

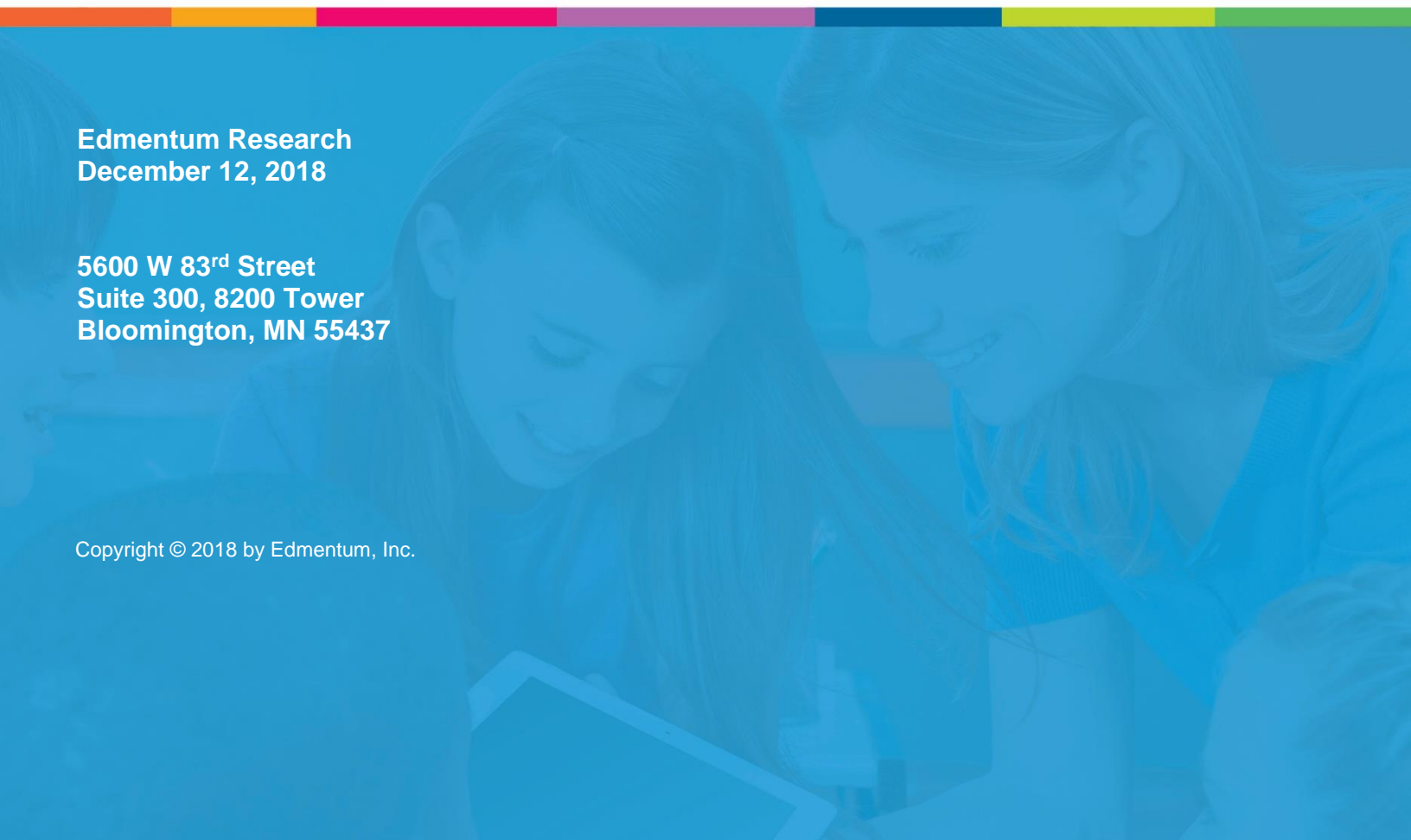


Research Report: Impacts of the Use of Study Island Practice – Arizona School District

Edmentum Research
December 12, 2018

5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

Copyright © 2018 by Edmentum, Inc.



Executive Summary

Study Island is a practice and assessment tool that provides state-standards-aligned opportunities for students to practice their skills. It features a system of continual assessments with immediate feedback to adjust instruction and learning. When educators integrate Study Island into their instructional practices, it acts as a formative, ongoing assessment tool that provides students with a platform to practice or demonstrate their knowledge of taught standards. This approach reflects the elements of formative assessments as a process for monitoring progress and adjusting instruction. Research on formative assessment and progress-monitoring practices has demonstrated positive outcomes for student achievement (Bangert-Drowns, Kulik, & Kulik, 1991; Black & Wiliam, 1998; Fuchs & Fuchs, 1986; January et al., 2018; Stecker, Lembke, & Foegen, 2008; Stiggins, 1999; Van Norman, Nelson, & Parker, 2016; Wolf, 2007).

Located in an Arizona metro area with an enrollment of almost 10,000 students, the Arizona school district under investigation was a Study Island partner during the 2016–17 school year. A large proportion of the students in the district live in households that fall below the poverty line, and a majority of students are Hispanic (National Center for Education Statistics, n.d.; Dun & Bradstreet, 2018). All districts in Arizona participate in the state's accountability system. The Arizona A–F Accountability System holds schools and districts accountable to a range of measures, including student proficiency and growth, English language proficiency and growth, high school readiness, career and college readiness, and high school graduation rates. As part of their accountability, the Arizona Measurement of Educational Readiness to Inform Teaching (AzMERIT) is administered annually to students in grades 3–8 for English language arts (ELA) and math and as end-of-course tests in high schools. AIMS Science, an assessment to measure science proficiency in Arizona Academic Content Standards in Science, is administered to students in grades 4, 8, and 10. Assessment data show that students from the district tend to perform at levels lower than the state average.

This study is intended to provide a research basis for Study Island in terms of the research literature and analyses of the district students' level of usage and performance data within Study Island compared to their performance on the AzMERIT.

Through a series of descriptive and statistical analyses, which include pseudo-controls through Propensity Score Matching (a process to create quasi control and treatment groups of equivalent ability), the findings in this study suggest there are discernable and statistically significant positive impacts on AzMERIT scores for students participating in Study Island, particularly in math.

Generally, implementation and use of Study Island in the district vary by grade and content area. Students appear to be answering a moderate number of questions and spending a fair amount of time over the course of the school year. Where students spend more time, answer more questions, and spread their time over active weeks, positive differences are observed, specifically in math. This is evident in the significant differences in mean scale scores and impact data for math in grades 4, 5, and 7.

These analyses are clearly impacted by the quality and approach with which schools use Study Island. It would be an important next step to understand the qualitative differences in implementation approaches, such as for grade 5 students. Understanding the methods will help guide implementations that drive evidence-based, positive outcomes for students.

Introduction

Education is a key indicator for individual and societal progress. As the Organisation for Economic Co-operation and Development (2012) put it, “School failure penalises a child for life . . . and imposes high costs on society” (p. 3). At Edmentum, our mission is to be educators’ most trusted partner in creating successful student outcomes everywhere learning occurs.

Over the years, legislation has been enacted to provide federal guidance and requirements to states in support of improving educational outcomes. From No Child Left Behind to the 2015 reauthorization of the Every Student Succeeds Act (ESSA), accountability of student achievement has been a critical focus. While ESSA continues to require states to assess students annually, the legislation now allows for some flexibility in the kinds of measures states may use, including measures of growth and of achievement. Specifically, assessments can now be “innovative” and include “multiple up-to-date measures of student academic achievement, including measures that assess higher-order thinking skills and understanding, which may include measures of student academic growth and may be partially delivered in the form of portfolios, projects, or extended performance tasks” (n.p.).

This new flexibility around accountability measures, particularly in terms of growth, has increased the focus on educational products to support educators in delivering targeted instruction and programs to monitor student progress throughout the school year, with particular attention to progress relative to state assessment expectations of standards-based achievement.

The Arizona A–F Accountability System holds schools and districts accountable to a range of measures, including student proficiency and growth, English language proficiency and growth, high school readiness, career and college readiness, and high school graduation rates. To support schools, Arizona’s Department of Education provides the Arizona College and Career Readiness Standards and the Arizona Department of Education Balanced Assessment Framework as resources to support student achievement, where the focus includes standards, assessments, curriculum framework, instruction, and materials and resources (as well as safe and supportive schools). As part of their accountability, the Arizona Measurement of Educational Readiness to Inform Teaching (AzMERIT) is administered annually to students in grades 3–8 for English language arts (ELA) and math and as end-of-course tests in high schools. AIMS Science, an assessment to measure science proficiency in Arizona Academic Content Standards in Science, is administered to students in grades 4, 8, and 10. The assessments have been built to align to Arizona State Standards and to provide student-level achievement scores and relevant placement into one of four proficiency categories: Minimally Proficient, Partially Proficient, Proficient, and Highly Proficient.

The Arizona school district highlighted in this paper was a Study Island partner during the 2016–17 school year. This study is intended to provide a research basis for Study Island in terms of the research literature and analyses of the district students’ level of usage and performance data within Study Island compared to their performance on the AzMERIT.

Literature Review

Formative assessment is a process for monitoring progress and adjusting instruction as a result of the feedback (Heritage, 2010). Research on formative assessment and progress-monitoring practices has demonstrated positive outcomes for student achievement (Bangert-Drowns, Kulik, & Kulik, 1991; Black & Wiliam, 1998; Fuchs & Fuchs, 1986; January et al., 2018; Stecker et al., 2008; Stiggins, 1999; Van Norman et al., 2016; Wolf, 2007), particularly for students with lower achievement (Black & Wiliam, 1998; January et al., 2018), as well as in building student confidence (Stiggins, 1999). Monitoring student progress is at the heart of such programs as Curriculum Based Measurement (CBM) (Deno, 1985; Fuchs & Fuchs, 1999), Response to Intervention (RtI), and the more recent movement to consider RtI as part of a Multi-Tier System of Supports (MTSS) (Gresham, Reschly, & Shinn, 2010).

Key to the success of monitoring progress is the action taken as a result of the feedback and information about progress that is provided (Duke & Pearson, 2002). Research shows that when an instructional feedback loop is applied in practice and instruction is modified based on student performance, student learning is accelerated and improved (Jinkins, 2001; Wiliam, Lee, Harrison, & Black, 2004), especially when feedback is used quickly and impacts or modifies instruction on a day-by-day or minute-by-minute basis (Leahy, Lyon, Thompson, & Wiliam, 2005) and provides students with opportunities to learn from the assessment (Kilpatrick, Swafford, & Bradford, 2001).

Although generally providing feedback to teachers and students regarding student performance can consistently enhance achievement (Adams & Strickland, 2012; Baker, Gersten, & Lee, 2002; Chase & Houtmanfar, 2009), meta-analytic research indicates that the timeliness and the type of feedback are critical within applied learning settings. Kulik and Kulik (1988) found that immediate feedback of results has a positive effect on student achievement within classroom settings, especially on applied learning measures such as frequent quizzes. Dihoff, Brosvic, Epstein, and Cook (2004) concluded that immediate feedback was even more effective when it immediately followed each answer a student provided. Bangert-Drowns, Kulik, Kulik, and Morgan (1991) showed that timely feedback can correct future errors when it informs the learner of the correct answer, and Kulhavy and Stock (1989) found immediate feedback especially helpful when students were confident in their answers. Multiple studies have found that feedback that also provides an explanation of the correct answer is the most effective (Adams & Strickland, 2012; Chase & Houtmanfar, 2009; Dihoff et al., 2004; Lee, Lim, & Grabowski, 2010; Marzano, Pickering, & Pollack, 2001). Through their meta-analysis, Marzano et al. (2001) additionally concluded that feedback is best when it encourages students to keep working on a task until they succeed and tells students where they stand relative to a target level of knowledge instead of how their performance ranks in comparison to other students.

