edmentum™

Research Report: Impacts of the Use of Study Island Practice and Benchmarks – Reading School District, Pennsylvania

Edmentum Research
September 12, 2018

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

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).

Reading School District (RSD) is a current Study Island partner located 60 miles northwest of Philadelphia and with a total enrollment of over 17,000 students. Ninety-one percent of students in the district are economically disadvantaged, and 83% are Hispanic (Pennsylvania Department of Education, n.d.-b). As a district in Pennsylvania, RSD participates in the state's accountability system. The Pennsylvania Accountability System (PAS) holds schools and districts accountable to a range of measures, including participation rate, graduation or attendance rate, and closing the achievement gap for all students, especially historically underperforming students. As part of this accountability, the Pennsylvania System of School Assessment (PSSA) is administered annually to students in grades 3 through 8 for English language arts (ELA) and math, as well as grades 4, 8, and 11 for science. Assessment data show that RSD tends to perform at levels lower than the state average.

In support of RSD's partnership with Edmentum, this study is intended to provide a research basis for Study Island in terms of the research literature and analyses of RSD students' level of usage and performance data within Study Island compared to their performance on the PSSA.

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 PSSA scores for students participating in Study Island practice and Benchmarks.

Generally, implementation and use of Study Island practice and Benchmarks in RSD vary by grade and content area. In practice, students appear to be answering a moderate number of questions and spending a fair amount of time using the product over the course of the school year. Grade 5, in both ELA and math, showed especially high use in the concentration of students using practice and the intensity of their usage. Where students spend more time, answer more questions, and spread their time over active weeks, positive differences are observed. These differences are evident in significantly different math mean scale scores and proficiency classification in grades 5, 6, and 7 when comparing users to non-users and the strongest users to the weakest. Statistically significant results are also found in grade 5 ELA. In addition, when students are exposed to the Benchmarks, which are widely used in grades 3–7 ELA and math, there is a strong and significant association between scores on the Benchmarks and scores on the PSSA. These significant observations remain even after controlling for student ability, based on students' prior-year PSSA scores.

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

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

Introduction

Education is a key indicator for individual and societal progress. As the Organisation for Economic Cooperation 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 Pennsylvania Accountability System (PAS) holds schools and districts accountable to a range of measures, including participation rate, graduation or attendance rate, and closing the achievement gap for all students, especially historically underperforming students. To support schools, Pennsylvania’s Department of Education provides the Standards Aligned System (SAS) as a resource 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 this accountability, the Pennsylvania System of School Assessment (PSSA) is administered annually to students in grades 3 through 8 for English Language Arts (ELA) and math, as well as grades 4 and 8 for science. The assessments have been built to align to Pennsylvania’s Core Standards and to provide student-level achievement scores and relevant placement into one of four proficiency categories: Advanced, Proficient, Basic, and Below Basic.

Reading School District (RSD) is a current Study Island partner located in Pennsylvania. In support of the district’s partnership with Edmentum, this study is intended to provide a research basis for Study Island in terms of the research literature and analyses of RSD students’ level of usage and performance data within Study Island compared to the students’ performance on the PSSA.

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).

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

Although generally providing feedback to teachers and students regarding student performance can consistently enhance achievement (Adams & Strickland, 2012; Baker, Gersten, & Lee, 2002; Chase & Houmanfar, 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 & Houmanfar, 2009; Dihoff et al., 2004; Lee, Lim, & Grabowksi, 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 (Adams & Strickland, 2012). 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 combining 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).

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

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 related to Study Island practice (questions 1–3) and Study Island Benchmarks (questions 4–5):

1. How did students in RSD use Study Island practice during the 2016–17 school year?
2. Were there significant mean differences in the PSSA state test scores between students who used Study Island practice and those who did not and between those who had a high level of usage and those who had a low level of usage?
3. Was there a significant relationship between PSSA performance-level classification and use of Study Island practice?
4. How did students in RSD use and perform on Study Island Benchmarks during the 2016–17 school year?
5. Was there a significant relationship between student scores on Study Island Benchmarks and their scores on summative, end-of-year PSSA state tests? If so, does the significant relationship between Study Island Benchmark scores and PSSA scores remain after accounting for a student's previous PSSA performance?

To answer these research questions, a description of Study Island and the PSSA is provided, followed by an analysis of the impact of Study Island usage on PSSA 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. While the current study focuses on students in grades 3 through 7, Study Island topics are also available in grades 2 and 8 through 11.

