Addendum #1 to the Texas SUCCESS Comprehensive Evaluation Report

for the Texas Education Agency

Addendum #1 to the Texas SUCCESS Comprehensive Evaluation Report

Contributing Authors

Marshall Garland
Joseph Shields
Eric Booth

Submitted By:

GIBSON
CONSULTING GROUP

Table of Contents

GIBSON
CONSULTING GROUP

Acknowledgements

GIBSON
CONSULTING GROUP

List of Tables

GIBSON
CONSULTING GROUP

List of Figures

GIBSON
CONSULTING GROUP

List of Acronyms Used in this Report

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

Education Service Center (ESC)

Gibson Consulting Group, Inc. (Gibson)

Istation Reading (Istation)

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

Over the 2012–13 to 2014–15 school years, the Texas Students Using Curriculum Content to Ensure Sustained Success (SUCCESS) program offered 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.

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. A few research questions that could not be answered because of data availability in the initial comprehensive evaluation report (published by TEA in January 2015) will be addressed in this addendum report.

This addendum to the Texas SUCCESS Comprehensive Evaluation Report addresses the following two research questions related to Istation and TTM:[1]

1) Among students who were retained in Grade 5 or 8 between 2012–13 and 2013–14, to what extent does 2013–14 academic performance differ between Texas SUCCESS participants and non-participants, and what is the relationship between intensity of usage (i.e., dosage) and academic performance for this student population?

2) What is the relationship between Texas SUCCESS program usage and intensity of usage in 2013–14 and the probability of being retained in Grades 5 and 8 between 2013–14 and 2014–15?

Grade 5 and Grade 8 were selected for these analyses because they are grade levels in which students are subject to high-stakes state assessments and are required to pass the State of Texas Assessments of Academic Readiness (STAAR)-Reading and STAAR-Mathematics exams to be promoted to the next grade level.[2] In addition, the research questions posed in this addendum are geared toward students in at-risk situations (i.e., students retained in grade between 2012–13 and 2013–14, and students at-risk of being retained in grade between 2013–14 and 2014–15).

---

[1] The Texas SUCCESS Comprehensive Evaluation report can be accessed at
http://tea.texas.gov/Reports_and_Data/Program_Evaluations/Reading_Math_Science_and_Technology_Initiatives/Program_Evaluation__Reading,_Math,_Science,_and__Technology_Initiatives/

[2] It is important to note that if a student does not pass the STAAR exam after three testing opportunities in Grades 5 and 8, the student may still be promoted to the next grade through a grade placement committee decision (Texas Education Code (TEC) 28.0211(e)).

GIBSON
CONSULTING GROUP

Key Findings

Reading

Istation is a supplemental reading program that provides computer-adaptive instruction in an animated environment that is designed to improve phonemic awareness, alphabetic knowledge, vocabulary, and reading comprehension. Although Istation includes curricular materials for Grades Pre-Kindergarten through 8, it was 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 use Istation curriculum for a minimum of 250 minutes and middle school students use the system for a minimum of 200 minutes throughout the school year.

Istation Student Outcomes

1) *Among students who were retained in Grade 5 or 8 for 2013–14, to what extent does 2013–14 academic performance differ between Istation participants and non-participants, and what is the relationship between intensity of usage (i.e., dosage) and academic performance for this student population?*

After examining Istation system usage patterns for Grades 5 and 8 students retained in grade between 2012–13 and 2013–14 and students not retained, Gibson explored the relationship between Istation usage and changes in student performance on STAAR-Reading for students in Grades 5 and 8 who were retained in grade between 2012–13 and 2013–14.[3] Analyses were confined to students in Grades 5 and 8 in 2013–14 who were retained in grade between 2012–13 and 2013–14 and to students who were enrolled in schools where Istation was used in 2013–14.

The first research questions explored the relationship between both Istation usage (i.e., whether a student used the system for one minute or more) and usage intensity (i.e., whether a student used the system for the recommended number of minutes, or an even higher threshold of usage), and STAAR-Reading decile-standardized gains (hereinafter referred to as STAAR-Reading gains).[4] While the primary research question explored the relationship between whether students retained in grade used Istation at all and STAAR-Reading gain scores, it is also important to understand if higher levels of system usage (i.e., usage intensity) are related to improved academic performance. Therefore, analyses were conducted for the following categories of Istation system usage:

---

[3] A student is classified as having been retained if their grade of enrollment in fall 2012 matched their grade of enrollment in fall 2013.

[4] Decile-standardized gain scores were used to control for prior STAAR performance differences among students included in the analyses. Gain scores reflect differences between 2012–13 and 2013–14 STAAR scale scores. This is explained in more detail in the Introduction section of this report.

GIBSON
CONSULTING GROUP

1) Students who used the system at least one minute (i.e., used the system at all) compared to students who did not use the system;

2) Students who met the vendor usage recommendation for the corresponding grade level (i.e., 250 minutes for Grade 5 and 200 minutes for Grade 8) compared to students who did not meet the recommended usage threshold; and

3) Students who used the system 300 minutes or more (i.e., a measure of high intensity usage) compared to students who did not use the system.

Key findings from these analyses are as follows:

Student Population and Istation System Usage

A relatively small proportion of students at campuses where Istation was used in 2013–14 were students who were retained in Grades 5 and 8 between 2012–13 and 2013–14, resulting in a small, high-need student population included in outcomes analyses presented in this report.

▪ Students who had been retained in grade from 2012–13 to 2013–14 accounted for 1.4% (n=5,216) of Grade 5 students and .94% (n=3,567) of Grade 8 students who used the system at Istation campuses in 2013–14. Therefore the subsequent analysis is on a small, high need population.

Students in Grades 5 and 8 who were retained in grade between 2012–13 and 2013–14 used Istation at higher levels during the 2013–14 school year than students who were not retained in grade during this period.

▪ On average, Grade 5 students who were retained in grade recorded almost an hour more Istation usage over the course of the 2013–14 school year (239 minutes compared to 182 minutes) than students who were not retained.

▪ System usage levels for students in Grade 8 were relatively low compared to elementary school students; however, Grade 8 students who were retained in grade between 2012–13 and 2013–14 recorded an average of 49 minutes of Istation usage in 2013–14 compared to 38 minutes for students who were in Grade 8 for the first time in 2013–14.

Reading Outcomes

No statistically significant relationship was found between the use of Istation by Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 and STAAR-Reading gains in 2013–14.

▪ After adjusting for other student, school, and district-level characteristics, there were no statistically significant improvements in 2013–14 STAAR-Reading score gains for retained Grade 5 and Grade 8 students when the results of Istation users were compared to those of nonusers.

GIBSON
CONSULTING GROUP

No statistically significant relationship was found between intensive usage (e.g., 300 or more minutes over the course of the 2013–14 school year) of Istation by Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 and STAAR-Reading gains in 2013–14.

- After adjusting for other student, school, and district-level characteristics, no statistically significant differences in 2013–14 STAAR-Reading gains were observed among Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 who used the system for 300 or more minutes and those who did not use the system at all.

While it was fairly rare for students to be retained in grade between 2012–13 and 2013–14, and for students to use Istation at the recommended levels (i.e., 250 minutes for Grade 5 students and 200 minutes for Grade 8 students) during the 2013–14 school year, when these thresholds were reached, Grade 8 students had significantly larger reading gains than students who did not reach the recommended threshold level. The opposite effect was observed for Grade 5 students.

- Retained Grade 8 students who used Istation for 200 or more minutes during 2013–14 had statistically significant, moderately larger reading gains (.29 standard deviations) than retained students who used the system at levels below the recommended usage threshold.[5]

- Contrary to the finding for retained Grade 8 students, among Grade 5 students, students who used Istation at or above the annual level of 250 minutes or more had smaller STAAR-Reading improvements when compared to Grade 5 students who did not use the system at the recommended level; the relationship was statistically significant.

2) What is the relationship between Istation program usage and the probability of being retained in Grades 5 and 8 between 2013–14 and 2014–15?

This research question assessed the relationship between Istation usage and usage intensity in 2013–14 and the probability of being retained in Grade 5 or 8 for 2014–15. Grade retention was determined by comparing students' fall 2013 grade of enrollment to their fall 2014 grade of enrollment: if the grade levels matched, students were flagged as having been retained in grade. If a student advanced grade levels between 2013–14 and 2014–15, the student was classified as having been promoted.

Analyses were disaggregated by both grade level and by students' risk of being retained, where students were classified as at risk of being retained if they had failed a STAAR-Reading assessment in 2011–12 (two years prior to 2013–14) or 2012–13 (one year prior to 2013–14).

After adjusting for other student, school, and district-level characteristics, no statistically significant differences in the probability of being retained in grade were found between participating and non-participating students.

---

[5] Only 234 (6.5%) Grade 8 students retained in grade between 2012–13 and 2013–14 used Istation for 200 or more minutes in 2013–14.

GIBSON
CONSULTING GROUP

- This finding was consistent regardless of grade level or a students' at-risk status, and held across all measures of Istation usage and usage intensity (i.e., dosage).

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 mathematics concepts and problem-solving skills. TTM includes instructional materials that cover mathematics content for Grades 3–8 and Algebra I. TTM was 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 to performance on quizzes given at the completion of lessons. The TTM vendor suggests that students attempt a minimum of 5 lessons but recommends students attempt 10 or more lessons.

TTM Student Outcomes

1) *Among students who were retained in Grade 5 or 8 between 2012–13 and 2013–14, to what extent does 2013–14 academic performance differ between TTM participants and non-participants, and what is the relationship between intensity of usage (i.e., dosage) and academic performance for this student population?*

After examining TTM usage patterns for Grade 5 and 8 students retained for 2013–14 and students not retained, Gibson explored relationship between TTM usage and changes in student performance on STAAR-Mathematics for students in Grades 5 and 8 who were retained in grade between 2012–13 and 2013–14.[6] Analyses were confined to students in Grades 5 and 8 in 2013–14 who were retained in grade between 2012–13 and 2013–14 and to students who were enrolled in schools where TTM was used in 2013–14.

The primary research question explored the relationship between students retained in grade between 2012–13 and 2013–14 who attempted one or more TTM lessons and 2013–14 STAAR-Mathematics gain scores for students; however, it is also important to understand if higher levels of system usage (i.e., usage intensity) are related to improved academic performance. Therefore, to fully address the first research question, analyses were conducted for the following categories of TTM system usage:

1) Students who attempted at least one TTM lesson (i.e., used the system at all) compared to students who did not use the TTM system at all;

2) The number of TTM lessons a student attempted (i.e., a continuous variable of attempted lessons);

---

[6] A student is classified as having been retained if their grade of enrollment in fall 2012 matched their grade of enrollment in fall 2013.

GIBSON
CONSULTING GROUP

3) Students who attempted less than the minimum number of five TTM lessons based on vendor recommendation (i.e., 1 to 4 lessons attempted) compared to students who did not attempt any lessons; and

4) Students who attempted either 5 to 9 lessons, 10 to 14 lessons, 15 to 19 lessons, or 20 or more lessons (i.e., four measures of increasingly high intensity usage) compared to students who did not attempt any lessons.

Key findings from these analyses are as follows:

Student Population and TTM System Usage

A relatively small proportion of students at campuses where TTM was used in 2013–14 were retained in Grades 5 and 8 between 2012–13 and 2013–14, resulting in a small, high-need student population included in outcomes analyses presented in this report.

- Students who had been retained in grade from 2012–13 to 2013–14 accounted for approximately 1.5% (n=5,224) of Grade 5 students and 1.0% (n=3,591) of Grade 8 students who used the system at TTM campuses in 2013–14. Therefore the subsequent analysis is on a small, high need population.

Students in Grades 5 and 8 who were retained in grade between 2012–13 and 2013–14 used TTM less frequently during the 2013–14 school year than students who were not retained in grade during this period.

- Approximately 43% of Grade 5 students who were retained in grade between 2012–13 and 2013–14 attempted at least one TTM lesson versus 46% of students who were not retained. The gap in percentage of students using the TTM system at the minimum recommended usage level (i.e., 5 attempted lessons) decreased by one percentage point for Grade 5 students who were retained in grade (36%) compared to students who were not retained (37%).

- On average, Grade 5 students who were retained in grade attempted a comparable number of TTM lessons throughout the course of the 2013–14 school year to students not retained (9.26 versus 9.77).

- TTM system usage levels for student in Grade 8 were relatively low compared to Grade 5 students. Approximately 19% of retained Grade 8 students attempted at least one TTM lesson compared to 25% of first-time Grade 8 students. Likewise, for students who attempted 5 or more TTM lessons, 10% of retained Grade 8 students and 15% of first-time Grade 8 students used at this level.

- On average, students in Grade 8 who were retained in grade between 2012–13 and 2013–14 attempted slightly fewer mathematics lessons in 2013–14 (2.70 TTM lessons) than students who were in Grade 8 for the first time in 2013–14 (3.79 TTM lessons).

GIBSON
CONSULTING GROUP

Mathematics Outcomes

No statistical relationship was found between the use of TTM by Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 and STAAR-Mathematics gains in 2013–14.

- After adjusting for other student, school, and district-level characteristics, there were no statistically significant differences in 2013–14 STAAR-Mathematics score gains for retained Grade 5 and Grade 8 students when the results of students who attempted at least one TTM lesson were compared to those of students who did not attempt any TTM lessons.

Among students in Grades 5 and 8 who were retained in grade between 2012–13 and 2013–14, a positive and statistically significant relationship was observed between the number of TTM lessons attempted and 2013–14 STAAR-Mathematics gains.

- The difference was statistically significant and moderate (.943 for Grade 5 and .905 for Grade 8) for each additional lesson attempted.

- This finding is slightly different from findings in the Texas SUCCESS Comprehensive Evaluation Report (Garland, Shields, Booth, Shaw, Samii-Shore, 2015), which showed that the relationship between being at-risk of being retained in grade and TTM usage was associated with smaller STAAR-Mathematics gains.

Among students in Grade 5 who were retained in grade between 2012–13 and 2013–14, a positive and statistically significant relationship was observed between attempting 1 to 4 TTM lessons and 2013–14 STAAR-Mathematics gains.

- After adjusting for other student, school, and district-level characteristics, Grade 5 students who attempted between one and four lessons experienced statistically significant differences relative to students who did not use the system (.174 standard deviations higher).

- This relationship was not observed for Grade 8 students.

Increasingly intensive TTM usage over the course of the 2013–14 school year by retained Grade 5 and 8 students was not related to larger STAAR-Mathematics gains in 2013–14 when compared to students who did not use the system at all.

- After adjusting for other student, school, and district-level characteristics, no statistically significant differences in 2013–14 STAAR-Mathematics gains were observed among Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 who used TTM at increasingly intensive levels (5 to 9 attempted lessons, 10 to 14 attempted lessons, 15 to 19 attempted lessons, or 20 or more attempted lessons) relative to students who did not use the TTM system at all.

GIBSON
CONSULTING GROUP

2) *What is the relationship between TTM program usage and the probability of being retained in Grades 5 and 8 between 2013–14 and 2014–15?*

The evaluation team explored the relationship between TTM usage in 2013–14 and the probability of being retained in Grade 5 or 8 between 2013–14 and 2014–15. Analyses were disaggregated by both grade level and by students' risk of being retained in grade, where students were classified as being at risk of being retained in grade if they had failed a STAAR-Mathematics assessment in 2011–12 (two years prior to 2013–14) or 2012–13 (one year prior to 2013–14).