Although most of the research literature has focused on the effect of teacher-provided feedback or feedback from classroom-based assessments, research has shown that computers are also effective tools for providing feedback. In their meta-analysis, Baker, et al. (2002) concluded that although using computers to provide ongoing progress monitoring feedback was effective (Effect Size [ES] = 0.29), using a computer to provide instructional recommendations based on these results was even more effective (ES = 0.51), suggesting that the combination of the two factors may be the most beneficial practice.

Taken together, these results suggest that a cycle of ongoing feedback followed by remediation and further assessment contributes to increases in student achievement. Study Island incorporates a short-cycle assessment feedback loop into its design through a system of continual assessment, immediate feedback, and quick remediation. When educators integrate Study Island into their instructional practices, it acts as a formative, ongoing assessment tool that provides students with a platform to practice or demonstrate their knowledge of taught standards. During program implementation, students answer questions that correspond to grade-specific state standards and learning objectives within state-tested content areas. When students answer a question, they immediately learn if the answer they provided is correct or not. When a student gets a question wrong, an explanation of the correct answer automatically appears, offering ongoing remediation to students who need it. At the end of each session, students can revisit the questions they missed and can seek learning opportunities for those questions. Students also have the option to engage in additional learning opportunities through lessons on the standards that are available at the beginning and end of a study session.

Additionally, Study Island provides in-depth reports of student performance data to students, teachers, and administrators. Specifically, reports provide the following information:

- Students can learn where they stand relative to specific proficiency goals

- Teachers can instantly use the reports of individual student performance data to provide additional remediation where needed within a general classroom instruction setting
- Administrators can use the reports to access summative data to determine if students are meeting benchmark standards over time

The availability of real-time achievement data allows for both quick remediation and the identification of trends in individual student performance, helping teachers to create personalized instructional paths based on demonstrated student need. Furthermore, technology-based programs such as Study Island that immediately utilize student performance data can also shift instruction or practice to the appropriate level required by a student to ensure more effective practice and to meet individual student needs. Such personalization of instructional materials promotes learning through a reduction of the cognitive load (i.e., working memory activity) required to complete a task (Kalyuga & Sweller, 2005), and research from a variety of learning environments shows that personalized instruction can lead to more efficient training and higher test performance than fixed-sequence, one-size-fits-all programs (Camp, Paas, Rickers, & van Merriënboer, 2001; Corbalan, Kester, & van Merriënboer, 2006; Kalyuga & Sweller, 2005; Salden, Paas, Broers, & van Merriënboer, 2004).

Study Island uses technology to provide students with both remediation or practice at lower levels and a customized learning experience based on demonstrated need. In many cases throughout the program, if students score 40% or lower in a session, the program cycles students down to lower levels to give them practice at levels that are building blocks for higher-level skills. Once students demonstrate success at a lower level, the program cycles students back up to the higher level.

Through this process, Study Island creates individual learning trajectories for students to follow. Study Island's administrative and reporting features allow teachers and administrators to constantly monitor how students are progressing through these personalized trajectories toward mastery of required benchmarks and standards. If students begin to fall below or exceed certain levels of achievement, teachers can prescribe additional practice at specific levels through the program and continue to monitor students' progress, or they can provide additional instruction or remediation within the classroom. Therefore, when teachers integrate Study Island into their curriculum, it essentially allows for individualized, differential instruction that could otherwise be difficult for one teacher alone to provide.

Using Study Island to track content mastery and individual changes in achievement concurrently, a teacher can efficiently determine if a student has significantly improved over time and if that improvement was enough to meet specific content benchmarks and standards. Weiss and Kingsbury (1984) concluded that the combination of these methods is particularly useful for identifying students who may begin the year at the same level but do not respond to instruction at the same rate. This methodology allows for the immediate notification of necessary remediation and intervention.

Research Questions

This study seeks to understand the association, if any, between students' use and their performance, both within the ongoing assessments in Study Island and on the state summative assessments. Specifically, this study seeks to answer the following research questions:

1. How did students in this Arizona school district use Study Island during the 2016–17 school year?

2. Were there significant mean differences in the AzMERIT state test scores between students who used Study Island and those who did not?
3. Was there a significant relationship between AzMERIT proficiency-level categorization and Study Island usage?

To answer these research questions, a description of Study Island and the AzMERIT is provided, followed by an analysis of the impact of Study Island usage on AzMERIT performance.

Components of Study Island

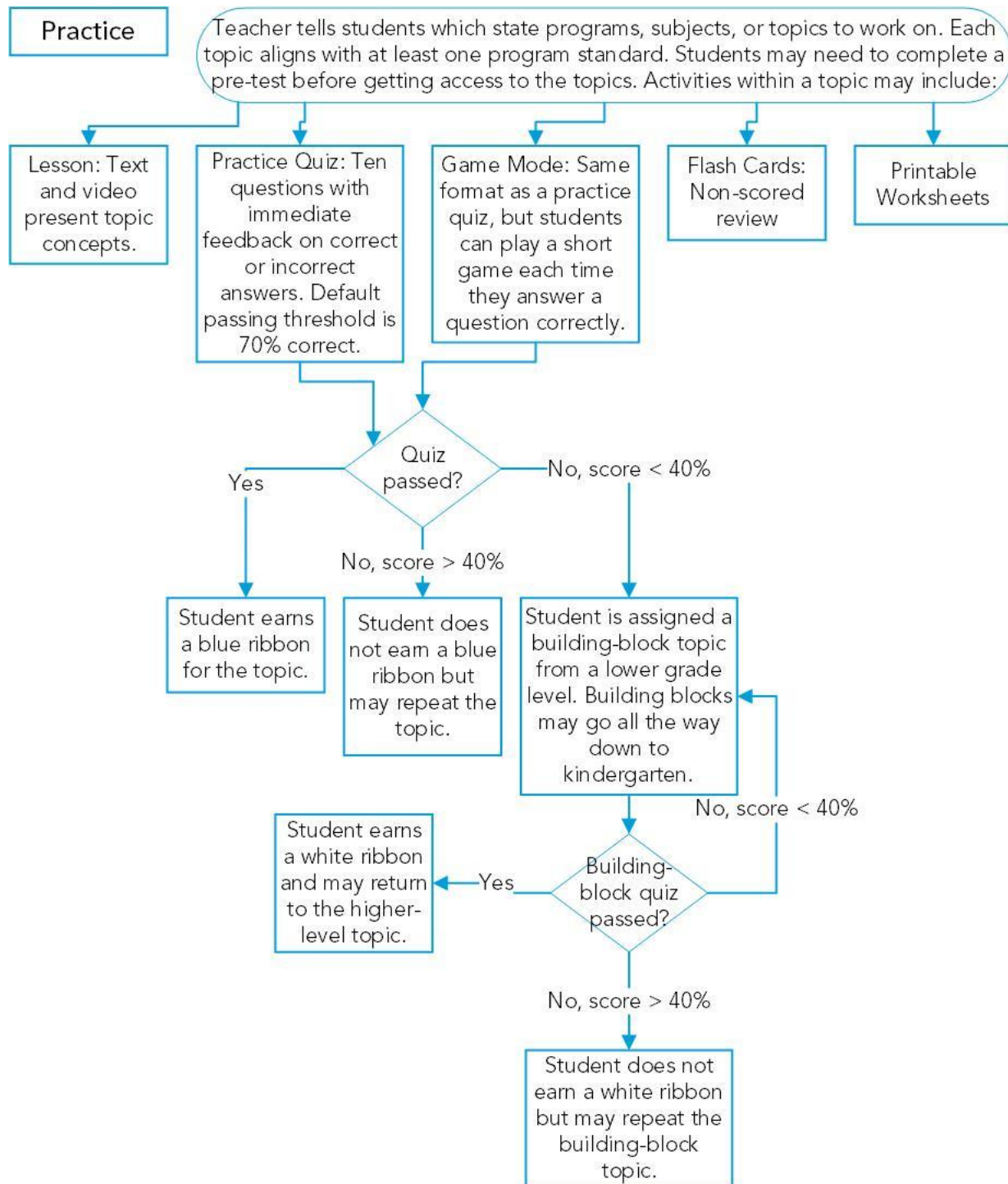
Study Island uses a comprehensive system of instructional and assessment tools to provide in-depth practice and feedback regarding student progress on content standards. The program is structured around *topics*. A *topic* is a grouping of conceptual material within a subject and grade level that is aligned to one or more state standards. Table 1 provides the total number of topics available by grade and content area for Arizona. While the current study focuses on grade 3–8, Study Island topics are also available in grades 2 and 9–11.

Table 1. Number of Study Island Topics Aligned to Arizona Standards

Grade	Number of Topics		
	ELA	Math	Science
2	43	20	
3	49	28	11
4	52	34	16
5	47	26	12
6	42	35	12
7	45	26	12
8	46	26	17
9	44	38	11
10	44	13	
11	45	28	

Resources offered within each topic may include assessments, practice tools, lessons, and instructional materials (games, flash cards, practice items, printables, etc.). The practice assessments are essentially ten-question quizzes. As students take a quiz, they receive immediate feedback on incorrect answers and earn a blue ribbon when they answer 70% of the questions correctly. (Teachers can adjust the 70% threshold as appropriate for their students.) Figure 1 visualizes the student experience within Study Island.

Figure 1. Student Experience, Study Island



The formative, short-cycle practice assessments may include multiple choice items (MC), technology-enhanced items (TE) and optional constructed response items. MC and TE items are scored by the system while CR items are scored by the teacher; results are incorporated into the system's information. Furthermore, Study Island includes reports of performance results that are instantly and constantly

available through the online system. These reports provide instructors and administrators with continual access to information regarding students' instructional weaknesses, their progress toward overcoming these weaknesses, and their eventual mastery of learning objectives.

Arizona's Measurement of Educational Readiness to Inform Teaching (AzMERIT)

Given the focus on accountability, a primary research question of this study relates how use of Study Island impacts students' end-of-year state test scores. The Arizona Measurement of Educational Readiness to Inform Teaching (AzMERIT) assesses students in grades 3–8 in mathematics (math) and English language arts (ELA), with a separate assessment, AIMS Science, administered to students in grades 4, 8, and 10 to assess science proficiency. The AzMERIT is a standards-based, criterion-referenced, summative assessment that is aligned with Arizona's College and Career Ready Standards (ACCRS), assessing basic knowledge, cognitive skills, and analytical thinking skills in writing, analysis, and problem-solving across subjects. These assessments are intended to provide information for use in school and district accountability systems and to improve curricular and instructional practice to help students achieve proficiency in the standards. For the purposes of this paper, we focus on the AzMERIT (ELA and math).

To measure math and ELA standards, the AzMERIT is composed of various types of assessment items and is developed according to a test blueprint specifying the range and depth for each content strand and standard that each test administration for each grade level is required to cover. AzMERIT assessments include a combination of multiple-choice (MC), technology-enhanced (TE), and constructed-response (CR) items. Specific TE items include the following:

- editing task (math)
- editing task choice (math, ELA)
- equation editor (math)
- evidence-based selected response (ELA)
- graphic response item display (math, ELA)
- hot text (math, ELA)
- matching item (math, ELA)
- multi-select (math, ELA)
- open response (math, ELA)
- table item (math)

Additionally, the ELA test includes a writing section where students are asked to write an essay in response to passages embedded in the assessment. Responses are rated by two human raters in three categories: purpose, focus, and organization; evidence and elaboration; and conventions and editing (Arizona Department of Education, 2016a; Arizona Department of Education, 2016b).