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

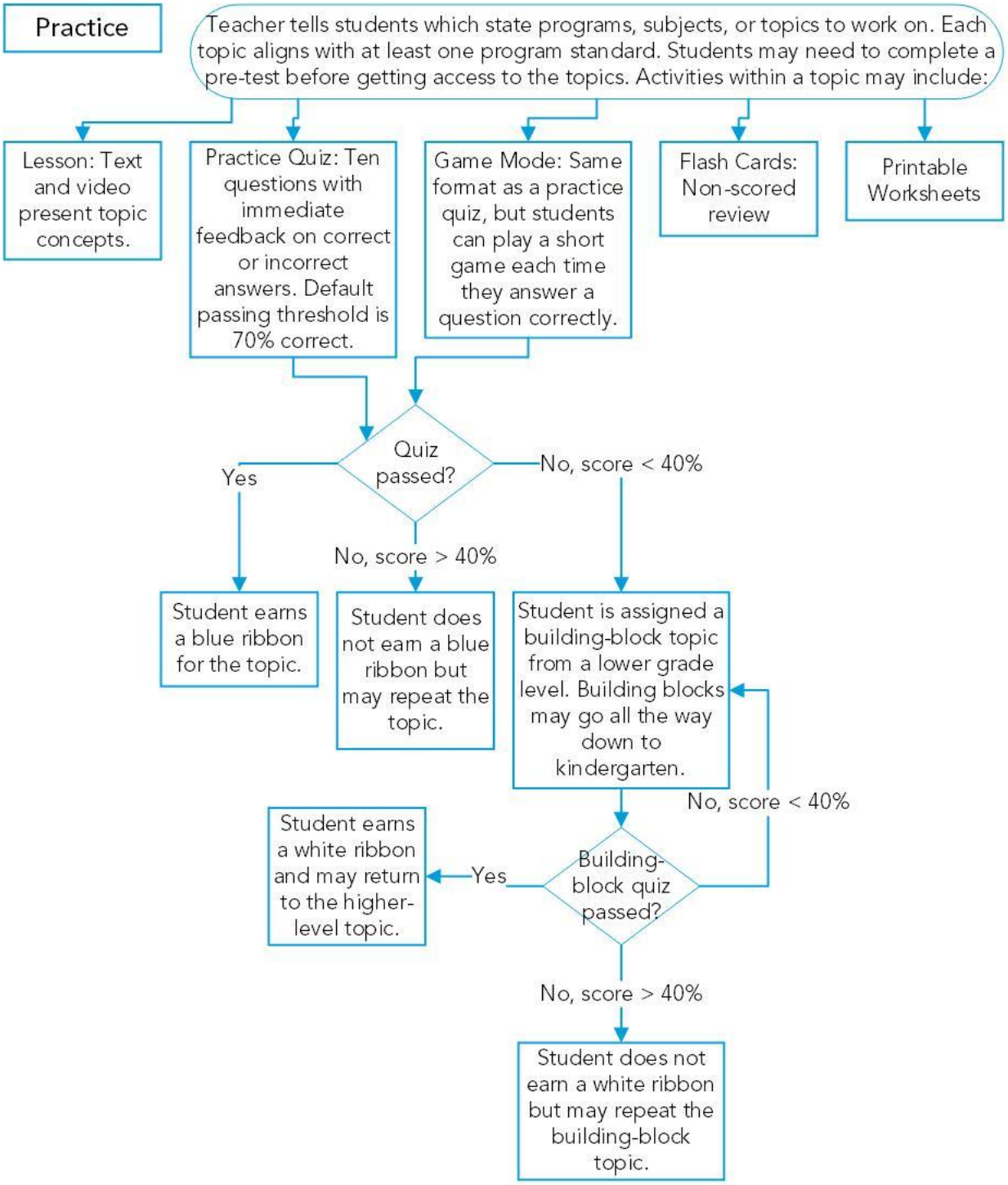
Table 1. Number of Study Island Practice Topics Aligned to Pennsylvania Standards

Grade	ELA	Math	Science
2	35	20	
3	39	27	
4	41	30	30
5	39	20	
6	35	26	
7	38	22	
8	42	20	40
9	19	24	21
10	30	23	20
11		23	

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 practice.