After adjusting for other student, school, and district-level characteristics, there were a few TTM usage and intensity of usage variables that were statistically significantly associated with retention outcomes.

- In Grade 5, students who attempted at least one lesson in 2013–14 were significantly less likely to be retained in grade between 2013–14 and 2014–15.

- For Grades 5 and 8 students, each additional attempted lesson was associated with a significantly lower likelihood of being retained in grade between 2013–14 and 2014–15. This was also the case for students in each grade who attempted greater than 20 TTM lessons.

- Grade 5 and 8 students were less likely to be retained in grade between 2013–14 and 2014–15 with high levels of TTM lesson attempts (20 or more lessons attempted).

- Each of these findings held for all first-time Grade 5 and 8 students, as well as first-time Grade 5 and 8 students who were at risk of being retained in grade between 2013–14 and 2014–15.

As was the case with reading-related findings, it is important to keep in mind that the population of students who met the usage criteria, and who had sufficient data to be included in the statistical model, was small (i.e., only 367 retained Grade 8 students attempted 5 or more TTM lessons) and observably different from the overall student population. For instance, in Grade 8, fewer than 3,000 students were retained in grade between 2012–13 and 2013–14. Consequently, the results should be interpreted with caution.

Additional limitations to the analyses can be found in Appendix C.

GIBSON
CONSULTING GROUP

Section 1 – 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 the Texas Education Agency (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.[7]

- 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 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.

---

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

GIBSON
CONSULTING GROUP

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.[8] 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

Over the 2012–13 to 2014–15 school years, the Texas Students Using Curriculum Content to Ensure Sustained Success (SUCCESS) program offered state-funded access to interactive web-based 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.[9] 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. 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 was determined and documented.[10]

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 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

---

[8] 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

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

[10] For further detail related to the correlation between Istation and TTM content and the 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

versions are not required by Texas SUCCESS, Istation is available in Spanish for some grade levels (K-3) and the benchmark tests for this program can also be completed in Spanish if districts choose to purchase it.[11] 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 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) 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. A few research questions that could not be answered in the initial comprehensive evaluation report (published by TEA in January 2015), due to data availability, will be addressed in this addendum report.

This addendum to the Texas SUCCESS Comprehensive Evaluation Report (Garland, Shields, Booth, and Shaw, and Samii-Shore, 2015), addresses the following two research questions related to the use of Istation and TTM by students in at-risk situations:[12]

1) Among students who were retained in Grade 5 or 8 between 2012–13 and 2013–14, to what extent does 2013–14 academic performance differ between Texas SUCCESS participants and non-participants, and what is the relationship between intensity of usage (i.e., dosage) and academic performance for this student population?

2) What is the relationship between Texas SUCCESS program usage and intensity of usage in 2013–14 and the probability of being retained in Grades 5 and 8 between 2013–14 and 2014–15?

Grade 5 and Grade 8 were selected for these analyses because they are years in which students are subject to high-stakes assessments, which they are required to pass in order to be promoted to the next grade level. In addition, the research questions posed in this addendum are geared toward students in at-risk situations (i.e., students retained in grade between 2012–13 and 2013–14, and students at-risk of being retained in grade between 2013–14 and 2014–15).

---

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

[12] The Texas SUCCESS Comprehensive Evaluation report can be accessed at http://tea.texas.gov/Reports_and_Data/Program_Evaluations/Reading_Math_Science_and_Technology_Initiatives /Program_Evaluation__Reading,_Math,_Science,_and__Technology_Initiatives/

GIBSON
CONSULTING GROUP

Analytic Methods

This addendum report expands on the prior analysis by presenting an analysis of the relationships between **Istation and TTM usage** (i.e., whether a student used the system at all) and **usage intensity** (i.e., how frequently a student used the system) and changes in performance on STAAR for students who were actually retained in Grades 5 and 8 between 2012–13 and 2013–14.

The following data were used to address the primary research questions in this addendum report:

- Online system usage data provided by the Istation and TTM vendors;
- Archival data on students contained in the Public Education Information Management System (PEIMS) from TEA; and
- Student outcomes data related to State of Texas Assessments of Academic Readiness (STAAR) performance (reading and mathematics) from TEA (2011–12 through 2013–14).

To address the student outcomes questions, the evaluation team utilized system usage 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.

To more accurately account for differences in prior performance on STAAR exams, the evaluation team calculated decile-standardized STAAR test-score gains for reading and mathematics. Decile-standardized STAAR-Reading and Mathematics test-score gains were calculated between the first test administration in 2012–13 and the first test administration in 2013–14. 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 or Mathematics 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 or Mathematics 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 or Mathematics score decile. The quantity of interest for the evaluation is the difference in the gain score between Istation and TTM 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.

Organization of the Report

Following this introductory section, Section 2 of this report explores each of the research questions related to reading (i.e., Istation), and Section 3 examines the same set of research questions for

GIBSON
CONSULTING GROUP

mathematics (i.e., TTM). Section 4 includes key findings and concluding observations from the Istation and TTM analyses.

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 a list of study limitations to keep in mind when interpreting the results of this evaluation. Appendix D contains a list of references cited in this report.

Section 2 – Reading Outcomes

While Garland, et al (2015) provided a broad assessment of Istation implementation and student outcomes, the results presented in this section focus on two key grade levels (Grades 5 and 8) and high-need students. Grade 5 and Grade 8 were selected for these analyses because students are subject to high-stakes assessments with grade placement implications in these grades. In addition, the research questions posed in this addendum are geared toward students in at-risk situations (i.e., students retained in grade between 2012–13 and 2013–14, and students at-risk of being retained in grade between 2013–14 and 2014–15).

This section of the report addresses the following two research questions related to the Istation program:

1) Among students who were retained in Grade 5 or 8 between 2012–13 and 2013–14, to what extent does 2013–14 academic performance differ between Istation participants and non-participants, and what is the relationship between intensity of Istation usage (i.e., dosage) and academic performance for this student population?

2) What is the relationship between Istation program usage and the probability of being retained in Grades 5 and 8 between 2013–14 and 2014–15?

Appendix A provides technical information related to the Istation analyses contained in this addendum report.

Guiding Question 1a: Among students who were retained in Grade 5 or 8 between 2012–13 and 2013–14, to what extent does 2013–14 academic performance differ between Istation participants and non-participants, and what is the relationship between intensity of Istation usage (i.e., dosage) and academic performance for this student population?

The descriptive analysis presented in Garland, et al (2015) found that students in Grades 5 and 8 who were at risk of being retained in grade used Istation at higher levels than students who were not at risk of being retained in grade. Being at risk for retention in grade was defined as having failed the STAAR-Reading exam in either the 2011–12 or 2012–13 school year. This section of the addendum report expands on the prior analysis by presenting an analysis of the relationships between **Istation usage** (i.e., whether a student used the system at least one minute) and **Istation usage intensity** (i.e., how frequently a student used the system) and changes in student performance on STAAR-Reading for students who were actually retained in Grades 5 and 8 between 2012–13 and 2013–14.

After examining Istation system usage patterns for Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 and students not retained, the evaluation team explored the relationship between Istation usage and changes in student performance on STAAR-Reading for students in Grades 5 and 8 who

GIBSON
CONSULTING GROUP

were retained in grade between 2012–13 and 2013–14. Analyses were confined to students in Grades 5 and 8 in 2013–14 who were retained in grade between 2012–13 and 2013–14 and to students who were enrolled in schools where Istation was used (in 2013–14).

While the primary research question explored the relationship between Istation usage (i.e., at least one minute of usage) and STAAR-Reading gain scores for students retained in grade, it is also important to understand whether higher levels of system usage (i.e., usage intensity) were related to improved academic performance. Therefore, analyses were conducted for the following categories of Istation system usage:

1) Students who used the system at least one minute (i.e., used the system at all) compared to students who used the system less than one minute;

2) Students who met the vendor usage recommendation for the corresponding grade level (i.e., 250 minutes for Grades 3–5 and 200 minutes for Grades 6-8) compared to students who did not meet the recommended usage threshold; and

3) Students who used the system 300 minutes or more (i.e., a measure of high intensity usage) compared to students who did not use the system.

Grade-level retention was calculated by comparing students' grade level in fall 2012 to their grade level in fall 2013.[13] The calculation only included students who were enrolled in Grade 5 or 8 in 2012–13. If students' grade level matched between these two periods, they were classified as having been retained in grade between the 2012–13 and 2013–14 school years.[14]

Istation Usage and Grade Retention

Table 2.1 displays the number of students (at campuses where at least one student was registered for Istation in 2013–14) in Grades 5 and 8 in 2013–14 who were identified as having been retained in grade

---

[13] The evaluation team followed TEA's operationalization of grade-level retention, with one difference. Students were classified as having been retained if their fall 2012 and fall 2013 grade level matched. The TEA definition utilizes attendance records from the baseline year (here, 2012–13) to include students who were not enrolled by the PEIMS snapshot date, where baseline grade level is determined by the last grade in which a student was enrolled (Texas Education Agency, 2015). For more information, see TEA's resources on retention here: http://tea.texas.gov/acctres/retention_index.html.

[14] Importantly, the inverse of this retention calculation does not imply promotion. That is, there are multiple reasons why a student not designated in the PEIMS data as being enrolled in the same grade level between 2012–13 and 2013–14 may not have been retained in grade. For example, the student may have either been promoted to the next grade (but was not present in the data) or they may have left the Texas public school system for a variety of reasons.

GIBSON
CONSULTING GROUP

between 2012–13 and 2013–14.[15] Overall, the percentage of Grade 5 and Grade 8 students who were retained in grade in 2013–14 was very low: approximately 1.4% of Grade 5 students and 1.0% of Grade 8 students were identified as being retained in grade between 2012–13 and 2013–14.

Table 2.1. Frequency Count and Percentage of Grade 5 and 8 Students at Istation Campuses Retained in Grade between 2012–13 and 2013–14

Student Grade Level	Total Student Count	Count of Students Retained	Percentage of Students
Grade 5	382,326	5,216	1.36%
Grade 8	378,948	3,567	0.94%

Source: Public Education Information Management System, 2012–13 and 2013–14, Texas Education Agency; Istation Session History Table, 2013–14.

Note: Frequency counts and calculations include all Grade 5 and 8 students in 2013–14 who were enrolled at a campus where at least one student was registered for Istation in 2013–14. Students denoted as retained in grade were students in Grades 5 and 8 in 2012–13 whose grade level in the fall of 2013 matched their grade level in the fall of 2012.

Table 2.2 provides the percentage and count of Grade 5 and 8 students who were in these grades for the first-time, or were retained in grade and repeating Grades 5 and 8 in 2013–14. A higher percentage of Grade 5 students who were retained in grade between 2012–13 and 2013–14 used the system for at least one minute (70% compared to 63% first-time Grade 5 students). Similarly a higher percentage of retained Grade 5 students used Istation at the recommended levels (32% compared to 24% first-time Grade 5 students) or the higher intensity level of 300 minutes (28% compared to 21% first-time Grade 5 students). Only minor differences in usage levels were observed between first-time and retained Grade 8 students.

Another important finding from Table 2.2 is the small percentage of Grade 8 students who used the system, or used the system at increasing levels of intensity. This is an important descriptive point to keep in mind when reviewing results from the forthcoming multivariate analyses.

---

[15] The retention rate calculation used in this report is slightly different than the retention rate reported by TEA, since the denominator in the calculation presented in Table 2.1 reflects all students in Grade 5 or 8 in 2013–14 who met the sample inclusion rules for the evaluation. Table 2.1 provides a description of the composition of the analytic sample that will be used in the outcomes section of this research question.

GIBSON
CONSULTING GROUP

Table 2.2. Frequency Count and Percentage of Grade 5 and 8 Students who Used Istation, by Grade-Level Retention Status and Istation Usage Intensity, 2013–14

Student Grade Level	Enrollment in 2013–14	Number and Percent of Students Who Used at Least One Minute	Number and Percent of Students Who Met Grade Level Usage Recommendation[16]	Number and Percent of Students Who Used at Least 300 Minutes[17]
Grade 5	First time	236,981 (62.8%)	91,682 (24.3%)	79,638 (21.1%)
Grade 5	Retained in grade	3,673 (70.4%)	1,657 (31.8%)	1,457 (27.9%)
Grade 8	First time	103,978 (27.7%)	20,809 (5.5%)	13,585 (3.6%)
Grade 8	Retained in grade	1,029 (28.8%)	234 (6.5%)	165 (4.6%)

Source: Public Education Information Management System, 2012–13 and 2013–14, Texas Education Agency. Istation Session History Table, 2013–14.

Note: Frequency counts and calculations include all Grade 5 and 8 students in 2012–13 who were enrolled at a campus where at least one student was registered for Istation in 2013–14. Students classified as retained in grade were students in Grades 5 and 8 in 2012–13 whose grade level in the fall of 2013 matched their grade level in the fall of 2012. First-time Grade 5 or 8 students reflect all students who were enrolled in Grade 5 or 8 for the first time in 2013–14, including students who were promoted, and students who were not enrolled in a Texas public school in the prior school year.

Istation Usage and Student Outcomes Related to STAAR-Reading Gains

First, the evaluation team explored descriptive system usage data to provide context for the student outcome findings which follow. Grade 5 and 8 students who were retained in grade between 2012–13 and 2013–14 used Istation at a higher level compared to students who were not retained in grade (Figure 2.1). In Grade 5, students who were retained in grade used the system for nearly an hour (approximately 57 minutes) more than students who were not retained in grade (239 minutes compared to 182 minutes). In Grade 8, students who were retained in grade used the system only eleven minutes more than students who were not retained in grade (49 minutes compared to 38) (Figure 2.1). *Istation recommends that students in Grade 5 use the system for at least 250 minutes per school year and Grade 8 students use the system for at least 200 minutes per school year.*

The findings in Figure 2.1 contrasts with findings presented in Garland, et al (2015) which examined the differences in usage intensity between Grade 5 and Grade 8 students who had failed STAAR-Reading in a prior school year compared to students who had not. Garland, et al (2015) found larger differences

---

[16] The recommended usage level from Grades 3–5 is 250 minutes per year, and the recommended usage level for Grades 6-8 is 200 minutes per year.

[17] The three categories of usage intensity are not mutually exclusive. That is, some of the students who used the Istation system for at least one minute also used it at the recommended level and for 300 or more minutes.

GIBSON
CONSULTING GROUP

between the student groups analyzed. For example, students in Grade 8 who were classified as having been at-risk of being retained in grade between 2013–14 and 2014–15 used the system nearly three times as many minutes (73 minutes compared to 20 minutes) relative to students who were not considered at risk of being retained in grade. Similarly, Grade 5 students who were at risk of being retained in grade between 2013–14 and 2014–15 used Istation approximately 87 minutes more compared to peers who were not at risk of being retained in grade (239 minutes compared to 152 minutes). This suggests that school staff were targeting students at risk of being retained in grade for more intensive usage of Istation relative to their peers who were not at risk of being retained in grade.