The AzMERIT reports student-level scale scores and performance-level classifications (Minimally Proficient, Partially Proficient, Proficient, and Highly Proficient). Scale scores were derived via the Rasch item response theory (IRT) model for each grade and content area. In order to help monitor student growth throughout time, a vertical linking design was also used to link common items across grades. This study will focus on scale scores within grade and performance-level classifications.

Sample

This study was conducted on a convenience sample of students from 18 schools within the district that were Study Island partners during the 2016–2017 academic year. The school district under study here is located in an Arizona metro area with a total enrollment of almost 10,000 students. A large proportion of the students in the district live in households that fall below the poverty line, and a majority of students are Hispanic (National Center for Education Statistics, n.d.; Dun & Bradstreet, n.d.). The district provided student-level AzMERIT data from the previous two years' administrations (Spring 2016 and Spring 2017) and demographic information for this study. The data were then matched to Study Island data via unique student identifiers.

As with any sample, it is important to understand how well the sample might generalize to other samples or the population overall. Table 2 provides the demographic make-up of the district overall compared to the state. More than a third of children in the district are in households below the poverty line, and 81% of students are Hispanic, many more than the 45% state average. Table 3 provides the demographic make-up of the sample for this study. It appears the students using Study Island in the sample are comparable to the district as a whole.

Table 2. District Demographics Compared to State Average

	District (%) ^a	State Average (%) ^a	Difference (District vs. State)
Individualized Education Program (IEP) ^a	13.1	12.0	+1.1
Percent of Children Below Poverty Line ^b	34.5	21.9	+12.6
Hispanic ^a	80.9	44.9	+36.0
Black ^a	12.6	5.3	+7.3
White ^a	3.1	39.5	-36.4
Two or More Races ^a	1.0	2.7	-1.7
Asian or Asian/Pacific Islander ^a	1.5	4.4	-2.9
American Indian/Alaska Native ^a	1.5	4.5	-3.0
Hawaiian/Pacific Islander ^a	0.1	0.3	-0.2

^a Ethnicity percentages may not add up to 100% because of rounding.

^a Data Source: National Center for Education Statistics Common Core of Data (CCD) "Local Education Agency (School District) Universe Survey LEP Data" 2015-16 v. 1a; "Local Education Agency (School District) Universe Survey Membership Data" 2015-16 v. 1a; "Local Education Agency (School District) Universe Survey Special ED Data" 2015-16 v. 1a; "Public Elementary/Secondary School Universe Survey Free Lunch Data" 2015-16 v. 1a; "Public Elementary/Secondary School Universe Survey Geo Data" 2014-15 v. 1a.

^b Data Source: Dun and Bradstreet, MDR Educational Data. 2018. License purchased by Edmentum

Table 3. Sample Demographics of Study Island Use

Variable	Category	Complete 2017 District Sample		Sample of SI Users	
		N	%	N	%
Race / Ethnicity	American Indian - Alaska Native	76	1	33	1
	Asian	47	1	19	1
	Black - African American	833	13	433	12
	Hispanic or Latino	5,060	80	2,892	81
	Multi-Racial	56	1	35	1
	Native Hawaiian - Other Pacific Islander	9	0	7	0

		Complete 2017 District Sample		Sample of SI Users	
Variable	Category	N	%	N	%
Gender	White	269	4	154	4
	Total	6,350	100	3,573	100
	Female	3,016	47	1,701	48
	Male	3,334	53	1,872	52
	Total	6,350	100	3,573	100
Special Education	No	5,570	88	3,190	89
	Yes	780	12	383	11
	Total	6,350	100	3,573	100
English Language Learner	No	5,118	81	2,905	81
	Yes	1,232	19	668	19
	Total	6,350	100	3,573	100
Migrant Status	Yes	88	1	16	0
	No	6,262	99	3,557	100
	Total	6,350	100	3,573	100

Definition of Usage

To evaluate just how much the district is using Study Island, “usage” is defined as answering questions for a quiz or “session,” in which a student answers questions associated with a ten-item practice quiz available for each topic. Students who answer at least one item in one quiz are considered Study Island users (**SI Users**). All other students with no practice questions answered are considered non-users (**SI Non-Users**).

Patterns of Use

Table 4 provides the total number of unique students answering any questions in any session, compared to the total number of students enrolled in the district. The far-right column, “ELA or Math,” shows the number of students using Study Island for at least one subject. Overall, a large proportion of the district’s 3rd to 8th graders are using Study Island. The concentration of users is particularly strong in grades 5, 7, and 8 in math and ELA, with around two-thirds of students using Study Island in at least one subject. By subject, math is more commonly used, with 51% of district students participating, ranging from 36% of students in 4th grade to 58% in both 7th and 8th grade.

Table 4. Total Number and District Proportion of Students Using Study Island

		ELA		Math		ELA or Math	
Grade	District Total Enrollment*	SI User (N)	Percent of District (%)	SI User (N)	Percent of District (%)	SI User (N)	Percent of District (%)
3	1,123	520	46	516	46	570	51
4	1,065	389	37	384	36	424	40
5	1,117	630	56	639	57	699	63
6	1,014	446	44	501	49	561	55
7	1,011	487	48	586	58	649	64
8	1,020	438	43	594	58	645	63
Total	6,350	2,910	46	3,220	51	3,548	56

*Total district enrollment counts based on AzMERIT data received from district.

		ELA		Math		ELA or Math	
Grade	District Total Enrollment*	SI User (N)	Percent of District (%)	SI User (N)	Percent of District (%)	SI User (N)	Percent of District (%)
Equivalent data for Science in Appendix B							

Data from across the district suggest that Study Island may be a tool used in preparation for the end-of-year assessments. See [Appendix A](#), which shows high usage across the district nearer the date of the state assessment.

Students in the district also used Study Island in science, but this analysis focuses exclusively on math and ELA because the district did not provide the AIM Science scores. Study Island science usage data is comparable to the data shown in Tables 2, 5–8, and Figures 2 and 3. [Appendix B](#) provides further information about students' use of Study Island for science.

Analyses: Study Island

Research Question 1: How did students in the district use Study Island during the 2016–17 school year?

As discussed earlier, students are considered Study Island users when they answer at least one practice question during the school year. To gauge the amount of student usage, we consider several measures including the number of items attempted, the amount of time spent, the number of active weeks within the product, and the amount of time spent per active week. Although performance in practice sessions is not a measure of usage, we also report the overall performance of students in practice.

Table 5 shows descriptive information about the total number of items attempted and the total number of those answered correctly aggregated over the course of the 2016–17 school year. The proportion of items students answered correctly hovers around 50% across the board, ranging from an average of 39% in grade 7 math to 59% in grade 3 math. Table 6 provides descriptive data on the amount of time spent by grade and content area to show how much time Study Island users spent answering these questions. Sixth and seventh graders in math spend, on average, the least amount of time and answer the fewest items with a median of 90–100 minutes (Table 6) and 93 and 121 items answered on average (Table 5). Grade 5 math students and grade 6 ELA students spent the most amount of time overall: about 320 minutes, or over 5 hours, answering almost 380 items and 359 items, on average. Third graders also had strong usage, answering the most items overall, 437 and 501 on average in ELA and math, and spending 288 minutes and 316 minutes on average.

Table 5. Descriptive Statistics for Total Number Attempted and Proportion Correct, Study Island Items, 2016–17 School Year

			Number of Items Attempted				Proportion Correct					
Subject	Grade	N	Min	Med	Max	Mean	SD	Min	Med	Max	Mean	SD
ELA	3	520	1	181.00	8090	437.03	692.58	0.00	0.48	1.00	0.47	0.17
	4	389	1	121.00	2681	226.92	309.80	0.00	0.43	1.00	0.43	0.19
	5	630	1	116.00	3886	238.10	346.50	0.00	0.54	1.00	0.53	0.17
	6	446	1	253.50	3938	359.06	401.35	0.00	0.52	1.00	0.51	0.18
	7	487	1	94.00	1105	126.04	131.99	0.00	0.50	1.00	0.48	0.19
	8	438	1	83.00	2304	132.21	180.40	0.00	0.49	1.00	0.48	0.19
	Total	2,910	1	126.00	8090	256.00	411.81	0.00	0.50	1.00	0.49	0.18
Math	3	516	3	325.00	5701	500.61	579.15	0.15	0.62	1.00	0.59	0.16
	4	384	1	221.00	3322	318.17	353.57	0.00	0.60	0.98	0.58	0.17

Subject	Grade	N	Number of Items Attempted				Proportion Correct					
			Min	Med	Max	Mean	SD	Min	Med	Max	Mean	SD
	5	639	1	224.00	4525	380.39	466.58	0.00	0.53	1.00	0.53	0.17
	6	501	1	99.00	1072	153.51	166.18	0.00	0.50	1.00	0.49	0.20
	7	586	1	56.00	1391	93.09	121.41	0.00	0.38	0.92	0.39	0.17
	8	594	1	71.50	1385	120.59	146.87	0.00	0.49	1.00	0.48	0.19
	Total	3,220	1	119.00	5701	256.73	381.57	0.00	0.52	1.00	0.50	0.19

Equivalent data for Science in Appendix B

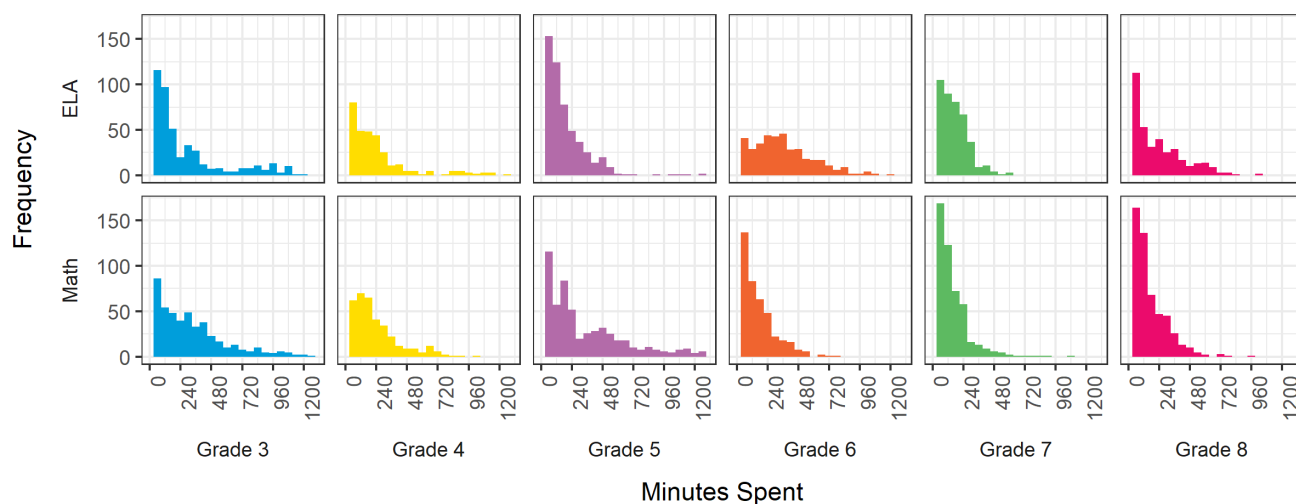
Table 6. Descriptive Statistics for Amount of Time (Minutes), Study Island Users

Subject	Grade	N	Min	Median	Max	Mean	SD
ELA	3	520	0.10	132.46	2,111.25	288.20	363.99
	4	389	0.25	133.63	1,340.13	207.57	252.69
	5	630	1.40	130.12	2,075.87	228.72	342.53
	6	446	0.17	282.92	1,355.07	318.46	250.57
	7	487	0.05	130.90	589.55	146.96	114.10
	8	438	0.58	127.56	1,855.22	202.18	212.96
	Total	2,910	0.05	144.72	2,111.25	232.60	281.52
Math	3	516	2.32	243.81	1,846.87	316.95	303.16
	4	384	1.93	179.62	1,008.75	222.88	183.45
	5	639	0.22	196.15	1,670.32	321.43	332.70
	6	501	0.05	98.62	756.92	141.03	133.56
	7	586	0.15	94.78	1,064.83	131.17	130.68
	8	594	0.25	112.63	939.87	153.79	135.00
	Total	3,220	0.05	138.94	1,846.87	215.34	237.95

Equivalent data for Science in Appendix B

Figure 2 shows how counts of students are distributed across the amount of time spent in Study Island. For example, there are many student users in grade 7, but they are spending less time using Study Island than the fewer users spending more time in grade 4.