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

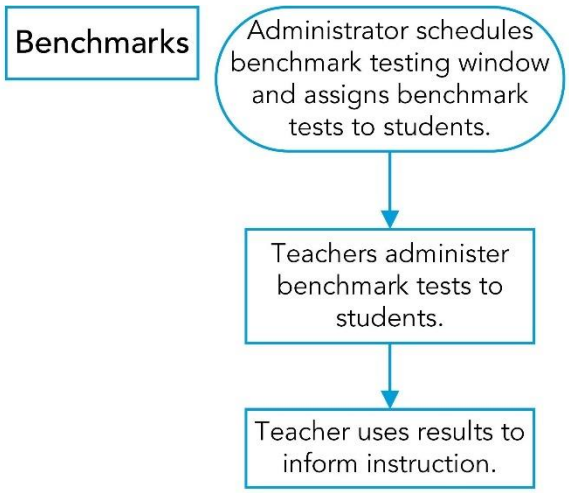
Figure 1. Student Experience, Study Island Practice



Students can also be assigned Benchmark assessments. These have been developed to mirror the content standards covered by the PSSA blueprint. Figure 2 shows the typical experience for Study Island Benchmarks.

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

Figure 2. User Experience, Study Island Benchmarks



The formative, short-cycle practice assessments include multiple-choice (MC) items. The fixed-form, interim-like Benchmark assessments include both MC and constructed-response (CR) items. All MC items are scored online and incorporated into the system's information, while all CR items are scored by the teacher.

Study Island practice and Benchmarks include 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.

Pennsylvania System of School Assessment (PSSA)

Given the focus on accountability, one of the primary research questions of this study relates to the impact of students' use of Study Island on their end-of-year state test scores. The Pennsylvania System of School Assessment (PSSA) assesses students in grades 3 through 8 in mathematics and English language arts (ELA) and students in grades 4 and 8 in science. The assessment is a standards-based (criterion-referenced) test measuring Pennsylvania Core Standards in English Language Arts and Mathematics and the Pennsylvania Academic Standards for Science. The assessment is 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.

To measure those standards, the PSSA is composed of various types of items and is developed according to a test blueprint indicating the proportion of the assessment measuring each set of standards. PSSA assessments include a combination of multiple-choice (MC) and constructed-response (CR) items. The MC items are dichotomously scored, and the CR items are scored on a 0–4-point scale using a scoring guideline. All CR items are scored by independent raters. While math assessments include only MC and CR, ELA assessments use several types of MC and CR items, including the following:

- standalone and passage-based MC, which has only one correct answer among four options and is dichotomously scored
- evidence-based MC, which allows students to select one or more answers and receive partial credit
- short answer (grade 3 only) scored on a 0–3-point scale
- text-dependent analysis (grades 4–8) scored on a 1–4-point scale
- mode-specific writing prompts scored on a 1–4-point scale

The PSSA reports student-level scale scores and performance-level classifications (Below Basic, Basic, Proficient, and Advanced). Scale scores were derived via the Rasch item response theory (IRT) model for each grade and content area. Because the scaled scores are not vertically scaled, meaning the scale does not translate across grades, they are only interpretable within grade and subject. (The Pennsylvania Value Added

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

Assessment System [PVAAS] tracks growth from year to year.) This study will focus on scale scores within grade and performance-level classifications.

Sample

This study was conducted on a convenient sample of students from 17 schools (13 elementary schools and four middle schools) from RSD that were Study Island partners during the 2016-2017 academic year. RSD is located 60 miles northwest of Philadelphia and has a total enrollment of over 17,000 students. Ninety-nine percent of these students qualify for free or reduced-price lunch (compared to a state average of 47%), and 83% are Hispanic, a much higher percentage of the student population than the 10% state average (Pennsylvania Department of Education, n.d.-b). The district provided student-level PSSA 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 practice and Benchmark data via unique student identifiers. For this study, while some high school students in the district used Study Island practice to practice skills aligned to Pennsylvania high school Keystone end-of-course exams, the sample was restricted to elementary and middle school students who are required by the state to take the PSSA. Furthermore, very few 8th graders, who in this district attend intermediate high schools with 9th graders, used Study Island practice. Given the low usage and different implementation context, 8th graders are not included in this analysis.

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.