Figure 2.1. Average Number of Istation Minutes Used by Grade 5 and 8 Students in 2013–14, by 2013–14 Grade-Level Retention Status

Source: Public Education Information Management System, 2012–13 and 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. Students classified as retained in grade were students in Grades 5 and 8 in 2012–13 whose grade level in the fall of 2013 matched their grade level in the fall of 2012. First-time Grade 5 or 8 students reflect all students who were enrolled in Grade 5 or 8 for the first time in 2013–14, including students who were promoted, and students who were not enrolled in a Texas public school in the prior school year.

Student decile-standardized scores were used to better account for differences in prior STAAR performance when examining changes in STAAR-Reading scores between 2012–13 and 2013–14, and to

GIBSON
CONSULTING GROUP

account for different average changes in student performance across the distribution of students' prior achievement score.[18] As noted earlier, the different measures of system usage analyzed were:

1) Student used the system for at least one minute (compared to students below one minute);

2) Student used the system at the dosage recommended by Istation (compared to students who used Istation below the prescribed threshold); and

3) Student used the system 300 minutes or more (compared to students who did not use Istation at all).[19]

The results are presented in Table 2.3.

In Grades 5 and 8, no statistical differences in STAAR-Reading gain scores were observed between students retained in grade between 2012–13 and 2013–14 who used the system at all (i.e., one or more minutes) and students who did not use the system at all.[20] Intensive Istation usage (i.e., 300 or more minutes) was also not significantly associated with differences in STAAR-Reading gains for this at-risk student population, when compared to students who did not use Istation in 2013–14.

Differences in STAAR-Reading gain scores were observed between retained students who used the system at the vendor-recommended levels for Grades 5 and 8, but the results were opposite for the two grade levels. Among retained Grade 8 students, those who used Istation at the recommended intensity level experienced statistically significant, larger STAAR-reading gains (.29 standard deviations) than students who did not meet the recommended number of minutes of use in 2013–14. This is consistent with findings presented in Garland, et al (2015), where the association between being at-risk of being retained in grade,

---

[18] The evaluation team calculated decile-standardized STAAR-Reading test-score gains between the first test administration in 2012–13 and the first test administration in 2013–14. 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. The same approach was utilized for the calculation of decile-standardized STAAR-Mathematics test-score gains reported in Section 3 of this report.

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

[20] Throughout this section, the terms "statistically significant" and "significant" are used interchangeably. They refer to instances in which, after controlling for other factors, differences in outcomes (e.g., STAAR-Readings gain scores) are found between student groups (e.g., those who used Istation for one or more minutes compared to students who did not use the system). 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.

GIBSON
CONSULTING GROUP

and Istation usage, was positively associated with STAAR-Reading gains. Conversely, among Grade 5 students retained in grade between 2012–13 and 2013–14, Istation usage at the recommended intensity level was statistically significant and associated with smaller STAAR-Reading gains than comparable students who failed to use Istation at the recommended intensity level. The difference was small (-.16) and statistically significant (*p<0.01)* among students who met the grade level Istation dosage threshold (250 minutes) relative to students who did not.

Table 2.3. Estimates of the Differences in the Relationship between 2013–14 Istation Usage/Intensity of Usage and STAAR-Reading Gains, for Students in Grades 5 and 8 who were Retained in Grade between 2012–13 and 2013−14

Grade	Usage Measure	B	SE	N
Grade 5	Used at least one minute (compared to those below one minute)	-0.002	0.076	4,683
	300 or more minutes (compared to those who did not use the system at all)	-0.124	0.095	
	Met Istation dosage threshold (compared to those who did not meet threshold)	-0.162**	0.061	
Grade 8	Used at least one minute (compared to those below one minute)	0.151	0.103	2,234
	300 or more minutes (compared to zero)	0.313	0.160	
	Met Istation dosage threshold (compared to those who did not meet threshold)	0.292*	0.129	

Source: Istation Session History table, 2013–14, Texas Education Agency. State of Texas Assessments of Academic Readiness (STAAR) and Public Education Information Management System, 2012–13 – 2013–14, Texas Education Agency.

Note: Only regular English and Spanish versions of STAAR-Reading (i.e., not modified or alternate versions) are included in the analyses. Only students with at a valid STAAR-Reading test score on the first administration in 2012–13 and 2013–14 are included in the multivariate model. Minutes are calculated prior to the administration of the first assessment, which is contingent on grade level. The Istation dosage threshold in Grade 5 is 250 minutes, and in Grade 8 it is 200 minutes. Estimates derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates * p<0.05; ** p<0.01; *** p<0.001

It is important to interpret these findings with caution. The population of students who met the usage criteria in both Grades 5 and 8, and who had sufficient data to be included in the statistical model, was small and observably different from the overall student population.[21] For instance, in Grade 8, fewer than 300 students who were retained in grade between 2012–13 and 2013–14 either used the system for 200 minutes or more or 300 minutes or more (Table 2.2). Because of these two issues, the amount of statistical uncertainty in these estimates (which is reflected in the size of the standard error of the estimates that are presented in Table 2.3) is large.

---

[21] For instance, students who were retained in grade had a higher rate of test abstention, or ineligibility based on the STAAR-Reading inclusion rules adopted in this evaluation, than students who were not retained. That is, students who were retained in grade took the first administration of the STAAR exam at lower rates than students not retained in grade.

GIBSON
CONSULTING GROUP

Guiding Question 2a: What is the relationship between Istation program usage and the probability of being retained in Grades 5 and 8 between 2013–14 and 2014–15?

Prior to describing the relationship between Istation usage during the 2013–14 school year and whether students in Grade 5 or 8 were retained in grade between 2013–14 and 2014–15, the research team received fall 2014 PEIMS student enrollment records to calculate actual retention and promotion between 2013–14 and 2014–15.[22,23]

Grade-Level Enrollment Status

Table 2.4 provides summary descriptive information about the enrollment status of students who were first-time Grade 5 and Grade 8 students in 2013–14. Less than 1.5% of students in Grades 5 and 8 in 2013–14 were retained in grade between 2013–14 and 2014–15, with a slightly higher percentage of Grade 5 students retained in grade than Grade 8 students (1.2% compared to 0.8%). In each grade, less than 4% of students who were enrolled in Grade 5 or 8 for the first time in 2013–14 were not enrolled in Texas public schools in the fall of 2014 (3.6% in Grade 5 and 3.7% in Grade 8).

Table 2.4. Student Promotion/Retention Results between 2013–14 and 2014–15, by Grade Level

	Grade 5			Grade 8		
	Promoted	Retained in Grade	Not Enrolled in Texas Public Schools in 2014–15	Promoted	Retained in Grade	Not Enrolled in Texas Public Schools in 2014–15
Count	358,925	4,461	13,718	358,472	3,055	13,826
Percentage of Students	95.18%	1.18%	3.64%	95.50%	0.81%	3.68%

Source: Public Education Information Management System (PEIMS), 2012–13, 2013–14, and 2014–15, Texas Education Agency.

Note: Frequency counts and calculations only include students who met the following inclusion rules: enrolled at a campus where at least one student was registered for Istation in 2013–14; and was a first-time Grade 5 or Grade 8

---

[22] Only sequential grade promotion is reflected in the promotion indicator. Students promoted more than one grade level between years are excluded from the calculation, since only Grade 6 and Grade 9 records were examined for the 2014–15 school year. Students in Grades 5 and 8 in 2013–14 who were not found in the fall 2014 enrollment records in the 2014–15 school year were excluded from the analysis since their grade enrollment status could not be determined.

[23] Grade retention was created by comparing students' fall 2013 grade of enrollment to their fall 2014 grade of enrollment. Students with matching 2013 and 2014 grade levels were marked as having been retained in grade. If the student advanced grade levels between 2013–14 and 2014–15, the student was classified as having been promoted. Students in Grades 5 and 8 in 2013–14 who were not found in the fall 2014 enrollment records in the 2014–15 school year were removed from the analysis since their enrollment status, or the reason for their absence from the enrollment files, could not be determined.

GIBSON
CONSULTING GROUP

student in 2013–14. Students deemed to not have been enrolled in Texas public schools consist of those for whom an enrollment record could not be found in PEIMS for both the 2012–13 and 2013–14 school years.

Next, students' enrollment status in 2014–15 was disaggregated by whether they were classified as being at risk of being retained in grade in 2014–15 (Table 2.5). Being at risk for retention in 2014–15 was defined as having failed the STAAR-Reading exam in either of the two school years prior to 2013–14 (2011–12 or 2012–13). This disaggregation was conducted because descriptive results presented in Garland, et al (2015) showed that students in Grades 5 and 8 who were at risk of being retained in grade used Istation more frequently relative to students who were not at risk of being retained in grade. Grade 5 students at risk of being retained in grade between 2013–14 and 2014–15 were much more likely to be retained in grade compared to students who were not at risk of being retained in grade (3.2% compared to 0.2%). Grade 8 students at risk of being retained in grade were also much more likely to be retained in grade than Grade 8 students who were not at risk of being retained (1.9% compared to 0.2%).

Table 2.5 Student Promotion/Retention Results between 2012–13 and 2013–14, by Grade Level and At-Risk Status

		Grade 5			Grade 8		
		Promoted	Retained in Grade	Not Enrolled in Texas Public Schools in 2014–15	Promoted	Retained in Grade	Not Enrolled in Texas Public Schools in 2014–15
Did not fail STAAR-Reading (i.e., Not At-Risk)	Count	220,261	354	6,861	230,229	522	6,357
	Percentage of students	96.83	0.16	3.02	97.10	0.22	2.68
Failed STAAR-Reading at least once (i.e., At-Risk)	Count	105,854	3,570	3,145	97,966	1,994	3,455
	Percentage of students	94.03	3.17	2.79	94.73	1.93	3.34

Source: Public Education Information Management System (PEIMS), 2012–13, 2013–14, and 2014–15; State of Texas Assessments of Academic Readiness (STAAR), 2011–12 to 2012–13, Texas Education Agency. Istation Session History table, 2013–14.

Note: Frequency counts and calculations only include students who met the following inclusion rules: enrolled at a campus where at least one student was registered for Istation in 2013–14; was a first-time Grade 5 or Grade 8 student in 2013–14; and had a valid STAAR-Reading test score in 2011–12 and 2012–13. Only regular English and Spanish versions of STAAR-Reading (i.e., not modified or alternate versions) were included. Only STAAR-Reading test data from 2011–12 to 2013–14 were examined. At risk of being retained in grade in 2014–15 was defined as having failed the STAAR-Reading exam in either of the previous two school years (2011–12 or 2012–13). Students deemed to not have been enrolled in Texas public schools consist of those for whom an enrollment record could not be found in PEIMS for both the 2012–13 and 2013–14 school years.

GIBSON
CONSULTING GROUP

Description of Istation Usage and Retention

Figure 2.2 displays the percentage of students who were retained in Grades 5 and 8 between 2013–14 and 2014–15 for students who either used Istation for one or more minutes in 2013–14 or did not use the system at all. In Grade 5, 1.3% of students who used Istation for one or more minutes were retained in grade between 2013–14 and 2014–15, compared to 1% of students who did not use the system at all. In Grade 8, 1% of students who used Istation for one or more minutes were retained in grade compared to 0.7% of students who did not use the system at all.

Figure 2.2. Percentage of First-Time Grade 5 and 8 Students Retained in Grade between 2013–14 and 2014–15, by Istation Usage in 2013–14
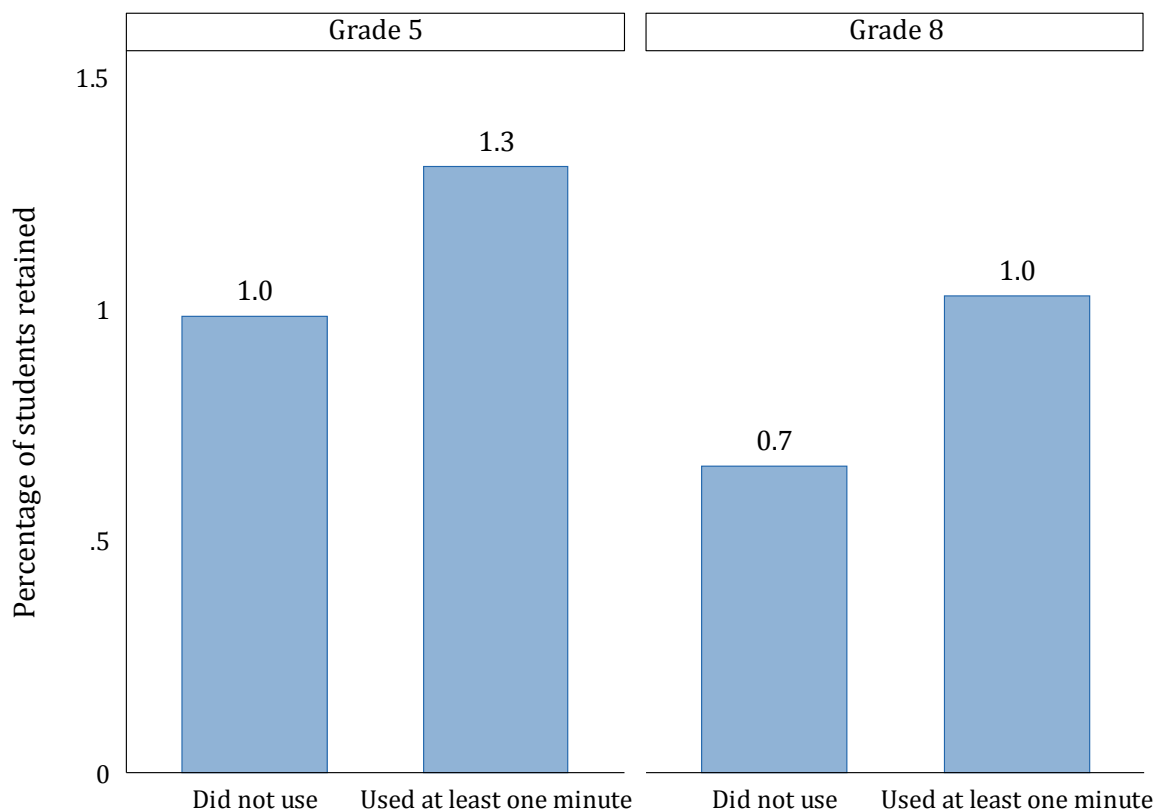
Source: Public Education Information Management System, 2013–14 and 2014–15, Texas Education Agency. Istation Session History Table, 2013–14.

Note: Frequency counts and calculations only include students who met the following inclusion rules: enrolled at a campus where at least one student was registered for Istation in 2013–14; and was a first-time Grade 5 or Grade 8 student in 2013–14.

As Figure 2.3 illustrates, grade retention results for students at-risk of being retained (i.e., they had failed the first administration of the STAAR-Reading exam in either 2011–12 or 2012–13) mirrored those of students who had actually been retained in grade between the 2013–14 and 2014–15 school years.

GIBSON
CONSULTING GROUP

Approximately 3.3% of first-time Grade 5 students who used the Istation system were retained in grade between 2013–14 and 2014–15 compared to 3.1% of comparable students who did not use Istation in 2013–14. Likewise, approximately 2.1% of first-time Grade 8 students who used the Istation system were retained in grade between 2013–14 and 2014–15 compared to 1.9% of comparable students who did not use Istation in 2013–14.

