreeing to participate in this study, I know this is a very busy time for you. Like other telephone interviews, I will be entering what you tell me in the computer as we talk. I would also like to record this if you don't mind so if I miss anything you say I can go back to it. Is that OK?

[If no then just tell the person that you may have to ask them to speak more slowly or to repeat so you get it all.]

Although we are recording and taking your answers individually, your name and individual school will not be shared as part of the study. We will be describing in the report, aggregate data only, so for example, 50% of schools used Istation/TTM this way. Does that make sense? Ok let's start.

Introduction

1. Please describe your role at the district.

   Probe: What is your title?

   Probe: Are you responsible for coordinating reading interventions, mathematics interventions, or both?

   Probe: Does your role involve coordinating the implementation of TTM and Istation?

[If there is a separate person at the district for the other program get their contact info here]

2. In addition to TTM and Istation, does your district currently use any other math/reading interventions or other math/reading online learning programs? If so, which programs?
3. To your knowledge, how many years has your district been using TTM and Istation?

Implementation

4. Is the decision to use one or both of these interventions (TTM for mathematics or Istation for reading) centralized at the district level or decentralized to the campus level?
5. (Ask if it is centralized, district decision) Why did the district decide to use TTM or Istation as online learning interventions?

   Probe: Is it being used as one of the district's intervention to ensure students are meeting SSI grade promotion requirements?
6. (Ask if decentralized) Do you routinely get feedback from campuses regarding TTM or Istation program usage, satisfaction, and effectiveness?
7. (Ask if decentralized) Does your district provide campuses with any guidance about which interventions or programs which they should utilize to address SSI grade promotion requirements, or provide them with a list of approved interventions/programs to use with struggling students? If so, please describe.
8. Are TTM or Istation being implemented district-wide or were specific schools in your district selected for the intervention?
9. (Ask If specific schools selected in Q5) Why were these specific schools selected for the TTM and Istation interventions? What criteria were used to select the schools for participation?
10. How much latitude do campuses in your district have in terms of how the TTM or Istation systems are being used (e.g., which students to target; which grade levels to target)?

11. What guidance (if any) does the district provide to your campus about how to use TTM or Istation, and how to supplement the online interventions with other academic support services?

    Probe for: Which students should be targeted?

    Probe for: Which grade levels should be targeted?

    Probe for: Suggested number of hours per week for students?

    Probe for: Making laptop or portable computers available for students to check out so they can use TTM/Istation at home?

    Probe for: What other academic services are provided to struggling students to further support mathematics and reading development?

12. Does the district coordinate training and technical assistance related to the TTM and Istation systems for campus staff, or is that done individually by each campus?
13. Please describe the professional development and/or technical assistance district staff have received for TTM and Istation.
    a. How was this professional development or technical assistance delivered? (e.g., by ESC staff, technical assistance vendor)
    b. Is the professional development or technical assistance you received sufficient?
    c. What other types of professional development or technical assistance would you like to receive?
14. In your opinion, has the TTM/Istation program been implemented or used with fidelity across campuses within your district? Please explain.
15. (Ask if Centralized or Response to Q3 is Yes) What have you found to be important in the effective implementation of TTM and Istation at campuses in your district?
16. (Ask if Centralized or Response to Q3 is Yes) Have you encountered any barriers to implementing TTM and Istation at campuses in your district? If so, please describe any barriers you encountered.

Resources

17. Do campuses in your district have enough of the following to effectively implement the TTM and Istation program?
    a) Instructional staff
    b) Educational aides
    c) Desktop and/or portable computers
    d) Internet connectivity

Coordination or Availability of other SSI Services

We understand that Texas SUCCESS online programs may not be the only way you are supporting struggling mathematics and reading students in your district. Now I am going to ask you about other programs or services you might use with your struggling students to meet the grade promotion requirements of the Student Success Initiative (SSI).

Interviewer Note: If the respondent asks about what the SSI requirements are, please use the following:

As part of the Student Success Initiative, students in Grades 3, 5, and 8 are required to meet state standards on the STAAR-Reading and mathematics exams, or through the action of a grade placement committee, in order to be promoted to the next grade level.

18. In what ways (if any) do you feel the various SSI-related interventions provided to struggling students across your school district may be impacting academic performance in mathematics and reading?
19. Can you describe other math/reading support services or programs that are provided to struggling students in your district to meet the Student Success Initiative grade promotion requirements?
20. What state or federal grants/funding streams are used to fund academic interventions/services for struggling students in your district to meet the Student Success Initiative grade promotion requirements?
21. Can you describe the process for coordinating academic interventions/services for struggling students at the district level?
    a) Are the decisions centralized at the district level or do campuses have decision making authority in how to address the needs of struggling students on their campuses?
    b) How do you determine which campus get access to which interventions/services?

