

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

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.

Copyright © Notice The materials are copyrighted © and trademarked ™ as the property of the Texas Education Agency (TEA) and may not be reproduced without the express written permission of TEA, except under the following conditions:

- 1) Texas public school districts, charter schools, and Education Service Centers may reproduce and use copies of the Materials and Related Materials for the districts' and schools' educational use without obtaining permission from TEA.
- 2) Residents of the state of Texas may reproduce and use copies of the Materials and Related Materials for individual personal use only without obtaining written permission of TEA.
- 3) Any portion reproduced must be reproduced in its entirety and remain unedited, unaltered and unchanged in any way.
- 4) No monetary charge can be made for the reproduced materials or any document containing them; however, a reasonable charge to cover only the cost of reproduction and distribution may be charged.

Private entities or persons located in Texas that are **not** Texas public school districts, Texas Education Service Centers, or Texas charter schools or any entity, whether public or private, educational or non-educational, located **outside the state of Texas** *MUST* obtain written approval from TEA and will be required to enter into a license agreement that may involve the payment of a licensing fee or a royalty.

For information contact: Office of Copyrights, Trademarks, License Agreements, and Royalties, Texas Education Agency, 1701 N. Congress Ave., Austin, TX 78701-1494; phone 512-463-9270 or 512-936-6060; email: copyrights@tea.texas.gov.

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

In 2013-14, the vast majority (87%) of students in Grades 3-8 across the state were registered to use the Istation

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

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.

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.

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

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.

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.

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.

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.

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

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

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

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

between Istation use by students in close proximity to the STAAR test and STAAR-Reading gains. This finding held across all grade levels.

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

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.

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

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.

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.

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.

TTM Student Outcomes

TTM Lesson Performance

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.

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.

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.

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

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.

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

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.

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

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.

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

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.