Figure 2. Distribution of Time in Minutes by Grade and Content Area for Study Island Users in the 2016–17 School Year.*



* Equivalent figure for Science in Appendix B

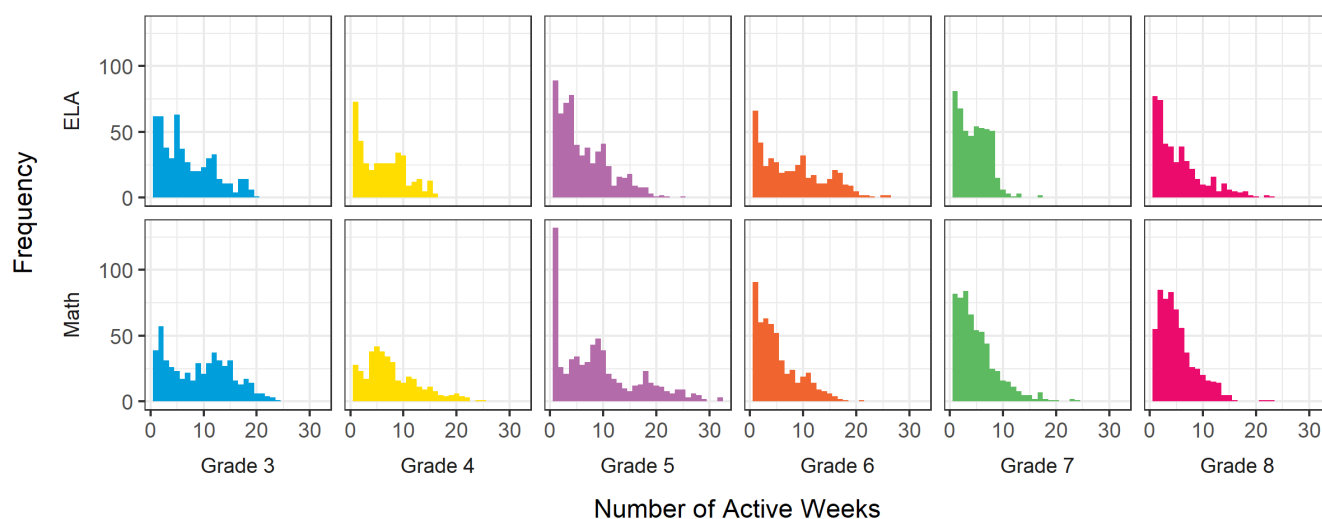
Such time durations are not likely to occur all at once. To get a sense of the dispersion of time in use across weeks, Table 7 shows the total number of weeks with any use, or “active weeks.” On average, the greatest number of weeks with usage are in math in grade 3 and grade 5, at about nine weeks. In ELA, 6th graders have the highest average number of active weeks, totaling eight weeks. These data show that the higher 5th grade math usage and 6th grade ELA usage are spread across more weeks during the school year as compared to other grades. Figure 3 shows the distribution of active weeks for each grade and subject. It shows that, generally, grades 3 and 5 have more students in both ELA and math with more active weeks. These views help to illustrate how the Study Island items are used across grades and subject areas. They help to show that, for example, there are many grade 8 students using Study Island, but not across as many active weeks as grade 5.

Table 7. Descriptive Statistics for Active Weeks Using Study Island

Subject	Grade	N	Min	Median	Max	Mean	SD
ELA	3	520	1	6	20	6.98	4.96
	4	389	1	6	16	6.07	4.19
	5	630	1	5	25	6.33	4.75
	6	446	1	7	26	8.15	5.99
	7	487	1	4	17	4.58	2.80
	8	438	1	4	23	5.56	4.60
	Total	2,910	1	5	26	6.28	4.77
Math	3	516	1	9	24	9.31	5.89
	4	384	1	7	25	7.87	5.12
	5	639	1	8	32	9.11	7.39
	6	501	1	4	21	5.20	3.95
	7	586	1	4	24	5.22	3.95
	8	594	1	4	23	5.28	3.54
	Total	3,220	1	5	32	6.97	5.52

Equivalent data for Science in Appendix B

Figure 3. Distribution of Active Weeks by Grade and Content area for Study Island Users in the 2016–17 School Year. *



* Equivalent figure for science in Appendix B

Finally, to see just how much of the time occurs within each active week, Table 8 provides the amount of time per week as a result of calculating the total time spent in Study Island divided by the number of active weeks. The average amount of time per active week ranges from about 24 minutes in grade 7 math to 40 minutes in grade 6 ELA. Fifth grade has the highest average in math with 34 minutes per active week.

Table 8. Descriptive Statistics for Time per Active Week (Minutes), Study Island Users

Subject	Grade	N	Min	Median	Max	Mean	SD
ELA	3	520	0.10	26.35	182.29	35.60	28.61
	4	389	0.25	23.51	144.52	28.38	19.65
	5	630	0.76	25.39	230.65	33.34	34.51
	6	446	0.17	35.83	159.76	39.93	24.85
	7	487	0.05	29.33	121.68	30.74	17.45
	8	438	0.58	29.08	105.76	33.29	20.75
	Total	2,910	0.05	28.07	230.65	33.65	26.02
Math	3	516	2.32	29.00	127.33	31.41	17.17
	4	384	1.58	25.10	89.25	27.68	14.31
	5	639	0.22	28.40	185.59	33.72	24.84
	6	501	0.05	22.76	138.58	26.58	17.82
	7	586	0.15	22.75	117.20	24.49	14.20
	8	594	0.25	25.29	114.14	28.41	16.76
	Total	3,220	0.05	25.85	185.59	28.86	18.49

Equivalent data for Science in Appendix B

AzMERIT Performance and Study Island Use

Research Question 2: Are there significant mean differences in the AzMERIT state test scores between students who use Study Island and those who do not?

In order to contextualize the district's AzMERIT performance, we begin with a descriptive look at mean AzMERIT scores. Table 9 compares performance on the 2017 AzMERIT in terms of scale scores for each of the Study Island user and Study Island non-user groups, compared to the district overall by

grade and content area. Because the Arizona Department of Education did not release mean state AzMERIT scale scores for the 2017 testing period, we also include Table 10 to compare the district performance on the 2016 AzMERIT to the state as a whole. Table 10 shows that in 2016, the district had lower mean scale scores compared to the state, and comparing Table 9 and Table 10 shows that the district mean scale scores are very similar in 2016 and 2017. Table 9 shows that Study Island users had higher AzMERIT scores than both the district as a whole and the Study Island non-user group. Specifically, Study Island users outperformed both the district and the Study Island non-users in all content areas and grades except for grade 8 ELA. Mean scores between users and non-users differed as much as 11 points in grade 3 ELA and 20 points in grade 3 and 5 math.

Table 9. Descriptive 2017 AzMERIT Scale Scores of Study Island Users, Study Island Non-Users, and District

Subject	Grade	Study Island User			Study Island Non-User			District		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
ELA	3	520	2492.13	29.93	604	2481.11	28.18	1,124	2486.26	29.52
	4	389	2507.95	27.93	677	2501.41	27.47	1,066	2503.81	27.81
	5	630	2521.84	29.27	488	2515.81	28.49	1,118	2519.22	29.07
	6	447	2530.91	31.03	573	2524.69	29.50	1,020	2527.43	30.33
	7	487	2539.97	28.66	524	2536.79	30.75	1,011	2538.33	29.78
	8	438	2536.78	29.99	581	2538.50	30.33	1,019	2537.75	30.18
Math	3	513	3516.55	43.44	604	3496.86	46.59	1,117	3505.94	46.20
	4	383	3542.01	42.13	682	3524.22	42.66	1,065	3530.63	43.30
	5	638	3578.46	42.62	478	3558.54	44.00	1,116	3569.95	44.30
	6	501	3602.39	44.45	522	3588.19	38.11	1,023	3595.14	41.92
	7	586	3614.85	41.28	426	3600.16	39.59	1,012	3608.70	41.20
	8	593	3645.54	37.27	425	3628.33	30.13	1,018	3638.37	35.49

Table 10. Descriptive 2016 AzMERIT Scale Scores of District and State

Subject	Grade	District			State*		
		N	Mean	SD	N	Mean	SD
ELA	3	1,038	2486.00	28.30	87,793	2501.38	31.81
	4	1,056	2497.64	28.79	86,325	2517.62	33.89
	5	1,051	2517.99	32.44	85,425	2536.70	34.41
	6	1,022	2521.95	27.97	84,651	2540.77	33.74
	7	989	2536.83	28.33	84,138	2551.77	31.12
	8	1,007	2540.42	28.15	82,779	2555.26	32.38
Math	3	1,053	3506.68	43.49	88,303	3524.19	44.86
	4	1,076	3531.67	37.37	86,711	3552.27	40.62
	5	1,059	3569.54	41.77	85,719	3587.75	41.57
	6	1,029	3594.71	39.52	84,675	3615.26	42.60
	7	1,001	3614.52	30.02	81,829	3632.28	35.61
	8	1,012	3640.34	35.09	69,858	3651.11	36.11

* State performance data from American Institutes for Research (2017)

To discern whether these differences shown in Table 9 between users and non-users are significant, we must take into account the differences in student ability across the user groups. If students using Study Island are generally higher-ability students, whether or not they are users may be meaningless

with regard to their performance on the AzMERIT. To understand the association between using Study Island and AzMERIT performance, only students with similar AzMERIT scores in 2016 should be compared across user groups. Holding their ability constant based on a prior score supports meaningful comparisons across the two groups.

A nearest neighbor propensity score matching (PSM) (Rosenbaum & Rubin, 1983) was conducted to align students in the user group to the students in the non-user group by ability (as measured by students' 2016 AzMERIT scores) so that statistical analyses of the 2017 AzMERIT mean score differences can be conducted. Although not causally conclusive, any discernable differences may reflect a difference in the impact of use rather than an inherent difference in ability from the start.

Only ELA and math grades 4–8 could be included in the analysis because third graders do not have a prior-year AzMERIT score. Some other users within these grades were also eliminated from the sample because they did not have an AzMERIT 2016 score. The size of the resulting matched sample depended on whether the user group or non-user group was smaller, with the total number of cases able to be matched determined by the group with the smaller size. The total resulting N is shown in both Tables 11 and 12. (Please see [Appendix C](#) for figures that show the spread of scores across Study Island users [True] and Study Island non-users [False] and the resulting PSM.)