Figure 2.3. Percentage of First-Time Grade 5 and 8 Students who were At Risk of Being Retained in Grade between 2013–14 and 2014–15 who were Actually Retained, by Istation Usage in 2013–14

Source: Public Education Information Management System, 2012–13, 2013–14, and 2014–15; State of Texas Assessments of Academic Readiness (STAAR), 2011–12 to 2012–13, Texas Education Agency. Istation Session History Table, 2013–14.

Note: Frequency counts and calculations only include students who met the following inclusion rules: enrolled at a campus where at least one student was registered for Istation in 2013–14; was a first-time Grade 5 or Grade 8 student in 2013–14; failed STAAR-Reading in either 2011–12 or 2012–13; and had a valid STAAR-Reading test score in 2011–12 and 2012–13. Only regular English and Spanish versions of STAAR-Reading (i.e., not modified or alternate versions) were included in the decision rule. Only STAAR-Reading test data from 2011–12 to 2013–14 were examined. Being at risk for retention in grade in 2014–15 was defined as having failed the STAAR-Reading exam in either of the previous two school years (2011–12 or 2012–13).

GIBSON
CONSULTING GROUP

Istation Usage and Outcome Results Related to Grade Retention

The results presented above are only descriptive, and do not adjust for differences in other characteristics between students who used the system and students who did not use Istation, nor does it account for differences across campus types or environments. To account for these differences, and to assess the relationship between Istation usage and the likelihood of being retained in Grade 5 or 8 between 2013–14 and 2014–15, the evaluation team used a logistic regression with the outcome measure capturing whether the student was retained in grade. Three measures of Istation usage/intensity (i.e. dosage) in 2013-14 were examined:

1) Whether a student used Istation at least one minute (compared to students below this level);

2) Whether a student used Istation at or above the recommended usage level for the respective grade (compared to students who used Istation at a level usage below the threshold); and

3) Whether a student used Istation 300 minutes or more (compared to students who did not use Istation).

The first measure above examines whether any usage is related to the likelihood of being retained in Grade 5 or 8 between 2013-14 and 2014-15, and the second and third measures are related to intensity of usage. In the model, the evaluation team adjusted for student attributes, including prior STAAR-Reading test scores, and whether they had been retained in grade in either of those two years. Analyses were conducted separately for Grades 5 and 8 to account for differences in the relationship between usage (i.e., one or more minutes of usage) and usage intensity (usage at or above each intensity threshold) and the likelihood of being retained in grade in 2014–15 across grade levels. Furthermore, within each grade level, the analytic models were run separately for two groups of students:

1) First-time Grade 5 and Grade 8 students in 2013–14; and

2) First-time Grade 5 and Grade 8 students who were at risk of being retained in grade between 2013–14 and 2014–15.

The results of these analyses are presented in Table 2.6 and 2.7.[24]

As Table 2.6 shows, no statistically significant differences in the probability of being retained in Grade 5 or 8 between 2013–14 and 2014–15 were found for students who used the Istation system when compared to nonusers. Similarly, no statistically significant differences in the probability of being retained in Grade 5 or 8 between 2013–14 and 2014–15 were found for students who used the Istation system at the recommended levels (of 250 minutes in Grade 5 and 200 minutes in grade 8) when compared to

---

[24] Additional technical detail that describes the statistical model is provided in Appendix A. In addition, supplementary statistical models assessing the sensitivity of the estimates presented in the main section of the report are provided in Appendix A. These models apply propensity score reweighting to account for differences in the probability of treatment assignment between participating and non-participating students based on available pre-treatment covariates.

GIBSON
CONSULTING GROUP

students who used the system at levels below the recommended thresholds. Similarly, no statistically significant differences in the likelihood of being retained in grade between students who used the system 300 minutes or more compared to students who did not use the system. These findings were consistent for both student groups described above. These findings suggest that 2013–14 Istation usage, at any level of intensity, is not related to whether or not Grade 5 and 8 students (first-time, and at-risk students) were retained in grade between 2013–14 and 2014–15.

Table 2.6. Estimated Relationship between Istation Use/Intensity of Usage and Grade Retention between 2013–14 and 2014–15, First-Time Grade 5 and 8 Students in 2013–14

Grade Level	Usage Measure	B	SE	N
Grade 5	Used at least one minute	0.112	0.066	322,342
	Met Istation usage threshold (compared to those who did not meet threshold)	0.068	0.061	322,342
	Used system at least 300 minutes (compared to no usage)	.114	.076	322,342
Grade 8	Used at least one minute	0.037	0.097	321,204
	Met Istation usage threshold (compared to those who did not meet threshold)	-0.147	0.121	321,204
	Used system at least 300 minutes (compared to no usage)	-.089	.157	321,204

Source: Istation Session History table, 2013–14, Texas Education Agency. State of Texas Public Education Information Management System, 2012–13, 2013–14, and 2014–15; State of Texas Assessments of Academic Readiness (STAAR), 2012–13, Texas Education Agency.
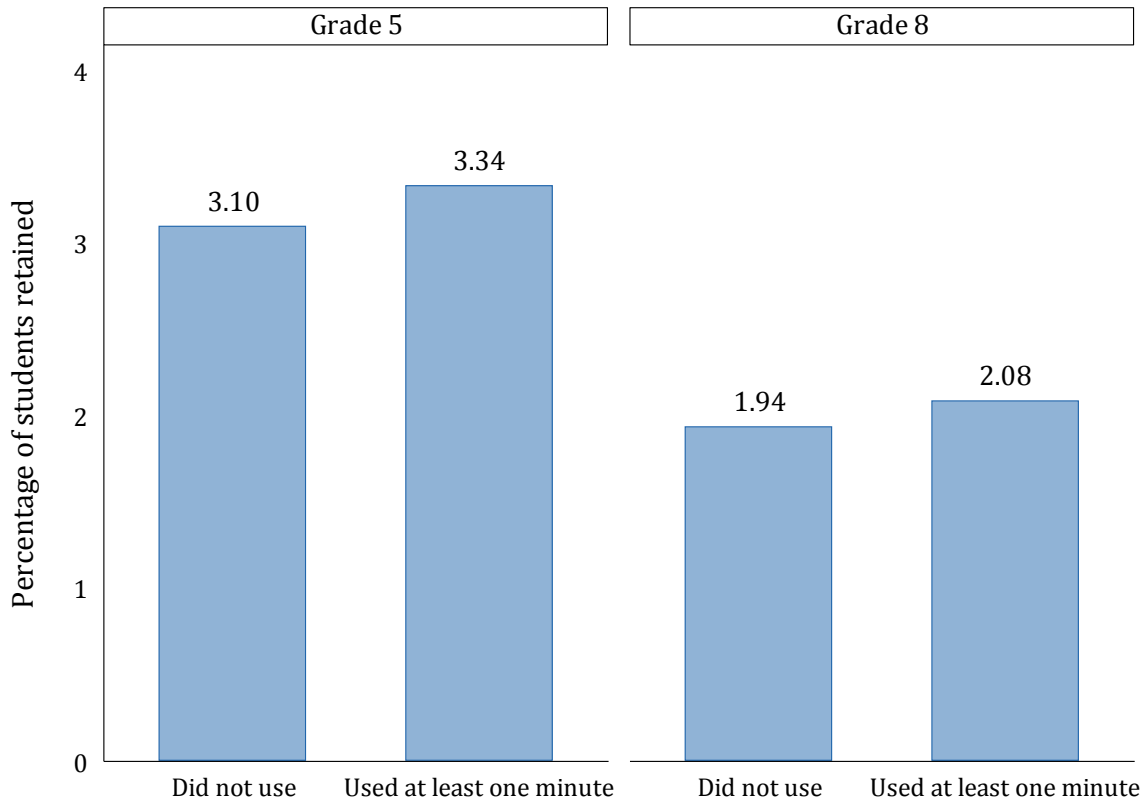
Note: Calculations only include students who met the following inclusion rules: enrolled at a campus where at least one student was registered for Istation in 2013–14; was a first-time Grade 5 or Grade 8 student in 2013–14; and had a valid STAAR-Reading test score in the prior year (2012–13). Only regular English and Spanish versions of STAAR-Reading (i.e., not modified or alternate versions) were included. 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. The notation, *B*, in this table refers to the logit coefficient from the statistical model and the notation, SE, refers to the standard error.

GIBSON
CONSULTING GROUP

Table 2.7. Estimated Relationship between Istation Use/Intensity of Usage and Grade Retention between 2013–14 and 2014–15, Grade 5 and 8 Students in 2013–14 who were at Risk of Being Retained in Grade

Grade Level	Usage Measure	B	SE	N
Grade 5	Used at least one minute	0.125	0.068	104,488
	Met Istation usage threshold (compared to those who did not meet threshold)	0.063	0.061	104,488
	Used system at least 300 minutes (compared to no usage)	.121	.079	104,488
Grade 8	Used at least one minute	0.038	0.109	96,336
	Met Istation usage threshold (compared to those who did not meet threshold)	-0.150	0.126	96,336
	Used system at least 300 minutes (compared to no usage)	-.099	.167	96,336

Source: Istation Session History table, 2013–14. Public Education Information Management System, 2012–13, 2013–14, and 2014–15; State of Texas Assessments of Academic Readiness (STAAR), 2011–12 to 2012–13, Texas Education Agency.

Note: Results only include students who met the following inclusion rules: enrolled at a campus where at least one student was registered for Istation in 2013–14; was a first-time Grade 5 or Grade 8 student in 2013–14; and had a valid STAAR-Reading test score in 2011–12 and 2012–13. Only regular English and Spanish versions of STAAR-Reading (i.e., not modified or alternate versions) were included. 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. The notation, B, in this table refers to the logit coefficient from the statistical model and the notation, SE, refers to the standard error.

Summary of Istation Findings

The two research questions in this chapter explored the relationships between Istation system usage and intensity of usage (i.e., dosage) in 2013–14 by at-risk students who were retained in grade between 2012–13 and 2013–14 and two important outcomes: 1) STAAR-Reading gains between 2012–13 and 2013–14; and 2) the probability of being retained in grade in 2014–15. Descriptive and multivariate findings presented in this report section are presented below:

Istation System Usage

- On average, Grade 5 students who were retained in grade recorded almost an hour more Istation usage over the course of the 2013–14 school year (239 minutes compared to 182 minutes) than students who were not retained.

- System usage levels for students in Grade 8 were relatively low compared to elementary school students; however, Grade 8 students who were retained in grade between 2012–13 and 2013–14 recorded an average of 49 minutes of Istation usage in 2013–14 compared to 38 minutes for students who were in Grade 8 for the first time in 2013–14.

- It was rare for students to be retained in grade between 2012–13 and 2013–14 and to use Istation at the recommended level of 250 minutes for Grade 5 students and 200 minutes for Grade 8 students during the 2013–14 school year.

GIBSON
CONSULTING GROUP

Reading Outcomes – Relationship between Istation and STAAR-Reading Gains

- After adjusting for other student, school, and district-level characteristics, no statistical relationship was found between the use of Istation by Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 and STAAR-Reading gains in 2013–14 (compared to not using the system at all).

- No statistical relationship was found between the intensive usage (e.g., 300 or more minutes over the course of the 2013–14 school year) of Istation by Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 and STAAR-Reading gains in 2013–14. That is, after adjusting for other student, school, and district-level characteristics, no statistically significant differences in 2013–14 STAAR-Reading gains were observed among Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 who used the system for 300 or more minutes, and those who did not use the system at all.

- Grade 8 students retained in grade between 2012–13 and 2013–14 who used Istation for 200 or more minutes during 2013–14 had statistically significant, moderately larger reading gains (.29 standard deviations) than their counterparts who used the system at levels below the recommended usage threshold.

- Contrary to the finding for Grade 8 students retained in grade between 2012–13 and 2013–14, among comparably retained Grade 5 students, Istation usage at or above the annual level of 250 minutes was found to be statistically significant, but negatively related to STAAR-Reading gains (when compared to students who did not use the system at the recommended level for students in this grade).

Reading Outcomes – Relationship between Istation Usage and the Likelihood of Being Retained in Grade

- After adjusting for other student, school, and district-level characteristics that may be associated with the likelihood of Istation usage and the likelihood of being retained in grade in 2014–15, no statistically significant differences in the probability of being retained in grade were found between participating and non-participating students. This finding was consistent across both grade levels, all measures of Istation usage (usage and usage intensity), and across Grade 5 and Grade 8 first-time students overall, as well as those first-time students at risk of being retained in grade.

The patterns of associations found in this section parallel those reported in Garland, et al (2015) indicating small to moderate, statistically significant associations between Istation dosage and STAAR-Reading gains among groups of Grade 8 students who are at risk of being retained in grade. The findings in this addendum report, however, should be interpreted with caution, given the small size of the subpopulation of students eligible to be included in the statistical model.

GIBSON
CONSULTING GROUP

Section 3 – Mathematics Outcomes

While Garland, et al (2015) provided a broad assessment of TTM implementation and program impacts, the results presented in this section are restricted to two important grade levels (Grades 5 and 8) and high-need students. Grade 5 and Grade 8 were selected for these analyses because students are subject to high-stakes assessments with grade placement implications in these grades. In addition, the research questions posed in this addendum are geared toward students in at-risk situations (i.e., students retained in grade or at-risk of being retained in grade).

This section of the report addresses the following two research questions related to the TTM program:

1) Among students who were retained in Grade 5 or 8 between 2012–13 and 2013–14, to what extent does 2013–14 academic performance differ between TTM participants and non-participants, and what is the relationship between intensity of TTM usage (i.e., dosage) and academic performance for this student population?

2) What is the relationship between TTM program usage and the probability of being retained in Grades 5 and 8 between 2013–14 and 2014–15?

Appendix B provides technical information related to the TTM analyses contained in this addendum report.

Guiding Question 1b: Among students who were retained in Grade 5 or 8 between 2012–13 and 2013–14, to what extent does 2013–14 academic performance differ between TTM participants and non-participants, and what is the relationship between intensity of TTM usage (i.e., dosage) and academic performance for this student population?

The descriptive analysis presented in Garland, et al (2015) found that students in Grades 5 and 8 who were at risk of being retained in grade used TTM at higher levels than students were not at risk of being retained in grade. Being at risk for retention in grade was defined as having failed the STAAR-Mathematics exam in either 2011–12 or 2012–13. This section of the addendum report expands on the prior analyses by exploring the relationships between **TTM usage** (i.e., whether a student used the TTM system at all) and **TTM usage intensity** (i.e., how frequently a student used the TTM system) and changes in student performance on STAAR-Mathematics for students who were actually retained in Grades 5 and 8 between 2012–13 and 2013–14.

After examining TTM usage patterns for Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 and students not retained, the evaluation team explored the relationship between TTM usage and changes in student performance on STAAR-Mathematics for students in Grades 5 and 8 who were retained in grade between 2012–13 and 2013–14. Analyses were confined to students in Grades 5 and 8

GIBSON
CONSULTING GROUP

in 2013–14 who were retained in grade between 2012–13 and 2013–14 and to students who were enrolled in schools where TTM was used (in 2013–14).