Table 2. District Demographics Compared to State Average

	District (%)*	State Average (%)*	Difference (District vs. State)
Individualized Education Program (IEP)	19.7	17.6	+2.1
Free and Reduced Lunch	99.4	46.7	+52.7
Hispanic	82.7	10.4	+72.3
Black	8.9	14.8	-5.9
White	5.9	67.5	-61.6
Two or More Races	2.0	3.4	-1.4
Asian or Asian/Pacific Islander	0.5	3.7	-2.2
American Indian/Alaska Native	0.0	0.2	-0.2
Hawaiian/Pacific Islander	0.0	0.1	-0.1
**Ethnicity percentages may not add up to 100% because of rounding.*			
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.			

Table 3 provides the demographic make-up of the sample for this study. It appears that students using Study Island in the sample are comparable to the district as a whole.

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

Table 3. Sample Demographics of Study Island Practice Use

		Complete 2017 District Sample		Sample of SI Users	
Variable	**Category**	**N**	**%**	**N**	**%**
Race / Ethnicity	American Indian / Alaskan Native	3	0	3	0
	Black / African American	565	8	440	8
	Hispanic (any race)	5,762	84	4,444	84
	White	369	5	258	5
	Multi-Racial	163	2	123	2
	Asian	29	0	21	0
	Native Hawaiian / Pacific Islander	1	0	1	0
	Total	6,892	100	5,290	100
Gender	Female	3,364	49	2,626	50
	Male	3,528	51	2,664	50
	Total	6,892	100	5,290	100
Special Education	No	5,457	79	4,245	80
	Yes	1,435	21	1,045	20
	Total	6,892	100	5,290	100
Economically Disadvantaged	No	2,081	30	1,471	28
	Yes	4,811	70	3,819	72
	Total	6,892	100	5,290	100
Title I	No	1	0	NA	NA
	Yes	6,891	100	5,290	100
	Total	6,892	100	5,290	100

Definition of Usage

To evaluate just how much the district is using Study Island, “usage” is defined in terms of two participatory factors: Study Island practice (or practice) and Study Island Benchmarks (or Benchmarks). In this paper, usage is defined differently for practice and Benchmarks.

Practice

For this study, usage in practice 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).

Benchmarks

Benchmarks offer four fixed-form formative assessments per subject, per grade level, aligned to state-specific and Common Core standards. These assessments are typically 30 to 40 items long and are designed to be taken periodically throughout the school year. Each Benchmark is built following the blueprint for the state summative test. The close alignment between state tests and Benchmarks suggests that Benchmark results may be predictive of how prepared students could be for their state tests.

For Benchmarks, student usage is defined as the completion of a Study Island Benchmark form.

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

Patterns of Use

Practice

Table 4 provides the total number of unique students answering any practice questions in any session for a grade, compared to the total number of students enrolled in the district. The far-right column, "ELA or Math," shows the number of students using practice for at least one subject. Overall, a large proportion of the district's 3rd through 7th graders are using practice. The concentration of users is particularly strong in elementary school, with 79–91% of students using practice in grades 3, 4, and 5 in either ELA or math. The proportion of middle schoolers in grades 6 and 7 using practice in either ELA or math ranges from 56 to 66%, with higher concentrations of students using math, compared to ELA. Eighth graders in the district attend separate intermediate high schools, with small proportions of these students using practice, so these students are omitted from this analysis.

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

		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	1,451	1,016	70	1,075	74	1,153	79
4	1,460	1,013	69	1,069	73	1,186	81
5	1,393	1,220	88	1,173	84	1,272	91
6	1,408	619	44	908	64	923	66
7	1,341	533	40	725	54	756	56
Total	7,053	4,401	62	4,950	70	5,290	75
Note. Total district enrollment counts from Pennsylvania Department of Education, Enrollment Reports and Projections							

Data from across the district suggest that Study Island practice 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.

Benchmarks

Table 5 shows the number of students responding to Benchmarks. The far-right column, "ELA or Math," shows the number of students using the Benchmarks for at least one subject. Benchmark usage is highly concentrated in the district, with 93% to 96% of students per grade participating. Administrations of Benchmark forms in RSD appear to have followed general testing windows in which Form 1 is delivered during the fall, Form 2 during the winter, and Forms 3 and 4 during the spring. The volumes of Benchmark test use by administration date are available in Appendix B.