Table 11 reports the results from a *t*-test comparing the mean 2016 AzMERIT scores of Study Island users to scores from Study Island non-users in order to determine whether equivalent matched groups were possible with nearest neighbor propensity score matching. Findings reported there show that the average 2016 AzMERIT scale scores for the matched samples were significantly different from each other for both grade 6 and grade 8 math, so any results for those groups should be treated cautiously. These differences suggest that in general, higher-ability students in those grades in the district were more likely to use Study Island math and that a group of similar ability non-users was not available for those grades.

Table 11. *t*-Test Comparisons of Propensity Score (2016 AzMERIT Score) Between Matched Study Island Users and Non-Users

		SI User		SI Non-User		Matched	AzMERIT 2016				
Subject	Grade	Mean	SD	Mean	SD	N	Mean Difference	95% CI		t	df
ELA	4	2490.75	27.44	2490.36	26.87	314	0.39	-4.65	3.86	-0.181	625.73
	5	2498.41	28.13	2496.97	28.00	350	1.44	-5.60	2.73	-0.677	697.99
	6	2523.08	32.91	2520.38	30.39	380	2.71	-7.22	1.80	-1.178	753.23
	7	2523.57	26.77	2523.17	28.19	402	0.40	-4.20	3.41	-0.205	799.85
	8	2538.45	26.24	2538.39	26.24	355	0.06	-3.92	3.81	-0.029	708.00
Math	4	3512.16	42.40	3511.77	41.88	315	0.40	-6.99	6.20	-0.118	627.91
	5	3535.07	37.32	3532.14	37.77	357	2.93	-8.45	2.58	-1.044	711.90
	6	3575.10	40.01	3564.26	38.42	404	10.84	-16.26	-5.43	-3.929***	804.68
	7	3597.57	40.20	3593.02	39.63	314	4.55	-10.81	1.70	-1.429	625.87
	8	3616.76	27.16	3610.41	28.13	329	6.34	-10.58	-2.11	-2.943**	655.19

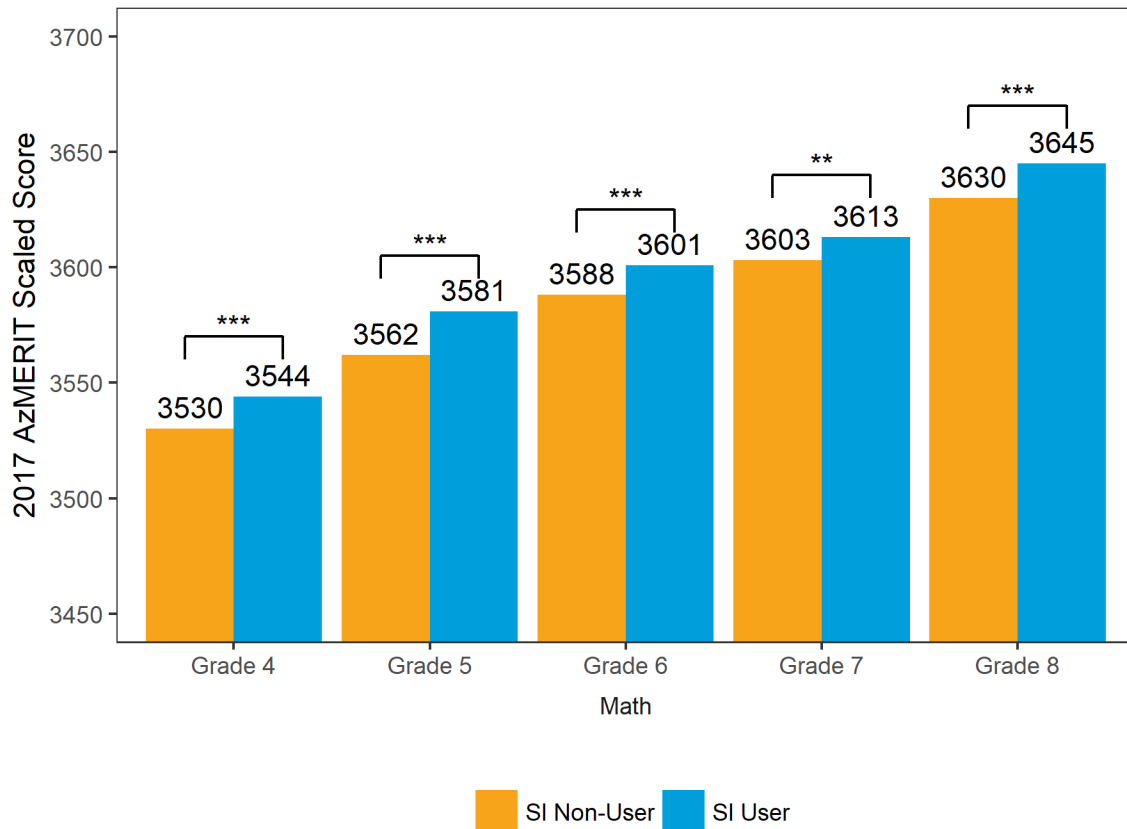
A *t*-test was conducted to compare the 2017 AzMERIT scores across the matched Study Island user and non-user groups. Results from the analysis are shown in Table 12, where *N* reports the equal size of the matched groups. While the mean AzMERIT scale score for the Study Island user group is larger than for the matched non-user group in every category except for grade 8 ELA, all the mean scale score differences for math but none for ELA are statistically significant once the groups are matched based on 2016 AzMERIT score. Recall that user groups in grade 6 and grade 8 math were not of equivalent ability, so the differences found in the 2017 AzMERIT cannot be confidently attributed to Study Island usage but could be associated with student ability. Table 12 includes a column that shows

the mean difference in 2017 AzMERIT scores between users and non-users. For example, in grade 5 math, the mean AzMERIT scores for the Study Island users is 20 points higher than for the matched sample of non-users, a statistically significant difference. Figures 4 and 5 show a visual representation of the mean scale scores by user group once they are matched based on 2016 ability.

Table 12. *t*-Test Comparisons of AzMERIT Scale Score Between Matched Study Island Users and Non-Users

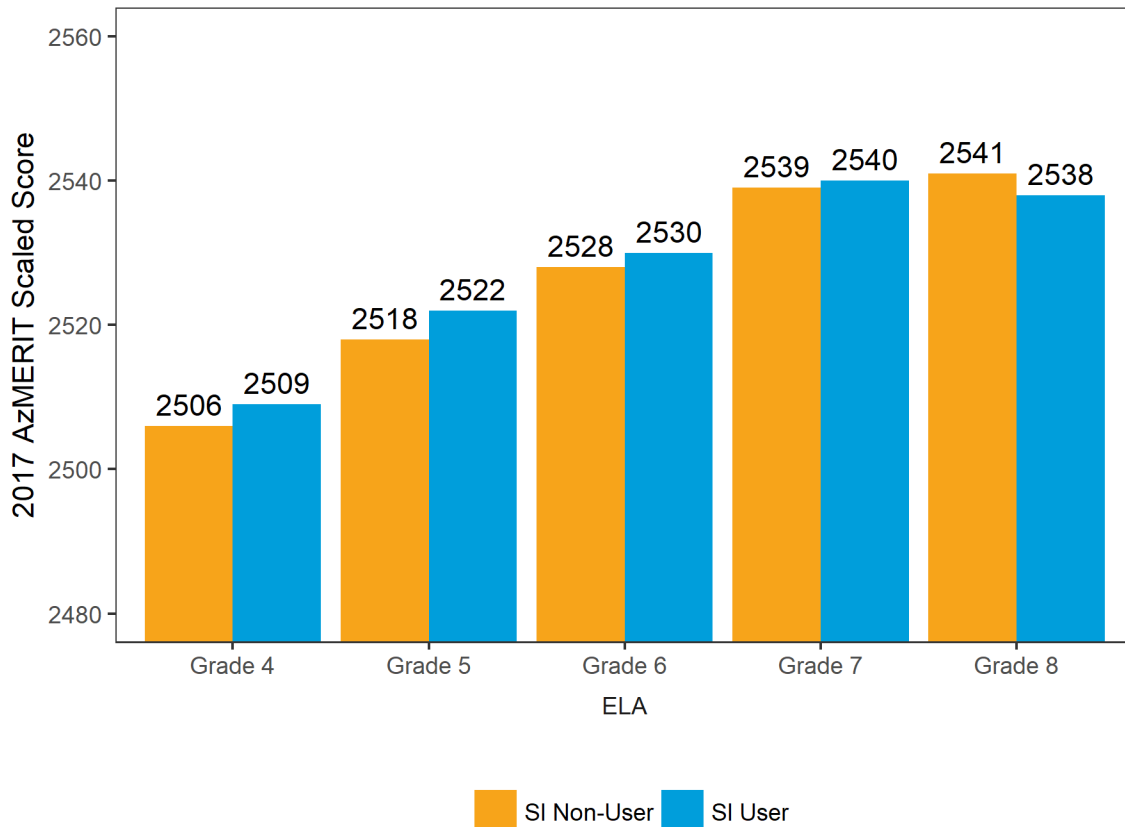
		SI User		SI Non-User		Matched	AzMERIT 2017				
Subject	Grade	Mean	SD	Mean	SD	N	Mean Difference	95% CI		t	df
ELA	4	2509.78	27.99	2506.31	27.80	314	3.47	-7.84	0.90	-1.559	625.97
	5	2522.01	27.62	2518.19	28.58	350	3.82	-7.99	0.35	-1.798	697.19
	6	2530.38	30.78	2528.10	29.49	380	2.28	-6.57	2.01	-1.042	756.62
	7	2540.26	28.19	2539.44	30.02	402	0.81	-4.84	3.22	-0.396	798.83
	8	2538.24	29.87	2541.07	30.01	355	-2.83	-1.58	7.24	1.258	707.98
Math	4	3544.70	43.30	3530.93	43.11	315	13.77	-20.53	-7.01	-4.000***	627.99
	5	3581.95	43.15	3562.13	44.11	357	19.82	-26.23	-13.41	-6.068***	711.66
	6	3601.72	42.36	3588.44	37.61	404	13.28	-18.81	-7.74	-4.711***	794.88
	7	3613.92	42.04	3603.99	39.90	314	9.93	-16.35	-3.51	-3.036**	624.30
	8	3645.92	35.01	3630.05	30.56	329	15.87	-20.90	-10.84	-6.195***	644.22

Figure 4. Mean 2017 AzMERIT Scores by Grade and Usage Group, Math



*/**/*** - Statistically significant difference

Figure 5. Mean 2017 AzMERIT Scores by Grade and Usage Group, ELA



*/**/*** - Statistically significant difference

Research Question 3: Is there a significant relationship between AzMERIT proficiency-level categorization and Study Island usage?

Because the proficiency level is a key variable in accountability, Table 13 provides descriptive data of the number and percentage of students performing in the top two proficiency categories as “overall proficiency” across the unmatched user groups, the district, and the state. (For total N counts by group on which the percentages are based, see Table 11.) The district has smaller proportions of students achieving at least proficient on the AzMERIT compared to the state as a whole, ranging from 15% in grade 8 math to 32% in grade 5 math, and in ELA, 16% in grade 8 to 25% in grade 4. Overall proficiency categorizations show similar results: in most groups, Study Island users tend to have higher proficiency percentages than non-users, with the district percentage falling between the percentages for the two user groups.