The primary research question explored the relationship between whether students used TTM at all and STAAR-Mathematics gain scores for students retained in grade; however, it is equally important to understand whether higher levels of system usage (i.e., usage intensity) are related to improved academic performance. Therefore, analyses were conducted for the following categories of TTM system usage:

1) Students who attempted at least one TTM lesson (i.e., used the system at all) compared to students who did not use the TTM system at all;

2) The number of TTM lessons a student attempted (i.e., a continuous variable of attempted lessons);

3) Students who attempted less than the minimum number of TTM lessons recommended by the vendor (i.e., 1 to 4 lessons attempted) compared to students who did not attempt any lessons; and

4) Students who attempted 5 to 9 lessons, 10 to 14 lessons, 15 to 19 lessons, or 20 or more lessons (i.e., three measures of increasingly high intensity usage) compared to students who did not attempt any lessons.

Grade-level retention was calculated by comparing students' grade level in fall 2012 to their grade level in fall 2013.[25] The calculation only includes students who were enrolled in Grade 5 or 8 in 2012–13. If students' grade level matched between these two periods, they were classified as having been retained in grade between the 2012–13 and 2013–14 school years.

TTM Usage and Grade Retention

Table 3.1 displays the count of students (at campuses where at least one student was registered for TTM in 2013-14) in Grades 5 and 8 in 2013–14 who were identified as having been retained in grade between 2012–13 and 2013–14.[26] Overall, the percentage of Grade 5 and Grade 8 students who were retained in

---

[25] We followed TEA's operationalization of grade-level retention, with one difference. Students were classified as having been retained if their fall 2012 and fall 2013 grade level matched. The TEA definition utilizes attendance records from the baseline year (here, 2012–13) to include students who were not enrolled by the PEIMS snapshot date, where baseline grade level is determined by the last grade in which a student was enrolled (Texas Education Agency, 2015). For more information, see TEA's resources on dropout and retention here: http://tea.texas.gov/acctres/retention_index.html.

[26] The retention rate calculation in this report is slightly different than the statewide retention rate, since the denominator in the calculation presented in Table 3.1 reflects all students in Grade 5 or 8 in 2013–14 who met the sample inclusion rules for the evaluation. Table 3.1 provides a description of the composition of the analytic sample that will be used in the outcomes section of this research question.

GIBSON
CONSULTING GROUP

grade in 2013–14 was low: approximately 1.5% of Grade 5 students and 1.0% of Grade 8 students were identified as being retained in grade between 2012-13 and 2013-14.

Table 3.1. Frequency Count and Percentage of Grade 5 and 8 Students at TTM Campuses Retained in Grade between 2012–13 and 2013–14

Student Grade Level	Total Student Count	Count of Students Retained	Percentage of Students
Grade 5	360,570	5,224	1.45%
Grade 8	350,956	3,591	1.02%

Source: Public Education Information Management System, 2012–13 and 2013–14, Texas Education Agency.

Note: Frequency counts and calculations include all Grade 5 and 8 students in 2013–14 who were enrolled at a campus where at least one student was registered for TTM in 2013–14 and attempted at least one TTM lesson. Students denoted as retained in grade were students in Grades 5 and 8 in 2012–13 whose grade level in the fall of 2013 matched their grade level in the fall of 2012.

Table 3.2 shows the percentage and number of Grade 5 and 8 students who were first-time Grade 5 or 8 students in 2013–14 or were retained in grade between 2012–13 and 2013–14. Furthermore, these summary statistics are disaggregated by two measures of TTM usage: 1) whether the student attempted at least one lesson; and 2) whether the student attempted 5 or more lessons. The 5-lesson cut point reflects 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.

As Table 3.2 shows, approximately 43% of Grade 5 students who were retained in grade between 2012–13 and 2013–14 attempted at least one TTM lesson versus 46% of students who were not retained. The percentage of retained students who used the TTM at the minimum recommended usage level (i.e., 5 attempted lessons) (36%) was not substantially different than students who were not retained (37%).

Another important finding of Table 3.2 is that a smaller percentage of Grade 8 students attempted TTM lessons than Grade 5 students. This is an important descriptive point to keep in mind when reviewing the results from the forthcoming multivariate analyses.

GIBSON
CONSULTING GROUP

Table 3.2. Frequency Count and Percentage of Grade 5 and 8 Students who Used TTM, by Grade-Level Retention Status, and Usage Intensity, 2013–14

Student Grade Level	Enrollment Status in 2013–14	At least 1 Attempted TTM Lessons	At least 5 Attempted TTM Lessons	Percentage at Least 1 Attempted TTM Lesson	Percentage at Least 5 Attempted TTM Lessons
Grade 5	First time	172,594	140,688	45.7	37.3
Grade 5	Retained in grade	2,221	1,887	42.5	36.1
Grade 8	First time	92,394	55,862	24.5	14.9
Grade 8	Retained in grade	666	367	18.5	10.2

Source: Public Education Information Management System, 2012–13 and 2013–14, Texas Education Agency. TTM Session History table, 2013–14.

Note: Frequency counts and calculations include all Grade 5 and 8 students in 2012–13 who were enrolled at a campus where at least one student who attempted a TTM lesson in 2013–14. Students classified as retained in grade were students in Grades 5 and 8 in 2012–13 whose grade level in the fall of 2013 matched their grade level in the fall of 2012. First-time Grade 5 or 8 students reflect all students who were enrolled in Grade 5 or 8 for the first time in 2013–14, including students who were promoted, and students who were not enrolled in a Texas public school in the prior school year.

TTM Usage and Student Outcomes Related to STAAR-Mathematics Gains

First, the evaluation team explored descriptive system usage data to provide context for the student outcome findings which follow. Grade 5 and 8 students who were retained in grade between 2012–13 and 2013–14 used TTM somewhat less frequently than students who were not retained in grade (Figure 3.1). ***The TTM vendor asserts that a student has reached a "gateway" point where they are more likely to continue using the system (when they attempt 5 or more lessons) and are at a "fidelity or impact" point where students may start experiencing gains in mathematics scores (when they attempt 10 or more lessons).*** In Grade 5, students who were retained in grade attempted an average of 9.26 lessons in 2013–14 compared to 9.77 lessons for students who were not retained in grade. In Grade 8, students who were retained in grade attempted an average of 2.70 lessons in 2013–14 compared to 3.79 lessons for students who were not retained in grade (Figure 3.1)**.**

These findings contrast with the descriptive results contained in Garland, et al (2015), which explored differences in usage intensity between Grade 5 and Grade 8 students who had failed STAAR-Mathematics in a prior school year compared to students who had not. For example, students in Grade 8 who were classified as having been at risk of being retained in grade between 2013–14 and 2014–15 attempted more lessons that peers who were not identified as being at risk of being retained in grade. This suggests that school staff were targeting students at risk of being retained in grade for more intensive usage of TTM relative to their peers who were not at risk of being retained in grade; this was not observed among students actually retained in grade between 2012–13 and 2013–14.

GIBSON
CONSULTING GROUP

Figure 3.1. Average Number of TTM Lessons Attempted by Grade 5 and 8 Students in 2013–14, by 2013–14 Grade-Level Retention Status
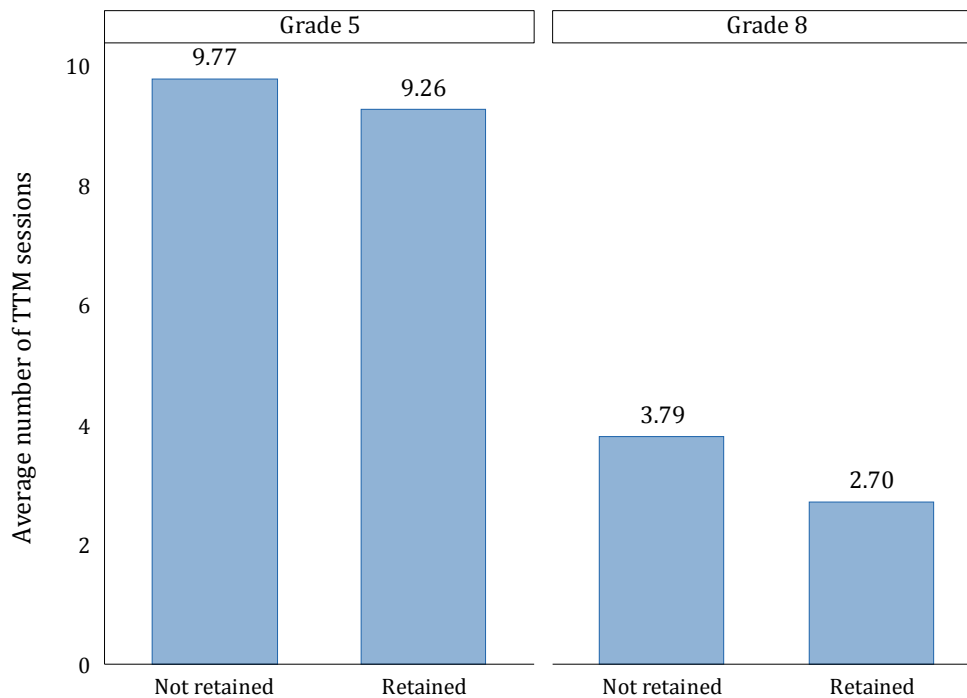
Source: Public Education Information Management System, 2012–13 and 2013–14, Texas Education Agency. TTM Session History table, 2013–14.

Note: Calculations only include students who were enrolled at a campus where at least one student attempted a TTM lesson in 2013–14. Students classified as retained in grade were students in Grades 5 and 8 in 2012–13 whose grade level in the fall of 2013 matched their grade level in the fall of 2012.

Student decile-standardized scores were used to better account for differences in prior STAAR performance when examining changes in STAAR-Mathematics scores between 2012–13 and 2013–14, and to account for different average changes in student performance across the distribution of students' prior achievement score.[27] As noted earlier, the different measures of system usage analyzed are:

---

[27] The evaluation team calculated decile-standardized STAAR-Mathematics test-score gains between the first test administration in 2012–13 and the first test administration in 2013–14. 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-Mathematics 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-Mathematics 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-Mathematics score decile*. The quantity of interest for the evaluation is the *difference* in the gain score between TTM users and non-users. This standardization facilitates comparisons across grade levels for

1) Student attempted at least one TTM lesson (compared to students who did not use TTM at all);

2) Student used the system at below recommended level, attempting at least 1 to 4 TTM lessons (compared to students who did not use TTM at all);

3) Student used the system at recommended "gateway" level, attempting 5 to 9 TTM lessons (compared to students who did not use TTM at all);

4) Student used the system at recommended "impact" level, attempting 10 to 14 TTM lessons (compared to students who did not use TTM at all);

5) Student used the system intensively at levels above recommended "impact" level, attempting 15 to 19, or 20 or more TTM lessons (compared to students who did not use TTM at all).[28]

In addition to the categorical measures presented above, the evaluation team also included the number of TTM lessons attempted as an independent variable in the statistical model to predict changes in STAAR-Mathematics gain scores.

The results are presented in Table 3.3.

After adjusting for other student, school, and district-level characteristics, there were no statistically significant differences in 2013–14 STAAR-Mathematics score gains (for Grade 5 and Grade 8 students retained in grade between 2012–13 and 2013–14) when the results of students who attempted at least one TTM lesson were compared to those of students who did not attempt any TTM lessons (Table 3.3).[29]

---

different grade-level assessments because the scale is standardized to have equal means and variance for each assessment, in each year and grade level.

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

[29] Throughout this section, the terms "statistically significant" and "significant" are used interchangeably. They refer to instances in which, after controlling for other factors, differences in outcomes (e.g., STAAR-Mathematics gain scores) are found between student groups (e.g., those who attempted one or more TTM lessons and those who did not use TTM at all). 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.

GIBSON
CONSULTING GROUP

Table 3.3. Estimates of the Differences in the Relationship between 2013–14 TTM Usage/Intensity of Usage and STAAR-Mathematics Gains, for Students in Grade 5 and 8 who were Retained in Grade between 2012–13 and 2013–14

Grade	Usage Measure	B	SE	N
Grade 5	Attempted at least one lesson (compared to did not use)	0.080	0.062	4,503
	Number of Attempts	*0.943****	0.131	
	1–4 Attempts (compared to zero attempted lessons)	*0.174**	0.083	
	5-9 Attempts (compared to zero attempted lessons)	0.101	0.086	
	10-14 Attempts (compared to zero attempted lessons)	0.064	0.083	
	15-19 Attempts (compared to zero attempted lessons)	0.086	0.087	
	20+ Attempts (compared to zero attempted lessons)	-0.007	0.073	
Grade 8	Attempted at least one lesson (compared to did not use)	0.056	0.095	1,977
	Number of Attempts	*0.905****	0.242	
	1–4 Attempts (compared to zero attempted lessons)	0.141	0.117	
	5-9 Attempts (compared to zero attempted lessons)	-0.132	0.128	
	10-14 Attempts (compared to zero attempted lessons)	0.178	0.210	
	15-19 Attempts (compared to zero attempted lessons)	0.017	0.341	
	20+ Attempts (compared to zero attempted lessons)	-0.021	0.163	

Source: State of Texas Assessments of Academic Readiness (STAAR) and Public Education Information Management System, 2011–12 and 2013–14, Texas Education Agency. TTM Session History table, 2013–14.

Note: Only regular English and Spanish versions (i.e., not modified or alternate versions) are included in the analyses. Only students with at a valid STAAR-Mathematics test score on the first administration in 2012–13 and 2013–14 are included in the multivariate model. Estimates derived from a linear regression with campus-level fixed effects and pre-treatment student-level covariates. $*$ $p<0.05$; $**$ $p<0.01$; $***$ $p<0.001$

However, the absolute number of TTM lessons attempted was found to be positive, and significantly associated with STAAR-Mathematics gains. This finding suggests that students who used TTM more often experienced larger STAAR-Mathematics gains than students using the system less frequently. The difference was statistically significant and moderate (.943 for Grade 5 and .905 for Grade 8 for each additional lesson attempted).

As Table 3.3 shows, among students in Grade 5, who were retained in grade between 2012–13 and 2013–14, a positive and statistically significant relationship was observed between attempting 1 to 4 TTM lessons (which reflects a level of usage below the 5 attempted lessons recommended by the TTM vendor) and 2013–14 STAAR-Mathematics gains (.174 standard deviations higher). This statistical relationship was not observed for Grade 8 students.

As Table 3.3 reflects, increasingly intensive TTM usage over the course of the 2013–14 school year by Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 was not related to larger STAAR-Mathematics gains in 2013–14 (when compared to students who did not used the system at all). No

GIBSON
CONSULTING GROUP

statistically significant differences in 2013–14 STAAR-Mathematics gains were observed among Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 who used TTM at increasingly intensive levels (5 to 9 attempted lessons, 10 to 14 attempted lessons, 15 to 19 attempted lessons, or 20 or more attempted lessons), relative to students who did not use the TTM system at all.

It is important to consider that the population of students who met the usage criteria, and who had sufficient data to be included in the statistical model, was small and observably different from the overall student population.[30] For instance, in Grade 8, fewer than 400 students who were retained in grade between 2012–13 and 2013–14 attempted 5 or more TTM lessons (Table 3.2). Consequently, the amount of uncertainty in these estimates (which is reflected in the size of the standard error of the estimates that are presented in Table 3.3) is large, and the results should be interpreted with caution.