Table 5. Total Number and District Proportion of Students Using Study Island Benchmarks

		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	1,451	1,369	94	1,381	95	1,384	95
4	1,460	1,348	92	1,357	93	1,358	93
5	1,393	1,324	95	1,330	95	1,335	96
6	1,408	1,307	93	1,313	93	1,315	93
7	1,341	1,244	93	1,250	93	1,251	93
Total	7,053	6,592	93	6,631	94	6,643	94
Note. Total district enrollment counts from Pennsylvania Department of Education, Enrollment Reports and Projections							

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

Analyses: Study Island Practice

Research Question 1: How did students in Reading School District use Study Island practice during the 2016-17 school year?

As discussed earlier, students are considered users of Study Island practice when they use practice at any level 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 6 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 45% in grade 4 ELA to 57% in grade 3 math.

Table 6. Descriptive Statistics for Total Number Attempted and Proportion Correct, Study Island Practice 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	1,016	1	141.00	5,096	258.58	367.42	0.00	0.47	1.00	0.46	0.17
	4	1,013	1	82.00	4,231	180.07	298.28	0.00	0.45	1.00	0.45	0.18
	5	1,220	1	206.50	2,959	321.42	348.04	0.00	0.54	1.00	0.53	0.16
	6	619	1	166.00	2,029	271.84	310.84	0.00	0.52	1.00	0.50	0.18
	7	533	1	177.00	1,535	256.94	253.05	0.00	0.54	1.00	0.53	0.17
	Total	4,401	1	150.00	5,096	259.60	330.17	0.00	0.50	1.00	0.49	0.17
Math	3	1,075	1	142.00	2,250	238.41	296.77	0.00	0.58	1.00	0.57	0.18
	4	1,069	1	87.00	3,593	179.40	260.99	0.00	0.56	1.00	0.53	0.18
	5	1,173	1	182.00	4,386	366.33	509.28	0.00	0.58	1.00	0.56	0.18
	6	908	1	135.00	2,461	303.81	385.55	0.00	0.56	1.00	0.53	0.20
	7	725	1	134.00	1,512	177.00	175.95	0.00	0.48	0.90	0.47	0.18
	Total	4,950	1	135.00	4,386	258.98	364.07	0.00	0.56	1.00	0.54	0.18

Table 7 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. Fourth graders spend, on average, the least amount of time and answer the fewest items with a median of about 70 minutes (Table 7) and 82 and 87 items answered in ELA and math, respectively (Table 6). Grade 5 students spent the most amount of time overall in both subjects: about 225 minutes in ELA and 245 minutes in math, or around four hours per subject, answering, on average, almost 320 items and 370 items respectively. Sixth graders also spend more time in math: about 192 minutes, or 3.2 hours, answering just over 300 items.

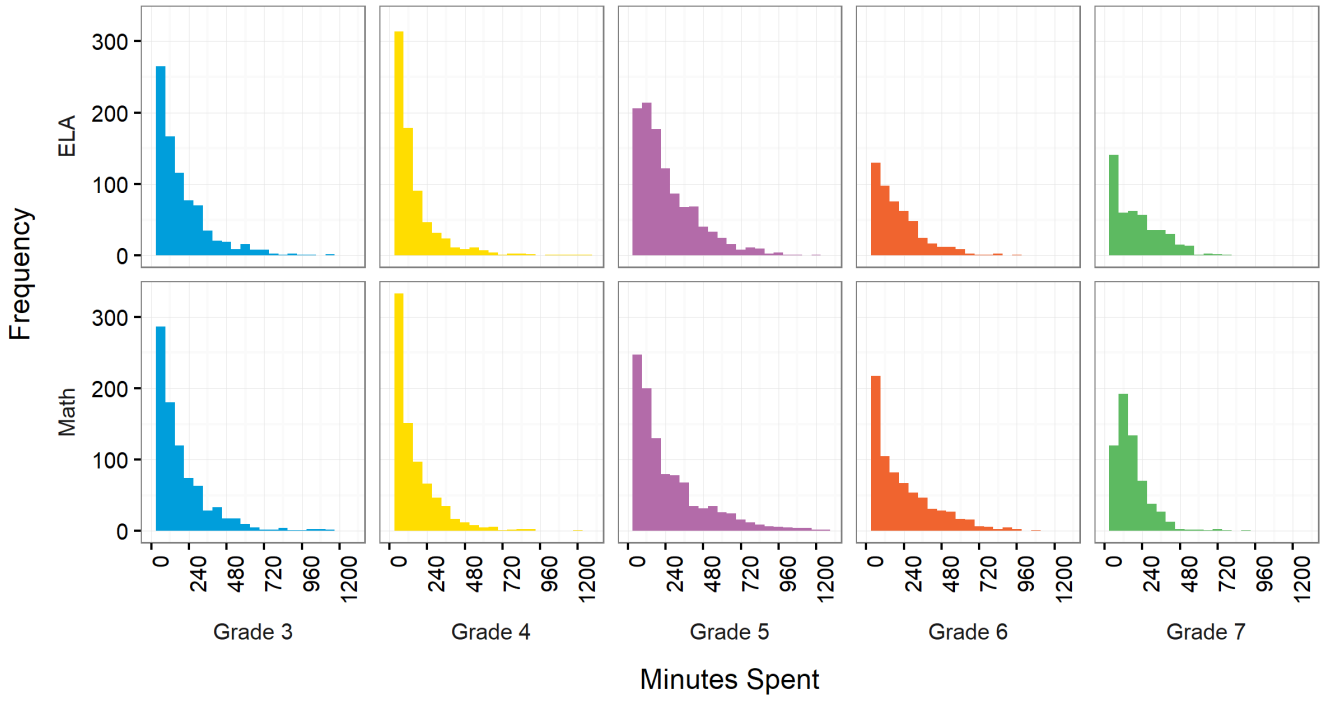
5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