Table 13. Percentage of Students in Grades 3–8 Scoring Proficient or Highly Proficient on the 2017 AzMERIT District vs State

Subject	Grade	SI User		SI Non-User		District		State	
		N	%	N	%	N	%	N	%
ELA	3	157	30	112	19	269	24	38,929	44
	4	117	30	147	22	264	25	43,543	49
	5	156	25	104	21	260	23	38,666	44
	6	112	25	106	18	218	21	36,251	42

Subject	Grade	SI User		SI Non-User		District		State	
		N	%	N	%	N	%	N	%
Math	7	124	25	123	23	247	24	37,690	44
	8	68	16	93	16	161	16	29,027	34
	3	194	38	148	25	342	31	41,615	47
	4	128	33	124	18	252	24	41,869	47
	5	246	39	115	24	361	32	41,329	47
	6	158	32	80	15	238	23	36,321	42
	7	110	19	44	10	154	15	28,441	34
	8	122	21	31	7	153	15	20,970	29

By using the same matched groups from the PSM that were used earlier to explore mean score difference (see Table 11 for cell counts), a chi-square test was run on the frequencies of students within each category to discern the relationship in proficiency-level categorization between Study Island users and non-users, findings for which are reported in Table 14. For ease of interpretation, Table 14 reports the percentage of students within each proficiency category for both the user and non-user groups rather than the frequencies within each group upon which the chi-square test is based. All groups in math except for grade 7 show a statistically significant relationship between Study Island user status and student proficiency-level categorization, all of which also have statistically different mean AzMERIT scale scores as shown earlier. These Study Island user groups have smaller proportions of students scoring at the “Below Basic” proficiency level and larger proportion scoring at the “Proficient” and “Highly Proficient” levels. As before, the results in grades 6 and 8 should be interpreted cautiously, since it was not possible to create matched groups of equivalent ability.

Table 14. Chi-Square Test Comparison of 2017 AzMERIT Proficiency Level Categorization Between Matched Study Island Users and Non-Users

Grade	Performance Level	ELA			Math		
		User (%)	Non-User (%)	Chi-Sq.	User (%)	Non-User (%)	Chi-Sq.
4	Minimally Proficient	53	59	2.545	34	44	15.695**
	Partially Proficient	14	14		30	34	
	Proficient	28	24		29	17	
	Highly Proficient	5	4		7	4	
5	Minimally Proficient	48	55	6.713	33	49	24.974***
	Partially Proficient	27	20		26	23	
	Proficient	23	22		29	23	
	Highly Proficient	2	3		12	5	
6	Minimally Proficient	51	52	2.512	51	63	31.801***
	Partially Proficient	23	27		18	22	
	Proficient	25	20		23	13	
	Highly Proficient	1	1		8	2	
7	Minimally Proficient	53	52	0.082	65	75	6.695
	Partially Proficient	21	21		16	13	
	Proficient	24	24		11	8	
	Highly Proficient	2	2		7	5	
8	Minimally Proficient	65	61	2.094	59	72	20.192***

	Partially Proficient	19	21		21	19	
	Proficient	15	15		12	6	
	Highly Proficient	1	3		8	2	

Conclusions

The findings in this study suggest there are discernable and statistically significant positive impacts on AzMERIT scores for students participating in Study Island. Generally, implementation and use of Study Island in the Arizona school district vary by grade and content area, with grade 3 math and ELA, grade 5 math, and grade 6 ELA exhibiting the strongest usage. Some groups of students appear to be answering relatively few questions and spending minimal time over the course of the year, while other groups have a stronger implementation. In math, where students spend more time, answer more questions, and spread their time over active weeks, positive differences are observed. This is particularly evident in grades 4 and 5. Similar results for grades with strong usage are not found in ELA.

These analyses are clearly impacted by the quality and approach by which schools use Study Island. It would be an important next step to understand the qualitative differences in implementation approaches, such as for grade 5 students. Understanding the methods will help guide implementations that drive evidence-based, positive outcomes for students.

References

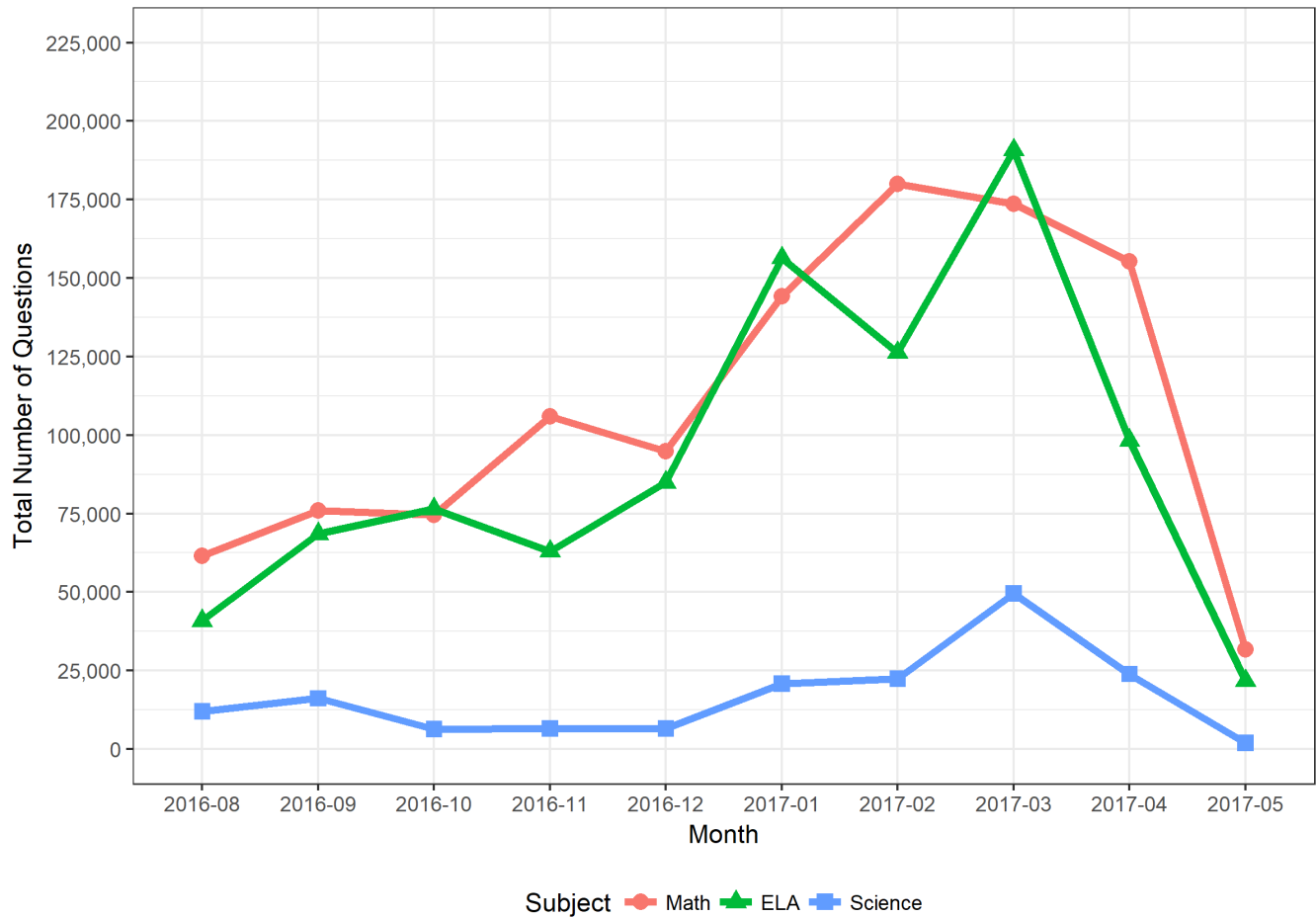
- Adams, R. H., & Strickland, J. (2012). The effects of computer-assisted feedback strategies in technology education: A comparison of learning outcomes. *Journal of Educational Technology Systems, 40*, 211–223.
- American Institutes for Research. (2017, February 24). *Annual technical report: Arizona statewide assessment in English language arts and math—2015-2016 school year*. Retrieved from <https://cms.azed.gov/home/GetDocumentFile?id=59c2e9ac3217e10dac1f2ffd>
- Arizona Department of Education. (2016a). *ELA item specifications*. Retrieved from <https://cms.azed.gov/home/GetDocumentFile?id=583deba7aadebf0a185fcf21>
- Arizona Department of Education. (2016b). *Math item specifications*. Retrieved from <https://cms.azed.gov/home/GetDocumentFile?id=583ca693aadebe13d87d4371>
- Baker, S., Gersten, R., & Lee, D. S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal, 103*, 51–73.
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research, 85*, 89–99.
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213–238.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 81*(2), 139–148.

- Camp, G., Paas, F., Rikers, R., & van Merriënboer, J. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior*, 17, 575–595.
- Chase, J. A., & Houmanfar, R. (2009). The differential effects of elaborate feedback and basic feedback on student performance in a modified, personalized system of instruction course. *Journal of Behavioral Education*, 18, 245–265.
- Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2006). Towards a personalized task selection model with shared instructional control. *Instructional Science*, 34, 399–422.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219–232.
- Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *The Psychological Record*, 54, 207–231.
- Duke, N. K., & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 205–243). Newark, DE: International Reading Association.
- Dun and Bradstreet. (2018). *MDR educational data* [Data file and code book]. License purchased by Edmentum.
- Every Student Succeeds Act of 2015. 20 U.S.C. §1111 Retrieved from <https://www.congress.gov/bill/114th-congress/senate-bill/1177/text>
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53, 199–208.
- Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review*, 28(4), 659–671.
- Gresham, F., Reschly, D., & Shinn, M. R. (2010). RTI as a driving force in educational improvement: Historical legal, research, and practice perspectives. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for academic achievement problems in a three-tier model, including RTI* (pp. 47–77). Bethesda, MD: National Association of School Psychologists.
- Heritage, M. (2010). *Formative assessment and next-generation assessment systems: Are we losing an opportunity?* Paper prepared for the Council of Chief State School Officers. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- January, S. A., Van Norman, E. R., Christ, T. J., Ardoin, S. P., Eckert, T. L., & White, M. J. (2018). Progress monitoring in reading: Comparison of weekly, bimonthly, and monthly assessments for students at risk for reading difficulties in grades 2–4. *School Psychology Review*, 47, 83–94.
- Jenkins, D. (2001). Impact of the implementation of the teaching/learning cycle on teacher decision making and emergent readers. *Reading Psychology*, 22, 267–288.
- Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to optimize the efficiency of e-learning. *Educational Technology, Research and Development*, 53(3), 83–93.
- Kilpatrick, J., Swafford, J., & Bradford, R. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1, 279–308.
- Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58, 79–97.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership*, 63(3), 19–24.
- Lee, H. W., Lim, K. Y., & Grabowski, B. L. (2010). Improving self-regulation, learning strategy use, and achievement with metacognitive feedback. *Education Technology Research and Development*, 58, 629–648.
- Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- National Center for Education Statistics. (n.d.). *Common core of data*. Retrieved from <https://nces.ed.gov/ccd/>
- Organisation for Economic Co-operation and Development. (2012). *Equity and quality in education: Supporting disadvantaged students and schools*. Retrieved from <http://dx.doi.org/10.1787/9789264130852-en>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.
- Salden, R. J. C. M., Paas, F., Broers, N. J., & van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional Science*, 32, 153–172.
- Stecker, P. M., Lembke, E. S., & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision making. *Preventing School Failure*, 52, 48–58.
- Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan*, 81(3), 191–198.
- Van Norman, E. R., Nelson, P. M., & Parker, D. C. (2016). Technical adequacy of growth estimates from a computer adaptive test: Implications for progress monitoring. *School Psychology Quarterly*, 22, 379–391.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21, 361–375.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education*, 11, 49–65.
- Wolf, P. J. (2007). Academic improvement through regular assessment. *Peabody Journal of Education*, 82, 690–702.