Guiding Question 2b: What is the relationship between TTM program usage and the probability of being retained in Grades 5 and 8 between 2013–14 and 2014–15?

Prior to describing the relationship between TTM usage during the 2013–14 school year and whether students in Grade 5 or 8 were retained in grade between 2013–14 and 2014–15, the research team received fall 2014 PEIMS student enrollment records to calculate actual retention and promotion between 2013–14 and 2014–15.[31,32]

Grade-Level Enrollment Status

Table 3.4 provides summary descriptive information about students who were first-time enrolled students in Grade 5 and Grade 8 in 2013–14. Only a very small percentage of students in Grades 5 and 8 in 2013–14 were retained in grade between 2013–14 and 2014–15, with a slightly higher percentage of Grade 5 students retained in grade than Grade 8 students (1.2% compared to 1.02%). In each grade, less than 4%

---

[30] For instance, students who were retained in grade had a higher rate of test abstention, or ineligibility based on the STAAR-Mathematics inclusion rules adopted in this evaluation, than students who were not retained. That is, students who were retained in grade took the first administration of the STAAR exam at lower rates than students not retained in grade.

[31] Only sequential grade promotion is reflected in the promotion indicator. Students promoted more than one grade level between years are excluded from the calculation, since only Grade 6 and Grade 9 records were examined for the 2014–15 school year. Students in Grades 5 and 8 in 2013–14 who were not found in the fall 2014 enrollment records in the 2014–15 school year were excluded from the analysis since their grade enrollment status, could not be determined.

[32] Grade retention was created by comparing students' fall 2013 grade of enrollment to their fall 2014 grade of enrollment. Students with matching 2013 and 2014 grade levels were marked as having been retained in grade. If the student advanced grade levels between 2013–14 and 2014–15, the student was classified as having been promoted. Students in Grades 5 and 8 in 2013–14 who were not found in the fall 2014 enrollment records in the 2014–15 school year were removed from the analysis since their enrollment status could not be determined.

GIBSON
CONSULTING GROUP

of students who were enrolled in Grade 5 or 8 for the first time in 2013–14 were not enrolled in Texas public schools in the fall of 2014.

Table 3.4. Student Promotion/Retention Results between 2012–13 and 2013–14, by Grade Level

	Grade 5			Grade 8		
	Promoted	Retained in Grade	Not Enrolled in Texas Public Schools in 2014–15	Promoted	Retained in Grade	Not Enrolled in Texas Public Schools in 2014–15
Count	338,405	4,275	12,856	332,187	2,882	12,612
Percentage of students	95.18	1.20	3.62	95.54	1.02	3.63

Source: Public Education Information Management System (PEIMS) 2012–13, 2013–14, and 2014–15, Texas Education Agency.

Note: Frequency counts and calculations only include students who met the following inclusion rules: enrolled at a campus where at least one student was registered for TTM in 2013–14; and was a first-time Grade 5 or Grade 8 students in 2013–14. Students deemed to not have been enrolled in Texas public schools consist of those for whom an enrollment record could not be found in PEIMS for both the 2012–13 and 2013–14 school years.

Next, students' enrollment status in 2014–15 was disaggregated by whether they were classified as being at risk of being retained in grade in 2014–15 (Table 3.5). Being at risk for retention in grade in 2014–15 was defined as having failed the STAAR-Mathematics exam in either of the two school years prior to 2013–14 (2011–12 or 2012–13). Descriptive results provided in Garland, et al (2015) showed that students in Grades 5 and 8 who were at risk of being retained in grade had a higher average number of TTM attempted lessons relative to students who were not at risk of being retained in grade. Grade 5 students at risk of being retained in grade between 2013–14 and 2014–15 were much more likely to be retained in grade compared to students who were not at risk of being retained in grade (2.7% compared to .13%). Grade 8 students at risk of being retained in grade were much more likely to be retained in grade than Grade 8 students who were not at risk of being retained in grade (1.9% compared to .22%).

GIBSON
CONSULTING GROUP

Table 3.5. Student Promotion/Retention Results between 2012–13 and 2013–14, by Grade Level and At-Risk Status

At-Risk Status		Grade 5			Grade 8		
		Promoted	Retained in Grade	Not Enrolled in Texas Public Schools in 2014–15	Promoted	Retained in Grade	Not Enrolled in Texas Public Schools in 2014–15
Did not fail STAAR-Mathematics (i.e., Not At-Risk)	Count	187,168	249	5,574	177,942	395	4,724
	Percentage of students	96.98	0.13	2.88	97.10	0.22	2.68
Failed STAAR-Mathematics at least once (i.e., At-Risk)	Count	121,093	3518	3796	94,572	1,938	3,457
	Percentage of students	94.30	2.73	2.95	94.60	1.94	3.45

Source: Public Education Information Management System (PEIMS), 2012–13, 2013–14, and 2014–15, Texas Education Agency.

Note: Frequency counts and calculations only include students who met the following inclusion rules: enrolled at a campus where at least one student was registered for TTM in 2013–14; was a first-time Grade 5 or Grade 8 student in 2013–14; and had a valid State of Texas Assessments of Academic Readiness (STAAR)-Mathematics test score in 2011–12 and 2012–13. Only regular English and Spanish versions of STAAR-Mathematics (i.e., not modified or alternate versions) were included. Only STAAR-Mathematics test data from 2011–12 to 2013–14 were examined. At risk of being retained in grade in 2014–15 was defined as having failed the STAAR-Mathematics exam in either of the previous two school years (2011–12 or 2012–13). Students deemed to not have been enrolled in Texas public schools consist of those for whom an enrollment record could not be found in PEIMS for both the 2012–13 and 2013–14 school years.

Description of TTM Usage and Retention

Figure 3.2 shows the percentage of students who were retained in Grades 5 and 8 between 2013–14 and 2014–15 for students who either used attempted one or more TTM lessons in 2013–14 or did not use the system at all. In both Grades 5 and Grade 8, students who attempted one or more TTM lessons were somewhat less likely to be retained in grade compared to students who did not use the system. In Grade 5, 1.1% of students who attempted one or more TTM lessons were retained in grade between 2013–14 and 2014–15, compared to 1.3% of students who did not use the system at all. In Grade 8, 0.8% of students who attempted one or more TTM lessons were retained in grade between 2013–14 and 2014–15, compared to 0.9% of students who did not use the system at all.

GIBSON
CONSULTING GROUP

Figure 3.2. Percentage of First-Time Grade 5 and 8 Students Retained in Grade between 2013–14 and 2014–15, by TTM Usage in 2013–14
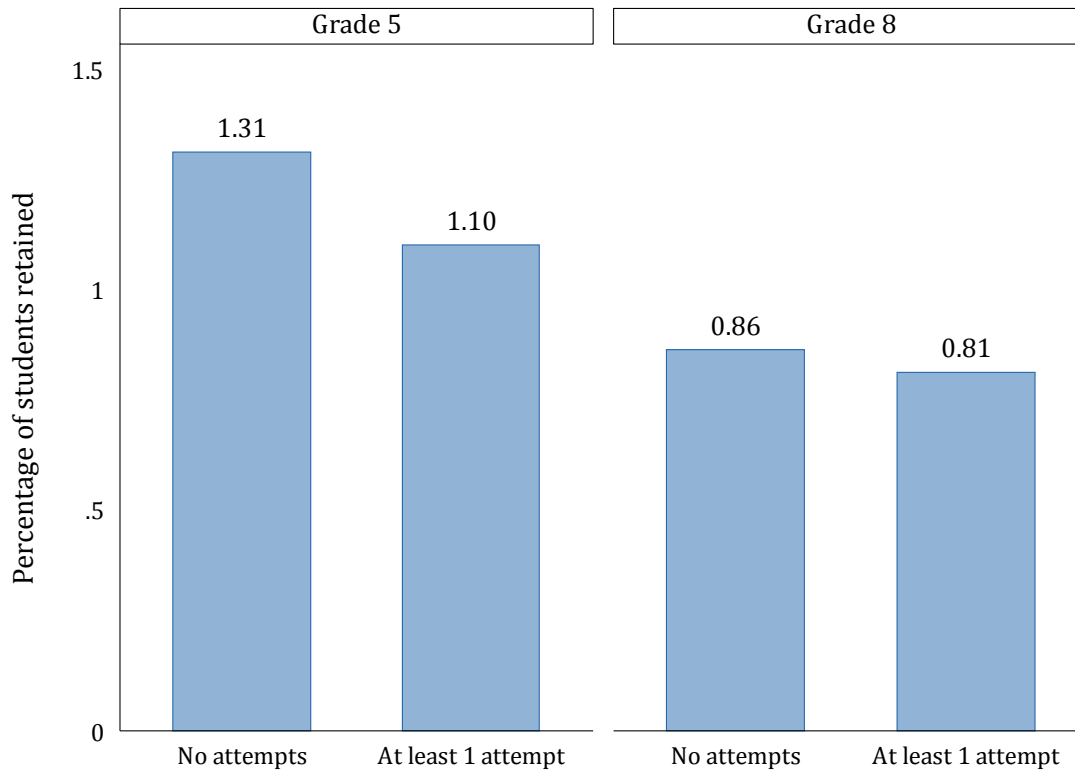
Source: State of Texas Public Education Information Management System, 2012–13, 2013–14, and 2014–15, Texas Education Agency. TTM Session History Table, 2013–14

Note: Frequency counts and calculations only include students who met the following inclusion rules: enrolled at a campus where at least one student attempted TTM lessons in 2013–14; was a first-time Grade 5 or Grade 8 student in 2013–14.

As Figure 3.3 illustrates, grade retention results for students at-risk of being retained (i.e., they had failed the first administration of the STAAR-Mathematics exam in either 2011–12 or 2012–13) mirrored those of students who had actually been retained in grade between the 2013–14 and 2014–15. Approximately 2.5% of first-time Grade 5 students who attempted at least one TTM lesson in 2013–14 were retained in grade between 2013–14 and 2014–15, compared to 3.1% of comparable students who did not use TTM in 2013–14. Similarly, approximately 1.7% of first-time Grade 8 students who attempted at least one TTM lesson on 2013–14 were retained in grade between 2013–14 and 2014–15 versus 2.2% of comparable students who did not use TTM in 2013–14.

GIBSON
CONSULTING GROUP

Figure 3.3. Percentage of First-Time Grade 5 and 8 Students who were At Risk of Being Retained in Grade between 2013–14 and 2014–15 who were Actually Retained, by TTM Usage in 2013–14
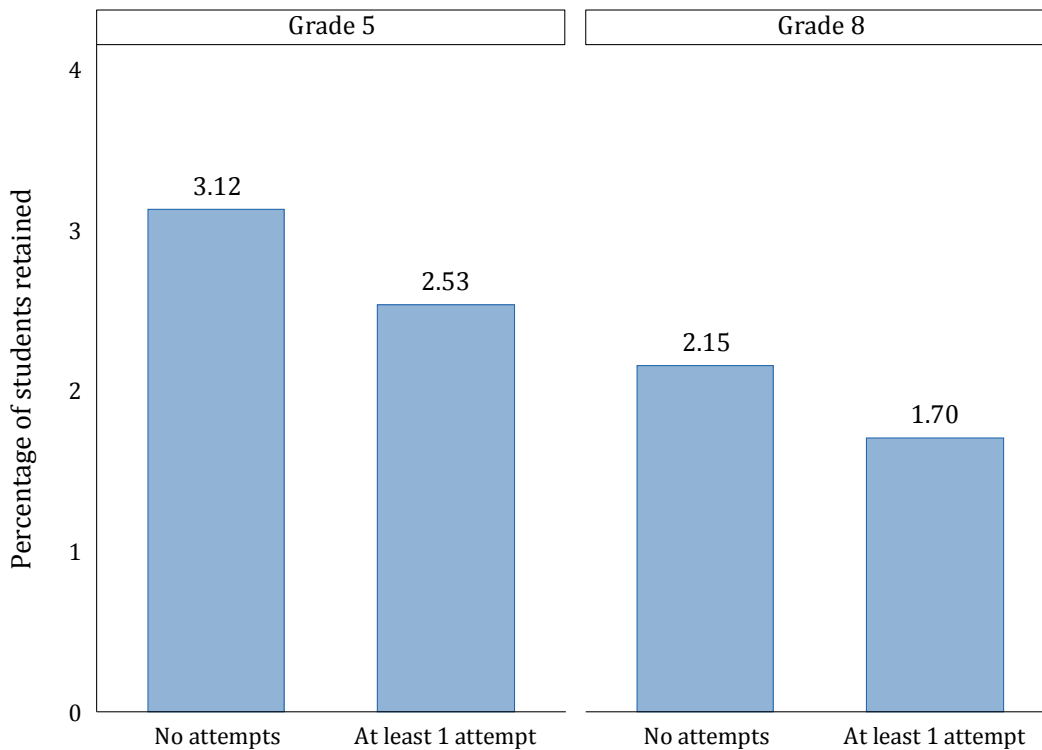
Source: Public Education Information Management System, 2012–13, 2013–14, and 2014–15; State of Texas Assessments of Academic Readiness (STAAR), 2011–12 to 2012–13, Texas Education Agency. TTM Session History Table, 2013–14.

Note: Frequency counts and calculations only include students who met the following inclusion rules: enrolled at a campus where at least one student attempted TTM lessons in 2013–14; was a first-time Grade 5 or Grade 8 student in 2013–14; failed STAAR-Mathematics in either 2011–12 or 2012–13; and had a valid STAAR-Mathematics test score in in 2011–12 and 2012–13. Only regular English and Spanish versions of STAAR-Mathematics (i.e., not modified or alternate versions) were included in the decision rule. Only STAAR-Mathematics test data from 2011–12 to 2013–14 were examined. Being at risk for retention in grade in 2014–15 was defined as having failed the STAAR-Mathematics exam in either of the previous two school years (2011–12 or 2012–13).

TTM Usage and Student Outcomes Related to Grade Retention

The results presented above are only descriptive, and do not adjust for differences in other characteristics between students who used the TTM system and students who did not use TTM, nor does it account for differences across campus types or environments. To account for these differences, and to assess the relationship between TTM usage and the likelihood of being retained in Grade 5 or 8 between 2013–14 and 2014–15, the evaluation team used a logistic regression with the outcome measure capturing whether the student was retained in grade. Similar the analyses presented for Research Question 1b above, statistical models were run using the following measures of system usage in 2013–14:

1) Student attempted at least one TTM lesson (compared to students who did not use TTM at all);

2) Student used the system at below recommended level, attempting at least 1 to 4 TTM lessons (compared to students who did not use TTM at all);

3) Student used the system at recommended "gateway" level, attempting 5 to 9 TTM lessons (compared to students who did not use TTM at all);

4) Student used the system at recommended "impact" level, attempting 10 to 14 TTM lessons (compared to students who did not use TTM at all);

5) Student used the system intensively at levels above recommended "impact" level, attempting 15 to 19, or 20 or more TTM lessons (compared to students who did not use TTM at all).