Table 7. Descriptive Statistics for Total Amount of Time (Minutes), Study Island Practice Users, 2016-17 School Year

Subject	Grade	N	Min	Median	Max	Mean	SD
ELA	3	1,016	0.12	107.50	2,595.78	178.00	230.16
	4	1,013	0.28	67.62	1,853.67	124.62	169.88
	5	1,220	0.15	172.00	1,184.82	224.64	193.27
	6	619	0.53	117.42	936.32	167.62	159.64
	7	533	1.62	139.85	806.78	180.42	155.58
	Total	4,401	0.12	119.28	2,595.78	177.47	192.32
Math	3	1,075	0.47	101.93	1,445.53	160.93	187.32
	4	1,069	0.28	69.12	2,141.13	128.08	161.89
	5	1,173	0.20	150.63	2,149.53	245.31	268.02
	6	908	0.10	110.53	1,086.80	191.95	199.73
	7	725	0.08	130.80	870.53	147.53	117.74
	Total	4,950	0.08	112.27	2,149.53	177.56	203.64

Figure 3 shows how many students are distributed across the amount of time spent in Study Island practice. There are many student users in grade 4, but they spent less time using Study Island than the fewer users spending more time in grade 6.

Figure 3. Distribution of Time in Minutes by Grade and Content Area for Study Island Practice Users in the 2016-17 School Year



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 8 shows the total number of weeks with any use, or “active weeks.” On average, the greatest number of active weeks are in grade 5, for both math and ELA, at about nine weeks. These data show that the higher usage in grade 5 is spread across more weeks during the school year as compared to other grades.