Satisfaction with and Perceived Impact of Istation/TTM

22. Overall, how satisfied are you with the TTM and or Istation systems?
    f) How satisfied are you with the level of support (PD/technical assistance) received from TTM and Istation vendors?
23. In what ways (if any) do you feel student participation in TTM/Istation may be impacting their academic performance?
24. Are you aware of any gaps or limitations with TTM or Istation programs that require supplemental interventions by other types of programs - either online or direct teach programs?

Those are all of the questions I have for you. Is there anything else that I have not asked you about that you think I need to know about the implementation of Istation/TTM in your district?

Thank you so much for your time. You have my contact information if you need to add anything or ask me about this study.

Appendix H – References

Agodini, R., & Dynarski, M. (2004). Are experiments the only option? A look at dropout prevention programs*. Review of Economics and Statistics*, *86*(1), 180-194.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, *119*(1), 249-275.

Figlio, D. N., & Rouse, C. E. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics*, *90*(1-2), 239-255.

Gelman, A. & Hill, J. (2007*). Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality*. NBER Working Paper 11154. National Bureau of Economic Research, Inc.

Heinrich, C. J., Meyer, R. H., & Whitten, G. (2010). Supplemental education services under No Child Left Behind: Who signs up, and what do they gain? *Educational Evaluation and Policy Analysis, 32*(2), 273-298.

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives, 16*(4), 91-114.

Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *Boston College Department of Economics, Statistical Software Components*. Retrieved from http://ideas.repec.org/c/boc/bocode/s432001.html

Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine, 23*(19), 2937-2960.

Mathes, P. (2014). *Istation's Indicators of Progress (ISIP) Advanced Reading: Technical manual*. Dallas, TX: Istation. Retrieved from http://www.istation.com/Content/downloads/studies/ar_technical_report.pdf

Nichols, A. (2007). Causal inference with observational data. *Stata Journal*, *7*(4), 507-541.

Patarapichayatham, C. (2014). *Istation Reading growth study grades 1 – 8*. Dallas, TX: Istation. Retrieved from http://www.istation.com/Content/downloads/studies/G1-8_TX_Growth.pdf

Patarapichayatham, C., Fahle, W., & Roden, T. R. (2014). *ISIP Reading versus STAAR Reading: The predictability study*. Dallas, TX: Istation. Retrieved from http://www.istation.com/Content/downloads/studies/Istation_ISIP_vsSTARR_Predictability_Study.pdf

Peterson, P. E., & Chingos, M. M. (2009). Impact of for-profit and nonprofit management on student achievement: The Philadelphia intervention, 2002-2008. *Program on Education Policy and Governance Working Paper Series*, PEPG 09-02. Retrieved from http://files.eric.ed.gov/fulltext/ED504449.pdf

Reback, R. L. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, *92*(5-6), 1394-1415.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688-701.

Springer, M. G., Pepper, M. J., & Ghosh-Dastidar, B. (2009). *Supplemental educational services and student test score gains: Evidence from a large, urban school district*. Working paper. Nashville: Vanderbilt University. Retrieved from http://www.vanderbilt.edu/schoolchoice/documents/supplemental_services.pdf

Taylor, L. L., Gronberg, T. J., Jansen, D. W., & Karakaplan, M. U. (2014). *Anticipating the consequences of school district consolidation in major metropolitan areas: A simulation based on cost function analysis*. Prepared for The University of Texas at Dallas Education Research Center and the Texas Education Agency. Retrieved from http://tea.texas.gov/Reports_and_Data/Program_Evaluations/Research_Reports/Program_Evaluation__Research_Reports/

Wilde, E. T., & Hollister, R. (2007). How close is close enough? Evaluating propensity score matching using data from a class size reduction experiment. *Journal of Policy Analysis and Management, 26*(3), 455-477.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Zimmer, R., Gill, B., Razquin, P., Booker, K., & Lockwood III, J. R. (2007). *State and local implementation of the No Child Left Behind Act: Volume I—Title I school choice, supplemental educational services, and student achievement*. Report to the U.S. Department of Education, Office of Planning, Evaluation and Policy Development. Retrieved from https://www2.ed.gov/rschstat/eval/choice/implementation/achievementanalysis.pdf