In addition to the categorical measures presented above, the evaluation team also included the number of TTM lessons attempted as an independent variable in the statistical model to predict changes in STAAR-Mathematics gain scores.

In the model, the evaluation team adjusted for student attributes, including two prior STAAR-Mathematics test scores (2012–13 and 2013–14), and whether they had been retained in grade in either of those two years. Analyses were conducted separately for Grades 5 and 8 to account for differences in the relationship between usage (i.e., one or more TTM lesson attempts) and usage intensity (i.e., number of lessons attempted and the following categories of TTM attempts: 5 to 9, 10 to 14, 14 to 19, and 20 or more), and the likelihood of being retained in grade in 2014–15. Furthermore, within each grade level, the analytic models were run separately for two groups of students:

1) First-time Grade 5 and Grade 8 students in 2013–14; and

2) First-time Grade 5 and Grade 8 students who were at risk of being retained in grade between 2013–14 and 2014–15.

The results of these analyses are presented in Table 3.6 and 3.7.[33]

After adjusting for other student, school, and district-level characteristics there were a few TTM usage and intensity of usage variables that were statistically significantly associated with retention outcomes. Among first-time Grade 5 students who attempted at least one TTM lesson in 2013–14 were significantly less likely to be retained in grade between 2013–14 and 2014–15. This finding, however, was not observed for Grade 8 students.

For first-time Grade 5 and Grade 8 students, each additional attempted lesson was associated with a significantly lower likelihood of being retained in grade between 2013–14 and 2014–15. This was also the case for first-time Grade 5 and Grade 8 students who attempted greater than 20 TTM lessons. The findings for TTM usage by first-time Grade 5 and Grade 8 students who were at risk of being retained in grade between 2013–14 and 2014–15, mirrored those presented above for all first-time Grade 5 and 8 students.

---

[33] Additional technical detail that describes the statistical model is provided in Appendix B.

GIBSON
CONSULTING GROUP

That is, the number of TTM lessons attempted and attempting 20 or more TTM lessons in 2013–14 was associated with a significantly lower likelihood of being retained in grade between 2013–14 and 2014–15 for both Grades 5 and 8. Attempting one or more TTM lessons was associated with a significantly lower likelihood of being retained in grade between 2013–14 and 2014–15 for Grade 5.

Table 3.6. Estimated Relationship between TTM Usage and Grade Retention between 2013–14 and 2014–15, First-Time Grade 5 and 8 Students in 2013–14

Grade	Usage Measure	B	SE	N
Grade 5	Attempted at least one lesson (compared to did not use)	*-0.133**	0.066	301,433
	Number of Attempts	*-0.656****	0.182	
	1–4 Attempts (compared to zero attempted lessons)	-0.076	0.080	
	5-9 Attempts (compared to zero attempted lessons)	-0.105	0.077	
	10-14 Attempts (compared to zero attempted lessons)	-0.100	0.104	
	15-19 Attempts (compared to zero attempted lessons)	-0.150	0.111	
	20+ Attempts (compared to zero attempted lessons)	*-0.184**	0.086	
Grade 8	Attempted at least one lesson (compared to did not use)	-0.167	0.096	224,954
	Number of Attempts	*-1.026****	0.245	
	1–4 Attempts (compared to zero attempted lessons)	-0.104	0.082	
	5-9 Attempts (compared to zero attempted lessons)	-0.094	0.079	
	10-14 Attempts (compared to zero attempted lessons)	0.111	0.103	
	15-19 Attempts (compared to zero attempted lessons)	0.049	0.114	
	20+ Attempts (compared to zero attempted lessons)	*-0.211**	0.088	

Source: TTM Session History table, 2013–14, Texas Education Agency. Public Education Information Management System, 2012–13, 2013–14, and 2014–15; State of Texas Assessments of Academic Readiness (STAAR), 2012–13, Texas Education Agency.

Note: Calculations only include students who met the following inclusion rules: enrolled at a campus where at least one student attempted TTM lessons in 2013–14; was a first-time Grade 5 or Grade 8 student in 2013–14; and 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. 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 B. The notation, B, in this table refers to the logit coefficient from the statistical model, SE refers to the standard error, and N refers to the number of observations used in the model.

GIBSON
CONSULTING GROUP

Table 3.7. Estimated Relationship between TTM Usage and Grade Retention between 2013–14 and 2014–15, First-Time Grade 5 and 8 Students in 2013–14 who were at Risk of Being Retained in Grade between 2013–14 and 2014–15

Grade	Usage Measure	B	SE	N
Grade 5	Attempted at least one lesson (compared to did not use)	*-0.146**	0.067	118,522
	Number of attempts	*-0.727****	0.199	
	1–4 Attempts (compared to zero attempted lessons)	-0.103	0.082	
	5-9 Attempts (compared to zero attempted lessons)	-0.094	0.079	
	10-14 Attempts (compared to zero attempted lessons)	-0.111	0.103	
	15-19 Attempts (compared to zero attempted lessons)	-0.148	0.114	
	20+ Attempts (compared to zero attempted lessons)	*-0.211**	0.088	
Grade 8	Attempted at least one lesson (compared to did not use)	-0.198	0.100	91,093
	Number of attempts	*-1.42****	0.271	
	1–4 Attempts (compared to zero attempted lessons)	-0.049	0.112	
	5-9 Attempts (compared to zero attempted lessons)	-0.210	0.170	
	10-14 Attempts (compared to zero attempted lessons)	-0.423*	0.175	
	15-19 Attempts (compared to zero attempted lessons)	*-0.578***	0.208	
	20+ Attempts (compared to zero attempted lessons)	-0.232	0.163	

Source: TTM Session History table, 2013–14. Public Education Information Management System, 2012–13, 2013–14, and 2014–15; State of Texas Assessments of Academic Readiness (STAAR), 2011–12 to 2012–13, Texas Education Agency.

Note: Results only include students who met the following inclusion rules: enrolled at a campus where at least one student was registered for TTM in 2013–14; was a first-time Grade 5 or Grade 8 student in 2013–14; and had a valid STAAR-Mathematics test score in 2011–12 and 2012–13. Only regular English and Spanish versions (i.e., not modified or alternate versions) were included. 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 the appendix B. The notation, B, in this table refers to the logit coefficient from the statistical model, SE refers to the standard error, and N refers to the number of observations used in the model.

As was the case with previous findings related to STAAR-Mathematics gains, it is important to keep in mind that the population of students who met the usage criteria, and who had sufficient data to be included in the statistical model, was small and observably different from the overall student population. For instance, in Grade 8, fewer than 3,000 students were retained in grade between 2012–13 and 2013–14. Consequently, the results should be interpreted with caution.

Summary of TTM Findings

The two research questions in this section explored the relationships between TTM system usage, and intensity of usage (i.e., dosage) in 2013–14 by at-risk students who were retained in grade between 2012–13 and 2013–14 and two important outcomes: 1) STAAR-Mathematics gains between 2012–13 and 2013–

GIBSON
CONSULTING GROUP

14; and 2) the probability of being retained in grade in 2014–15. Key descriptive and multivariate findings presented in this report section are presented below:

TTM System Usage

- Students in Grades 5 and 8 who were retained in grade between 2012–13 and 2013–14 used TTM less frequently during the 2013–14 school year than students who were not retained in grade during this period.

- On average, Grade 5 students who were retained in grade attempted a comparable number of TTM lessons throughout the course of the 2013–14 school year to students not retained (9.26 versus 9.77).

- On average, students in Grade 8 who were retained in grade between 2012–13 and 2013–14 attempted slightly fewer mathematics lessons in 2013–14 (2.70 TTM lessons) than students who were in Grade 8 for the first time in 2013–14 (3.79 TTM lessons).

- TTM system usage levels for student in Grade 8 were relatively low compared to Grade 5 students. Approximately 19% of Grade 8 students who were retained in grade between 2012–13 and 2013–14 attempted at least one TTM lesson compared to 25% of first-time Grade 8 students. Likewise, 10% of Grade 8 students who were retained in grade between 2012–13 and 2013–14 attempted 5 or more TTM lessons compared to 15% of Grade 8 students not retained in grade during this time period.

Mathematics Outcomes – Relationship between TTM Usage and STAAR-Mathematics Gains

- After adjusting for other student, school, and district-level characteristics, no statistical relationship was found between attempting one or more TTM lessons by Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 and STAAR-Mathematics gains in 2013–14. The comparison group for this analysis was students who did not use the TTM program at all.

- However, students who used TTM more often experienced larger STAAR-Mathematics gains than students using the system less frequently. The difference was statistically significant and moderate (.943 for Grade 5 and .905 for Grade 8) for each additional lesson attempted.

- Among students in Grades 5, who were retained in grade between 2012–13 and 2013–14, a positive and statistically significant relationship was observed between attempting 1 to 4 TTM lessons (i.e., a level of usage below that recommended by the TTM vendor) and 2013–14 STAAR-Mathematics gains. After adjusting for other student, school, and district-level characteristics only Grade 5 students who attempted between one and four lessons experienced statistically significant differences relative to students who did not use the system (.174 standard deviations higher). This statistical relationship did not hold for Grade 8 students.

- After adjusting for other student, school, and district-level characteristics, no statistically significant differences in 2013–14 STAAR-Mathematics gains were observed among Grade 5 and

GIBSON
CONSULTING GROUP

8 students retained in grade between 2012–13 and 2013–14 who used TTM at increasingly intensive levels 5 to 9 attempted lessons, 10 to 14 attempted lessons, 15 to 19 attempted lessons, or 20 or more attempted lessons), relative to students who did not use the TTM system at all.

Mathematics Outcomes – Relationship between TTM Usage and the Likelihood of Being Retained in Grade

- After adjusting for other student, school, and district-level characteristics there were a few TTM usage and intensity of usage variables that were statistically significantly associated with retention outcomes. First-time Grade 5 students who attempted at least one TTM lesson in 2013–14 were significantly less likely to be retained in grade between 2013–14 and 2014–15. This finding was not observed for Grade 8 students.

- For first-time Grade 5 and Grade 8 students, each additional attempted lesson was associated with a significantly lower likelihood of being retained in grade between 2013–14 and 2014–15. This was also the case for first-time Grade 5 and Grade 8 students who attempted greater than 20 TTM lessons.

- The findings for TTM usage by first-time Grade 5 and Grade 8 students who were at risk of being retained in grade between 2013–14 and 2014–15, mirrored those presented above for all first-time Grade 5 and 8 students. That is, for Grades 5 and 8, the number of TTM lessons attempted and attempting 20 or more TTM lessons in 2013–14 was associated with a significantly lower likelihood of being retained in grade between 2013–14 and 2014–15. Additionally, for Grade 5, attempting one or more TTM lessons was associated with a significantly lower likelihood of being retained in grade between 2013–14 and 2014–15.

Section 4 – Conclusion

This addendum report focuses on the relationship between Istation and TTM usage (and intensity of usage) in 2013–14 for students in Grades 5 and 8, and two key outcome measures: 1) STAAR-Reading and Mathematics gains between 2012–13 and 2013–14; and the likelihood of being retained in grade in 2014–15. Further, the analyses contained in this addendum focus on at-risk students in grades subject to high stakes testing (i.e., Grade 5 and 8 where students are required to meet state standards to be promoted to the next grade level).

Istation Results

Statistical analyses did not reveal a strong relationship between Istation system usage in 2013–14 and STAAR-Reading gains for Grade 5 and 8 students retained in grade between 2012–13 and 2013–14. However, among Grade 8 students who were retained in grade between 2012–13 and 2013–14, Istation usage at or above the annual level recommended by the vendors was found to be statistically significant and positively related to STAAR-Reading gains (when compared to students who did not use the system at the recommended level of 200 minutes for the 2013–14 school year). However, the proportion of at-risk students who used the Istation program at recommended levels was very low.

In addition, no statistical relationship was found between Istation usage in 2013–14 for first-time Grade 5 and 8 students at risk of being retained in grade, and the likelihood of being retained in grade in 2014–15.

TTM Results

A number of compelling findings were revealed in the TTM analyses conducted for this addendum report. While no statistical relationship was found between attempting one or more TTM lessons by Grade 5 and 8 students retained in grade between 2012–13 and 2013–14 and STAAR-Mathematics gains in 2013–14 (when compared to students who did not use the TTM program at all), a number of TTM usage metrics were associated with positive student outcomes related to STAAR-Mathematics gains or grade promotion. Students in Grade 5 and 8 who used TTM more often experienced larger STAAR-Mathematics gains than students using the system less frequently. The difference was statistically significant and moderate for each additional lesson attempted.

Among students in Grades 5, who were retained in grade between 2012–13 and 2013–14, a positive and statistically significant relationship was observed between attempting 1 to 4 TTM lessons (i.e., a level of usage below that recommended by the TTM vendor) and 2013–14 STAAR-Mathematics gains.

Lastly, after adjusting for other student, school, and district-level characteristics, first-time Grade 5 students who were at risk of being retained in grade between 2013–14 and 2014–15 and attempted at least one TTM lesson in 2013–14 were significantly less likely to be retained in grade between 2013–14 and 2014–15. Among Grade 5 and 8 students at risk of being retained in grade between 2013–14 and

GIBSON
CONSULTING GROUP

2014–15, each additional attempted lesson was associated with a significantly lower likelihood of being retained in grade between 2013–14 and 2014–15. This was also the case for comparable Grade 5 and 8 students who attempted greater than 20 TTM lessons.

Conclusion

The results presented in this addendum to Garland, et al (2015) suggest that Istation and TTM usage can positively impact student achievement results for at-risk students who use the systems at the recommended dosage levels. However, the proportion of students retained in grade, or at risk of being retained in grade who use the system intensively is quite low.

Appendix A – Istation Technical Section

Guiding Question 1a

To address Guiding Question 1a, the base model described in Appendix A of the Texas Students Using Curriculum Content to Ensure Sustained Success (SUCCESS) Comprehensive Evaluation Report (Garland, Shields, Booth, Shaw, Samii-Shore, 2015) was amended to investigate whether Grade 5 and Grade 8 students who were retained in grade between 2012–13 and 2013–14 and who used Istation at varying levels of dosage achieved larger STAAR-Reading gains relative to a comparison group of students who did not meet the usage criteria. This was done by restricting the analytic sample to only students who were retained in Grade 5 or Grade 8 between 2012–13 and 2013–14. Thus, the estimate produced from this analysis describes the association between Istation usage and STAAR-Reading scores *only for students who were retained in grade*.

Technical Material for Econometric Models

More formally,

$$A_{ijt} - A_{ijt-1} = \alpha SUCCESS_{it} + \beta X_{it-1} + \beta X_i + \pi_j + \upsilon_{ijt}$$

Where:

- $A_{ijt} - A_{ijt-1}$ is the difference in standardized assessment scores (subject-grade-year-prior decile standardized scores) at time *t*, for student *i*, attending school *j*

- $\alpha SUCCESS_{it}$ is an indicator of whether student *i* participated in a SUCCESS program at time *t*, and represents the mean program effect for students compared to the reference category, which depends upon the measure of program usage included in the statistical model

- 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 Istation 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 ν_{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 2a

Supplementary analysis of the relationship between Istation usage and the probability of being retained in grade

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 5 and 8 in 2013–14 with at least two valid test scores

1. This sample was used in Guiding Research Question 1a since only one valid prior test score was required. All valid cases in Grades 5 and 8 were included.