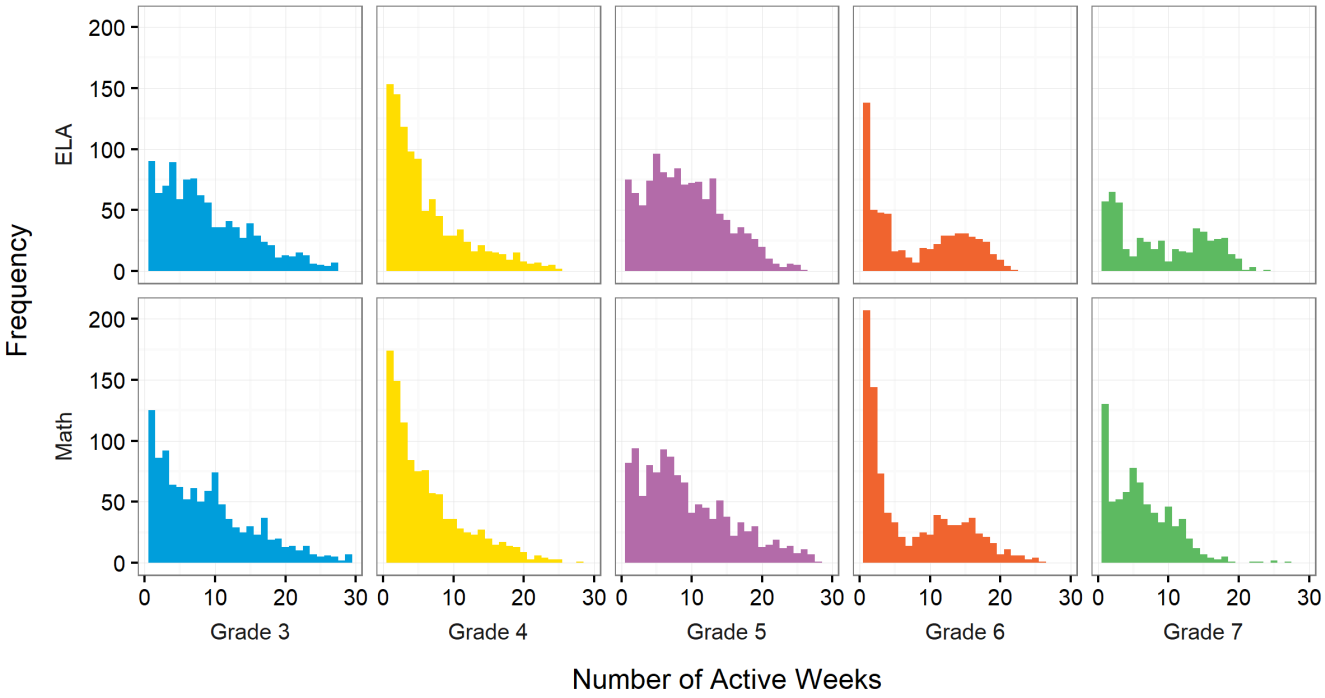
5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

Table 8. Descriptive Statistics for Active Weeks Using Study Island Practice, 2016-17 School Year

Subject	Grade	N	Min	Median	Max	Mean	SD
ELA	3	1,016	1	7	27	8.76	6.20
	4	1,013	1	4	25	6.22	5.35
	5	1,220	1	9	26	9.22	5.47
	6	619	1	6	22	8.04	6.36
	7	533	1	8	24	8.87	6.27
	Total	4,401	1	7	27	8.21	5.96
Math	3	1,075	1	7	29	8.69	6.57
	4	1,069	1	5	28	6.33	5.37
	5	1,173	1	8	28	9.34	6.44
	6	908	1	4	26	7.27	6.50
	7	725	1	5	27	6.16	4.36
	Total	4,950	1	6	29	7.70	6.13

Figure 4 shows the distribution of active weeks for each grade and subject. Here we see that, generally, grades 3 and 5 have more students in both ELA and math with more active weeks. These views help to illustrate how students are using the practice items across grades and across subject areas. They help to show that, for example, there are many grade 4 students using Study Island practice, but they are not using it across as many active weeks as grade 5 students are.

Figure 4. Distribution of Active Weeks by Grade and Content Area for Study Island Practice Users in the 2016-17 School Year



Finally, to see just how much of the time occurs within each active week, Table 9 provides the amount of time per week, calculating the total time spent in practice divided by the number of active weeks.

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

Table 9. Descriptive Statistics for Time per Active Week (Minutes), Study Island Practice

Subject	Grade	N	Min	Median	Max	Mean	SD
ELA	3	1,016	0.12	15.57	96.14	16.91	10.36
	4	1,013	0.28	16.34	84.00	17.46	9.72
	5	1,220	0.15	19.67	95.41	23.07	13.67
	6	619	0.53	20.00	90.65	22.33	13.32
	7	533	1.62	18.32	61.25	19.72	8.77
	Total	4,401	0.12	17.59	96.14	19.85	11.81
Math	3	1,075	0.47	14.75	64.27	16.25	8.32
	4	1,069	0.28	17.20	85.65	17.79	8.16
	5	1,173	0.20	21.11	92.23	22.36	11.68
	6	908	0.10	24.88	188.18	27.54	17.05
	7	725	0.08	22.68	105.42	24.33	11.34
	Total	4,950	0.08	18.98	188.18	21.29	12.28

PSSA Performance and Use of Study Island Practice

Research Question 2: Are there significant mean differences in the PSSA state test scores between students who use Study Island practice and those who do not and between those who have a high level of usage and those who have a low level of usage?