Appendix A: Study Island Practice Questions Answered by Month (Grades K–12), 2016–17 School Year

Figure A1: Total Number of Practice Questions by Month



Appendix B: Study Island Usage in Science

Table B1. Total Number and District Proportion of Students Using Study Island

Grade	District Total Enrollment*	Science	
		SI User (N)	Percent of District (%)
3	1,123	146	13
4	1,065	251	24
5	1,117	150	13
6	1,014	155	15
7	1,011	263	26
8	1,020	363	36
Total	6,350	1,328	21

**Total district enrollment counts based on AzMERIT data received from district.*

Table B2. Descriptive Statistics for Total Number Attempted and Proportion Correct, Study Island Items

			Number of Items Attempted					Proportion Correct				
Subject	Grade	N	Min	Med	Max	Mean	SD	Min	Med	Max	Mean	SD
Science	3	146	1	13.50	369	26.64	44.30	0.00	0.32	1.00	0.34	0.23
	4	251	1	81.00	1764	180.68	262.41	0.00	0.48	1.00	0.48	0.19
	5	150	1	39.50	1619	73.87	160.69	0.00	0.46	1.00	0.45	0.19
	6	155	1	25.00	528	44.39	74.16	0.00	0.35	1.00	0.37	0.18
	7	263	1	47.00	646	67.13	73.77	0.00	0.48	1.00	0.47	0.20
	8	363	1	114.00	1294	164.63	191.17	0.00	0.55	1.00	0.53	0.19
	Total	1,328	1	50.00	1764	108.90	176.94	0.00	0.47	1.00	0.46	0.21

Table B3. Descriptive Statistics for Amount of Time (Minutes), Study Island Users

Subject	Grade	N	Min	Median	Max	Mean	SD
Science	3	146	0.05	6.84	118.08	13.09	17.36
	4	251	0.03	76.93	594.43	111.45	118.19
	5	150	0.05	29.52	469.82	43.15	55.66
	6	155	0.10	17.33	328.37	33.15	44.62
	7	263	0.08	48.08	228.65	57.89	48.02
	8	363	0.23	142.23	1,121.08	175.98	177.97
	Total	1,328	0.03	43.96	1,121.08	90.81	126.21

Figure B1. Distribution of Time in Minutes by Grade for Study Island Users in the 2016–17 School Year

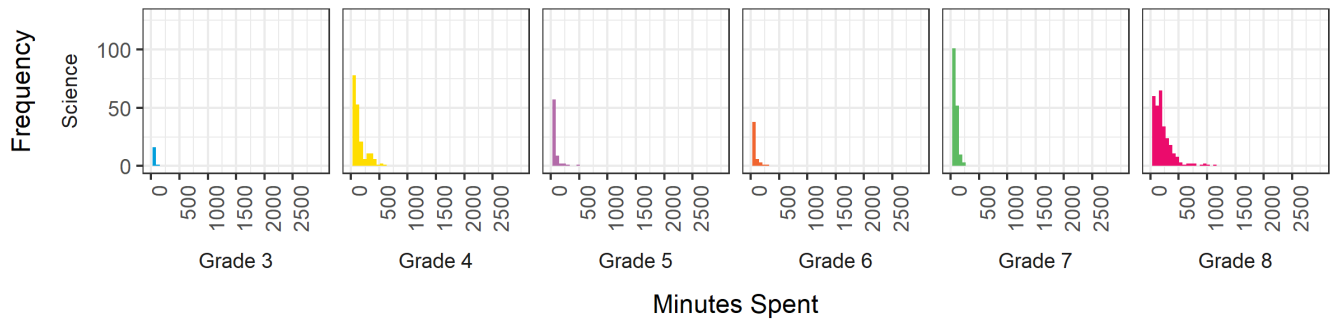


Table B4. Descriptive Statistics for Active Weeks Using Study Island

Subject	Grade	N	Min	Median	Max	Mean	SD
Science	3	146	1	1	11	1.85	1.58
	4	251	1	3	8	2.90	1.61
	5	150	1	2	15	2.83	2.07
	6	155	1	2	13	2.23	1.64
	7	263	1	2	10	2.70	1.91
	8	363	1	5	14	4.96	3.13
	Total	1,328	1	2	15	3.22	2.49

Figure B2. Distribution of Active Weeks by Grade for Study Island Users in the 2016–17 School Year

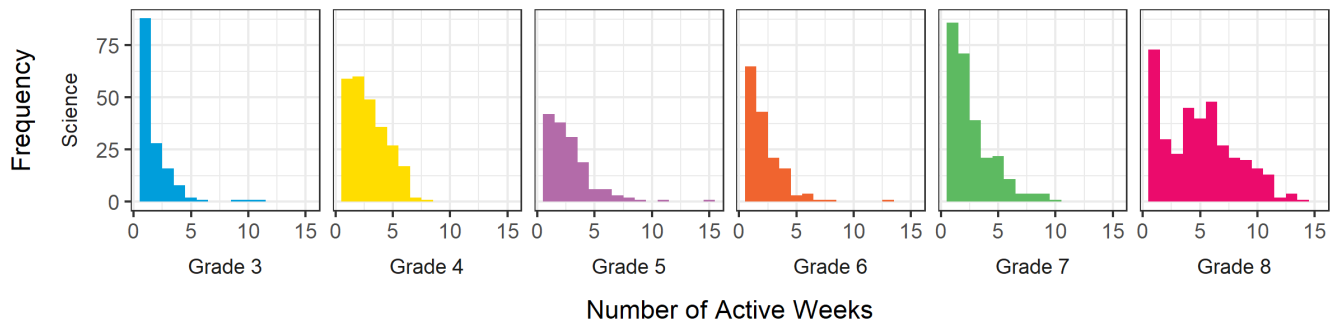


Table B5. Descriptive Statistics for Time per Active Week (Minutes), Study Island Users

Subject	Grade	N	Min	Median	Max	Mean	SD
Science	3	146	0.05	5.50	56.03	6.68	6.79
	4	251	0.03	27.78	174.41	34.04	29.36
	5	150	0.05	11.74	81.12	13.94	11.65
	6	155	0.10	10.83	82.15	12.96	12.03
	7	263	0.08	18.93	228.65	21.81	19.54
	8	363	0.23	28.96	166.38	30.74	20.02
	Total	1,328	0.03	17.76	228.65	22.97	21.78