GIBSON
CONSULTING GROUP

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 300 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, and 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

Technical Material for Econometric Models

A logistic regression with student-, school-, and district-level covariates was fit for Guiding Research Question 2a. Logistic regression estimates the log odds of the outcome (here, being retained) 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 ($n_{ijd} = \log(\frac{\phi_{ijd}}{1-\phi_{ijd}})$), where 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

$$n_{ijd} = \beta_0 + \beta_1 SUCCESS_{ijd} + \beta_k X_{ijdt-1} + \beta_k Q_{jdt-1} + \beta_k Y_{dt-1} + e_{ijd} + \mu_{jd} + \nu_d \qquad (3)$$

Where:

- $\beta_1 SUCCESS_{ijd}$ is the measure of student *I's* participation in a SUCCESS program in 2013–14

- $\beta_k X_{ijdt-1}$ is a vector of student-level characteristics from the prior year and current year, including

GIBSON
CONSULTING GROUP

- – 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

- – STAAR-Reading scaled score from the first administration in March 2012–13

- – Sex

- – Race

- – Current ELL student

- – Test language of 2013–14 STAAR test

- – Test language of 2012–13 STAAR test

- – Economic Disadvantaged status

- – Received any accommodation on the STAAR administration in 2013–14

- – Received any accommodation on the STAAR administration in 2012–13

- – Student was ever retained in grade

- – Student was considered at risk

- – Immigrant indicator

- – Special Education indicator

- $\beta_k Q_{jdt-1}$ is a vector of school-level characteristics

  - – Title I status in 2013–14

  - – Campus accountability rating in 2013–14

  - – 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 2013–14

  - – Percentage of students who are classified as at-risk in 2013–14

  - – Percentage of students classified as economically disadvantaged in 2013–14

- $\beta_k Y_{dt-1} \pi_{jd}$ is a vector of district-level characteristics

  - – Education Service Center (ESC) region

  - – District type derived from http://goo.gl/gSoiog

- e_i is a random error term for student i in school j in district d

GIBSON
CONSULTING GROUP

Propensity Score Reweighted Estimates of the Relationship Between Istation Usage and Retention

To investigate the robustness of the multivariate results presented in the main section of the response for Guiding Research Question 2a, 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, or meeting the grade level prescribed Istation usage threshold (Table A1). 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. The results are consistent with the unweighted results.

Table A1. Propensity Score Reweighted Effects of Istation Usage on the Likelihood of Grade 5 or Grade 8 Retention between 2013–14 and 2014–15, by Student Grade Level, First-Time Grade 5 or Grade 8 Students

		B	SE	N
Grade 5	Used at least once	0.129	0.082	320,066
	Met Istation usage threshold (compared to below threshold)	0.053	0.060	320,066
Grade 8	Used at least once	0.013	0.102	318,948
	Met Istation usage threshold (compared to below threshold)	-0.174	0.116	318,948

Source: Istation Session History table, 2013–14. Public Education Information Management System, 2012–13, 2013–14, and 2014–15; State of Texas Assessments of Academic Readiness (STAAR), 2011–12 to 2012–13, Texas Education Agency.

Note: Results only include students who met the following inclusion rules: enrolled at a campus where at least one student was registered for Istation in 2013–14; was a first-time Grade 5time Grade 5 or Grade 8 student 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. Estimates were derived from a logistic regression with cluster-robust standard errors at the campus level. 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) or meeting the Istation usage threshold (compared to below the threshold). 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. The notation *B* refers to the logit coefficient from the statistical model and SE refers to the standard error.

GIBSON
CONSULTING GROUP

Table A2. Propensity Score Reweighted Effects of Istation Usage on the Likelihood of Grade 5 or Grade 8 Retention between 2013–14 and 2014–15, by Student Grade Level, First-Time Grade 5 or Grade 8 Students at Risk of Being Retained in Grade

		B	SE	N
Grade 5	Used at least once	0.151	0.085	103,844
	Met Istation usage threshold (compared to below threshold)	0.051	0.061	103,844
Grade 8	Used at least once	0.011	0.113	95,464
	Met Istation usage threshold (compared to below threshold)	-0.178	0.119	95,464

Source: Istation Session History table, 2013–14. Public Education Information Management System, 2012–13, 2013–14, and 2014–15; State of Texas Assessments of Academic Readiness (STAAR), 2011–12 to 2012–13, Texas Education Agency.

Note: Results only include students who met the following inclusion rules: enrolled at a campus where at least one student was registered for Istation in 2013–14; was a first-time Grade 5time Grade 5 or Grade 8 student 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. Estimates were derived from a logistic regression with cluster-robust standard errors at the campus level. 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) or meeting the Istation usage threshold (compared to below the threshold). 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. * $p<0.05$, ** $p<0.01$, *** $p<0.001$. The notation *B* refers to the logit coefficient from the statistical model and SE refers to the standard error.

GIBSON
CONSULTING GROUP

Appendix B – TTM Technical Section

Research Question 1b

To address Guiding Question 1b, the base model described in the Texas Students Using Curriculum Content to Ensure Sustained Success (SUCCESS) Comprehensive Evaluation Report has been amended as outlined below. This model assesses whether Grade 5 and Grade 8 students who were retained in grade between 2012–13 and 2013–14 and who used TTM at varying levels of dosage achieved larger STAAR-Mathematics gains relative to a comparison group of students who did not meet the usage criteria.

These models only included students who were retained in Grade 5 or Grade 8 between 2012–13 and 2013–14. Thus, the estimate produced from this analysis describes the association between TTM usage and STAAR-Mathematics scores *only for students who were retained in grade*. Models were fit separately by grade and program participation dosage measure.

Econometric Specification

$$A_{ijt} - A_{ijt-1} = \alpha SUCCESS_{it} + \beta X_{it-1} + \beta X_i + \pi_j + \upsilon_{ijt}$$

Where:

- $A_{ijt} - A_{ijt-1}$ difference in standardized assessment scores (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.

- βX_{it-1} 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 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

GIBSON
CONSULTING GROUP

–   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 school fixed effect

▪   and υ_{ij} random disturbance term

The base, or reference, category for student groups were:

▪   ELL status: Not ELL

▪   Economic disadvantaged status: Not economically disadvantaged

▪   Ethnicity: Hispanic

Research Question 2b

Construction of the Propensity Score and Implementation of the Regression

The analysis in this section utilized propensity score reweighting (Nichols, 2007) where a student's likelihood of being a member of the treatment group, 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 5 and 8 in 2013–14 with *at least* two valid test scores

1. This sample was previously used since only one valid prior test score was required. All valid cases in Grades 5 and 8 were included.

2. Fit propensity score model estimating the conditional probability of treatment using a logistic regression based on three separate measures of TTM use in 2013–14:

    a. Students who recorded at least attempted lesson compared to students who did not record any attempted lessons

    b. The number of attempted student lessons compared to students who did not record any attempted lessons

c. An ordinal measure of categories of the number of attempted lessons, bounded by TTM recommended usage threshold (e.g., 5 to 10 or 10 to 15 lessons attempted in a year) compared to students who did not record any attempted lessons

3. 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 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: $\hat{\lambda}/(1-\hat{\lambda})$

GIBSON
CONSULTING GROUP

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

Technical Material for Econometric Models

A logistic regression with student-, school-, and district-level covariates was fit for Guiding Research Question 2b. Logistic regression estimates the log odds of the outcome (here, being retained in grade) 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 ($n_{ijd} = \log(\frac{\phi_{ijd}}{1-\phi_{ijd}})$), where 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

$$n_{ijd} = \beta_0 + \beta_1 SUCCESS_{ijd} + \beta_k X_{ijdt-1} + \beta_k Q_{jdt-1} + \beta_k Y_{dt-1} + e_{ijd} + \mu_{jd} + v_d \qquad (3)$$

Where:

- $\beta_1 SUCCESS_{ijd}$ is the measure of student *I's* participation in a SUCCESS program in 2013–14

- $\beta_k X_{ijdt-1}$ 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

  – STAAR-Mathematics scaled score from the first administration in March 2012–13

  – Sex

- – Race

- – Current ELL student

- – Test language of 2013–14 STAAR test

- – Test language of 2012–13 STAAR test

- – Economic Disadvantaged status

- – Received any accommodation on the STAAR administration in 2013–14

- – Received any accommodation on the STAAR administration in 2012–13

- – Student was ever retained in grade

- – Student was considered at risk

- – Immigrant indicator

- – Special Education indicator

- ▪ $\beta_k Q_{jdt-1}$ is a vector of school-level characteristics

  - – Title I status in 2013–14

  - – Campus accountability rating in 2013–14

  - – 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 2013–14

  - – Percentage of students who are classified as at-risk in 2013–14

  - – Percentage of students classified as economically disadvantaged in 2013–14

- ▪ $\beta_k Y_{dt-1} \pi_{jd}$ is a vector of district-level characteristics

  - – Education Service Center (ESC) region

  - – District type derived from http://goo.gl/gSoiog

- ▪ e_i is a random error term for student *i* in school *j* in district *d*

Propensity Score Reweighted Estimates of the Relationship Between TTM Usage and Retention

To investigate the robustness of the multivariate results presented in the main section of the response for Guiding Research Question 2b, 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 TTM for at least lesson during the 2013–14 school year

GIBSON
CONSULTING GROUP

(parallel to Table B3). 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 TTM match the distribution of participant students. The results presented in Table B1 and B2 are consistent with the unweighted results.

Table B1. Propensity Score Reweighted Effects of TTM Usage on the Likelihood of Grade 5 or Grade 8 Retention between 2013–14 and 2014–15, by Student Grade Level, First-Time Grade 5 or Grade 8 Students

Grade	Usage Measure	B	SE	N
Grade 5	Attempted at least one lesson (compared to did not use)	*-0.133**	0.066	301,433
	1–4 Attempts (compared to zero attempted lessons)	-0.076	0.080	
	5-9 Attempts (compared to zero attempted lessons)	-0.105	0.077	
	10-14 Attempts (compared to zero attempted lessons)	-0.100	0.104	
	15-19 Attempts (compared to zero attempted lessons)	-0.150	0.111	
	20+ Attempts (compared to zero attempted lessons)	*-0.2184**	0.086	
Grade 8	Attempted at least one lesson (compared to did not use)	-0.169	0.098	212,337
	1–4 Attempts (compared to zero attempted lessons)	-0.077	0.109	
	5-9 Attempts (compared to zero attempted lessons)	-0.379	0.518	
	10-14 Attempts (compared to zero attempted lessons)	0.135	0.159	
	15-19 Attempts (compared to zero attempted lessons)	0.386	0.207	
	20+ Attempts (compared to zero attempted lessons)	*-0.341**	0.171	

Source: TTM Session History table, 2013–14. Public Education Information Management System, 2012–13, 2013–14, and 2014–15; State of Texas Assessments of Academic Readiness (STAAR), 2011–12 to 2012–13, Texas Education Agency.

Note: Results only include students who met the following inclusion rules: enrolled at a campus where at least one student was attempted TTM lessons in 2013–14; was a first-time Grade 5 or Grade 8 student in 2013–14; and 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. 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 the appendix. Observations were reweighted by a propensity score derived for a logistic regression model predicting the probability of using the system (compared to not at all) or meeting the TTM usage categories (compared to no usage). 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. * $p<0.05$,

GIBSON
CONSULTING GROUP

** p<0.01, *** p<0.001. B refers to the logit coefficient from the statistical model, SE refers to the standard error, and N refers to the number of observations used in the model.

Table B2. Propensity Score Reweighted Effects of TTM Usage on the Likelihood of Grade 5 or Grade 8 Retention between 2013–14 and 2014–15, by Student Grade Level, First-Time Grade 5 or Grade 8 Students at Risk of Being Retained in Grade

Grade	Usage Measure	B	SE	N
Grade 5	Attempted at least one lesson (compared to did not use)	*-0.173***	0.063	118,522
	1–4 Attempts (compared to zero attempted lessons)	-0.105	0.083	
	5-9 Attempts (compared to zero attempted lessons)	-0.134	0.080	
	10-14 Attempts (compared to zero attempted lessons)	-0.183	0.105	
	15-19 Attempts (compared to zero attempted lessons)	-0.181	0.107	
	20+ Attempts (compared to zero attempted lessons)	*-0.228***	0.087	
Grade 8	Attempted at least one lesson (compared to did not use)	*-0.201**	0.099	84,676
	1–4 Attempts (compared to zero attempted lessons)	-0.062	0.113	
	5-9 Attempts (compared to zero attempted lessons)	-0.189	0.175	
	10-14 Attempts (compared to zero attempted lessons)	-0.471	0.295	
	15-19 Attempts (compared to zero attempted lessons)	*-0.514**	0.211	
	20+ Attempts (compared to zero attempted lessons)	-0.194	0.168	

Source: TTM Session History table, 2013–14. Public Education Information Management System, 2013–14, 2012–13, 2013–14, and 2014–15; State of Texas Assessments of Academic Readiness (STAAR) data, 2011–12 to 2013–14, Texas Education Agency.

Note: Results only include students who met the following inclusion rules: enrolled at a campus where at least one student was attempted TTM lessons in 2013–14; was a first-time Grade 5 or Grade 8 student in 2013–14; and 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. 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 the appendix. Observations were reweighted by a propensity score derived for a logistic regression model predicting the probability of using the system (compared to not at all) or meeting the TTM usage categories (compared to no usage). 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 logit coefficient from the statistical model, SE refers to the standard error, and N refers to the number of observations used in the model.

Appendix C – Study Limitations

It is important to consider the following caveats related to this study's methodology when considering implications of the results discussed above, and as presented in the rest of this addendum to the Texas Students Using Curriculum Content to Ensure Sustained Success (SUCCESS) Comprehensive Evaluation Report (Garland, Shields, Booth, Shaw & Samii-Shore, 2015).

Non-Random assignment of students to the SUCCESS interventions: Through the Texas SUCCESS program, 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 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.[34]

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.

Small number of students who used the systems at threshold levels: In each of the findings presented, it is important understand that the population of students who met the usage criteria, and who had sufficient data to be included in the statistical models, was small and observably different from the overall student population. Consequently, the amount of uncertainty in these estimates (which is reflected in the size of the standard error of the estimates) is large, and the results should be interpreted with caution.

---

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

GIBSON
CONSULTING GROUP

Appendix D – References

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press.

Garland, M., Shields, J., Booth, E., Shaw, S, and Samii-Shore, K. (2015). Texas SUCCESS Comprehensive Evaluation Report. Austin, TX: Texas Education Agency. Retrieved from http://tea.texas.gov/Reports_and_Data/Program_Evaluations/Reading_Math_Science_and_Technology_Initiatives/Program_Evaluation__Reading,_Math,_Science,_and__Technology_Initiatives/

Gelman, A. & Hill, J. (2007*). Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

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

Nichols, A. (2007). Causal inference with observational data. *Stata Journal*, *7*(4), 507-541.

Texas Education Agency (2015). Grade-Level Retention and Student Performance in Texas Public Schools, 2012-13. Austin, TX: Texas Education Agency. Retrieved from http://tea.texas.gov/acctres/retention_index.html#reports_2010-11