Appendix C: Propensity Score Matching

Figure C1. ELA Grade 4 User vs Non-User

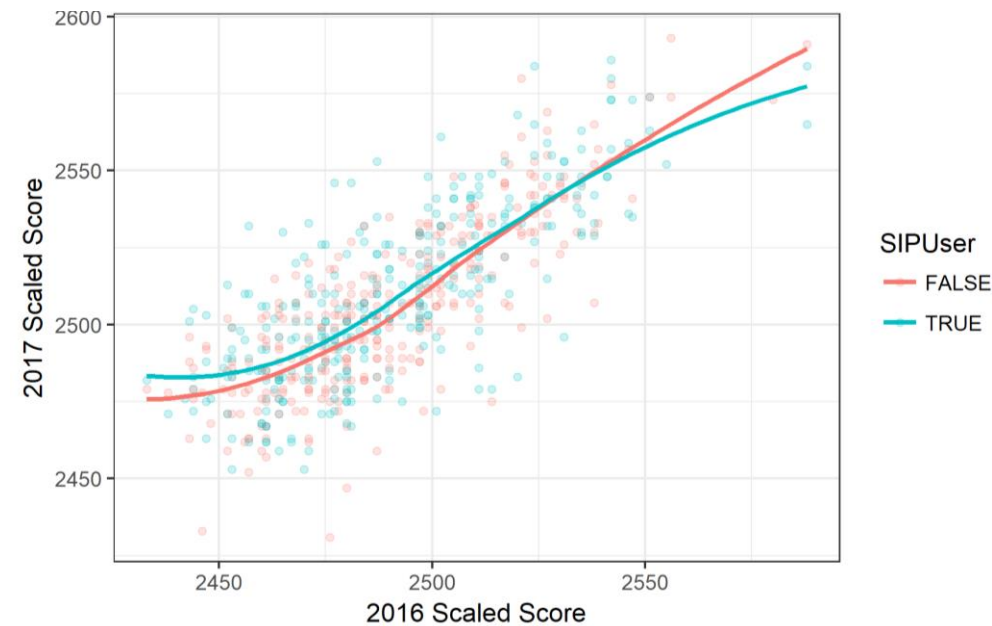


Figure C2. ELA Grade 5 User vs Non-User

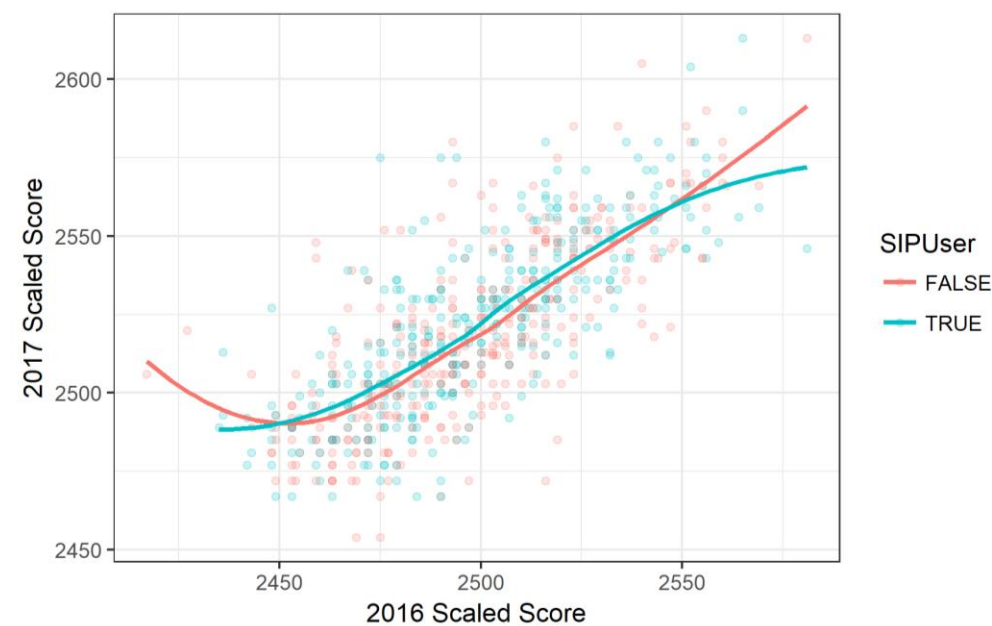


Figure C3. ELA Grade 6 User vs Non-User

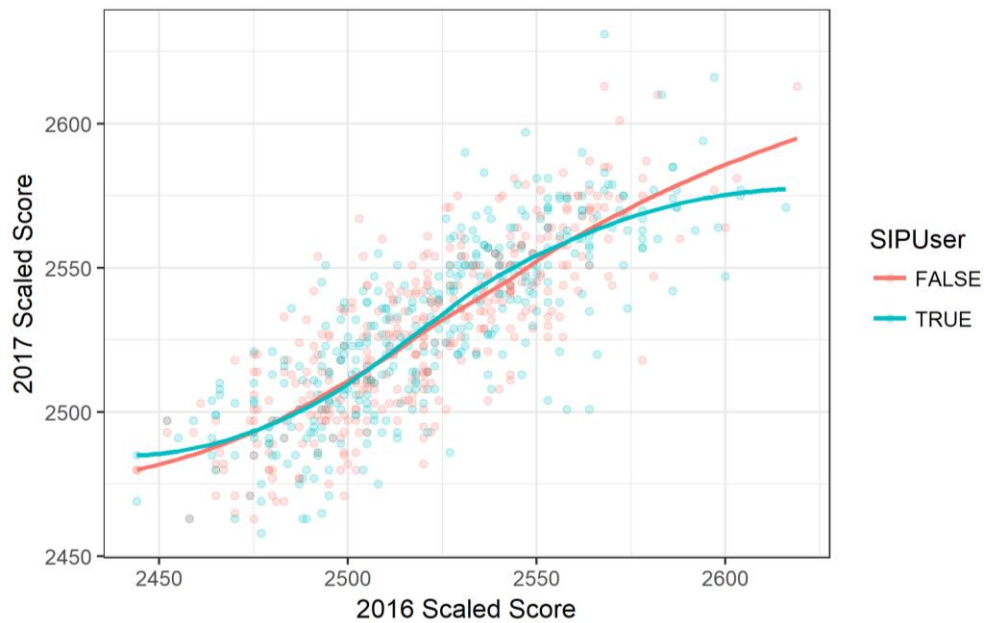


Figure C4. ELA Grade 7 User vs Non-User



Figure C5. ELA Grade 8 User vs Non-User

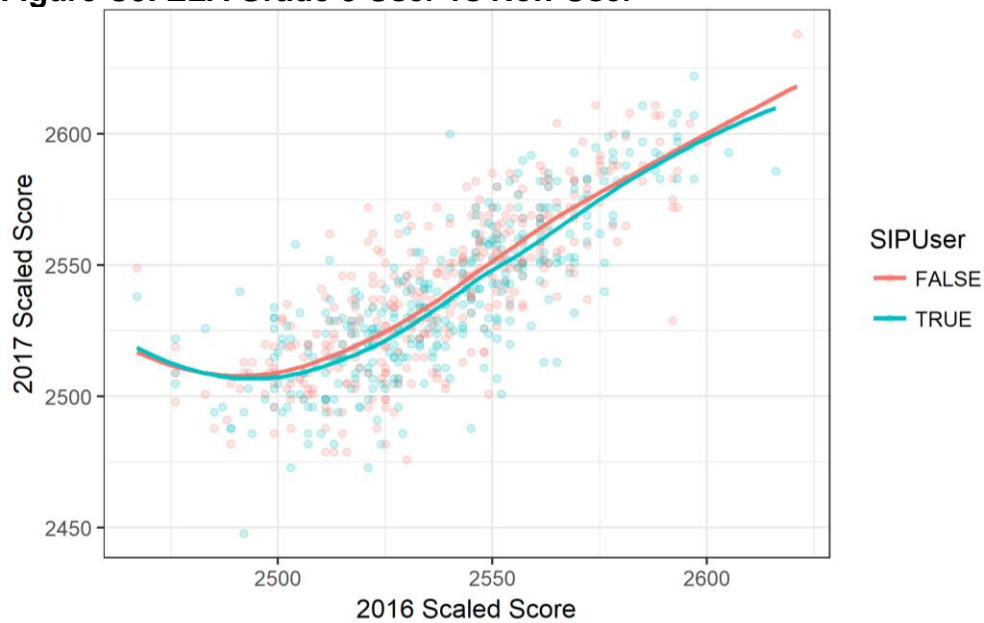


Figure C6. Math Grade 4 User vs Non-User

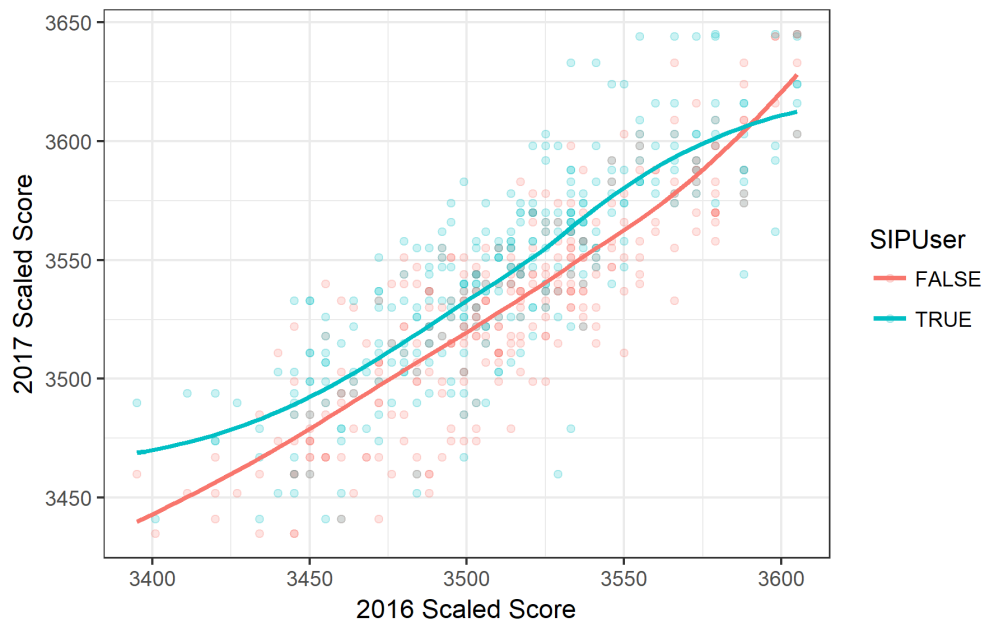


Figure C7. Math Grade 5 User vs Non-User

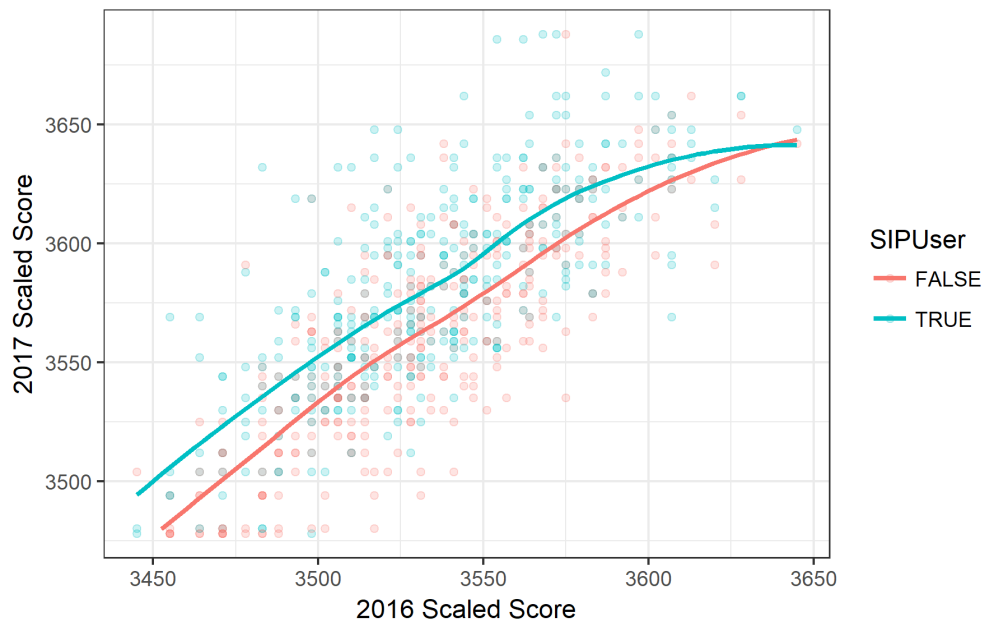


Figure C8. Math Grade 6 User vs Non-User

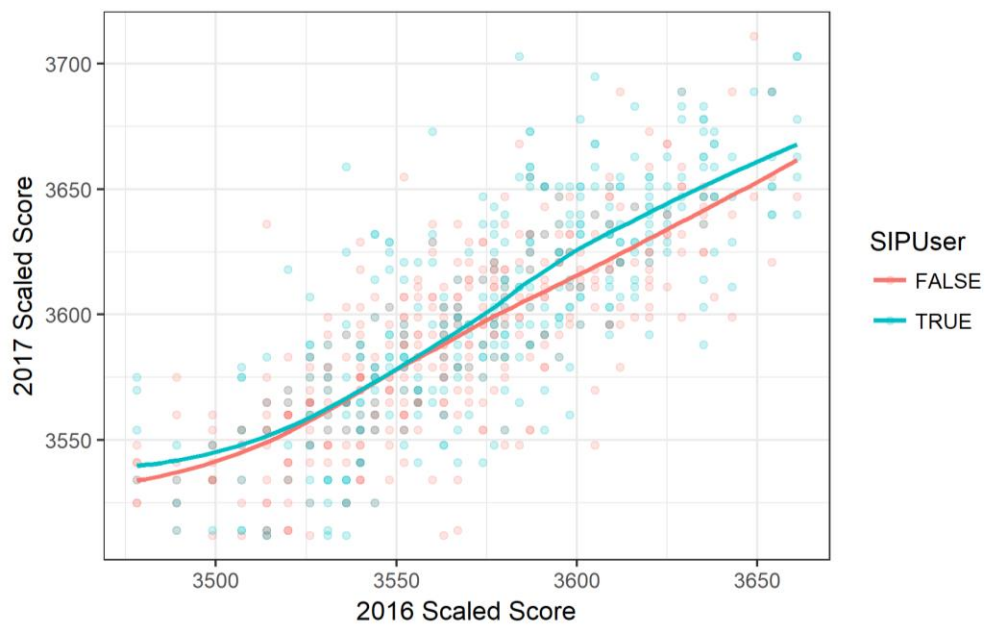


Figure C9. Math Grade 7 User vs Non-User

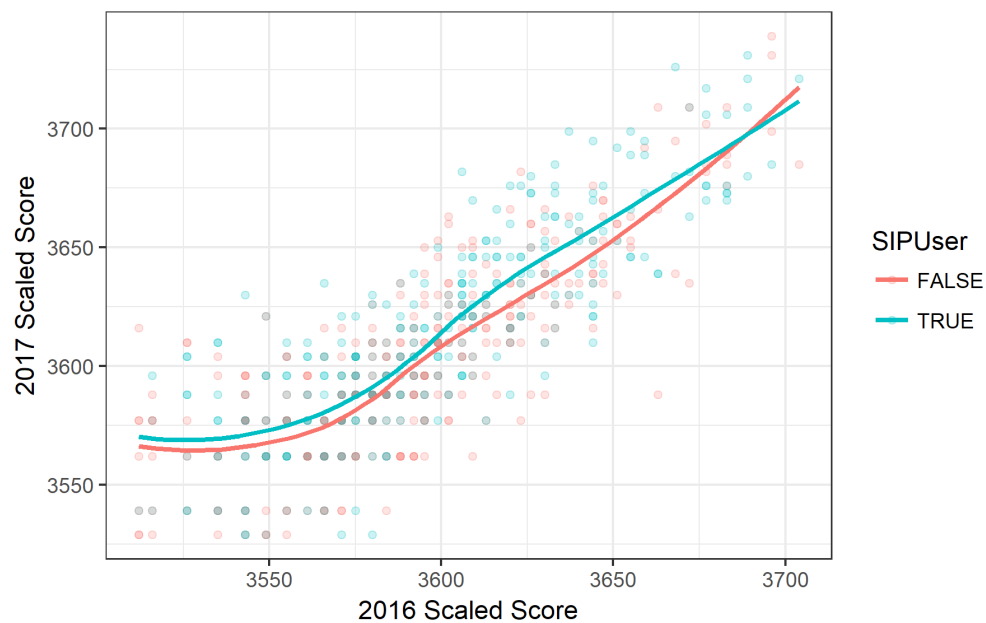


Figure C10. Math Grade 8 User vs Non-User