Users vs. Non-Users

To contextualize RSD's PSSA performance, we begin with a descriptive look at mean PSSA scores. Table 10 compares performance on the PSSA in terms of scale scores for each of the Study Island user and Study Island non-user groups, compared to the district and state by grade and content area. The table shows that RSD has substantially lower mean scale scores compared to the state. Study Island users have similar PSSA scores to RSD overall, and those are higher than non-users' scores. Specifically, Study Island users outperform the district slightly and also outperform Study Island non-users in all content areas and grades. Mean scores differ as much as 30 points in grade 5 ELA and 26 points in grade 5 math.

Table 10. Descriptive 2017 PSSA Scale Scores of Study Island Users, Study Island Non-Users, RSD, and State

		SI User			SI Non-User			District			State		
Subject	Grade	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
ELA	3	1,016	953.27	88.10	357	932.31	84.30	1,373	947.71	87.57	124,923	1039.30	111.21
	4	1,013	936.28	89.52	348	933.86	90.88	1,361	935.65	89.85	125,200	1030.55	112.72
	5	1,220	944.05	84.36	110	914.06	86.38	1,330	941.50	84.91	124,183	1029.58	112.26
	6	619	943.09	84.19	684	937.73	76.57	1,303	940.25	80.27	123,170	1035.08	106.23
	7	533	947.08	94.37	708	932.51	86.01	1,241	938.67	89.89	125,744	1031.71	113.46
Math	3	1,075	924.01	94.48	362	901.69	84.40	1,437	918.75	92.66	125,205	1019.85	129.66
	4	1,069	902.19	82.86	348	892.93	82.92	1,417	900.13	82.93	125,575	993.58	118.67
	5	1,173	905.99	76.35	197	889.36	75.02	1,370	903.82	76.35	124,405	991.82	119.70
	6	908	892.06	80.37	454	865.51	59.66	1,362	883.55	75.37	123,112	976.25	115.64
	7	725	886.53	90.32	567	873.39	82.88	1,292	880.88	87.40	125,584	968.65	126.69

To discern whether these differences are significant, we must consider the differences in student ability across the user groups. If students using Study Island are generally higher-ability students, whether or not they use Study Island may be meaningless with regard to their performance on the PSSA. To understand the association between using Study Island and PSSA performance, only students with similar PSSA 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.

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

A nearest neighbor propensity score matching (PSM) (Rosenbaum & Rubin, 1983) was conducted to match students in the user group with students in the non-user group by ability (as measured by students' 2016 PSSA scores) so that statistical analyses of the 2017 PSSA 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 grades 4 through 7 ELA and math could be included in the analysis because third graders do not have a prior-year PSSA score. Some other students within these grades were also eliminated from the sample because they did not have a 2016 PSSA score. The resulting matched sample size was dependent on the size of the user and non-user groups, with the total number of cases able to be matched determined by the group with the smaller size. Because Study Island practice was widely used in many grades within the district with few non-users, the PSM process resulted in discarding substantial proportions of the user data in all grades in math and in grades 4 and 5 in ELA. The total resulting N is included in Table 11. (See Appendix C for figures that show the spread of scores across Study Island users [High True] and Study Island non-users [High False] and the resulting PSM.)

Table 11 reports the results from a *t*-test comparing the mean 2016 PSSA scores of Study Island users to Study Island non-users to determine whether it was possible to create equivalent matched groups with nearest neighbor PSM. These findings show that the average 2016 PSSA scale scores for the matched samples were not significantly different from each other, thus enabling us to compare 2017 PSSA outcomes between groups of equivalent ability.

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

		SI User		SI Non-User		Matched	PSSA 2016				
Subject	**Grade**	**Mean**	**SD**	**Mean**	**SD**	**N**	**Mean Difference**	**95% CI**		***t***	**df**
ELA	4	944.36	92.72	936.87	89.77	285	7.49	-22.51	7.52	-0.980	567.40
	5	922.47	89.26	912.34	99.06	74	10.14	-40.77	20.50	-0.654	144.45
	6	942.50	89.82	941.25	88.05	492	1.24	-12.37	9.89	-0.219	981.61
	7	940.03	88.39	939.84	88.24	444	0.20	-11.83	11.44	-0.033	886.00
Math	4	916.44	94.87	914.93	93.69	229	1.51	-18.83	15.80	-0.171	455.93
	5	891.21	75.88	904.76	91.36	132	-13.55	-6.81	33.90	1.310	253.46
	6	905.53	77.16	906.21	70.81	346	-0.69	-10.37	11.74	0.122	684.98
	7	870.53	81.61	860.85	86.13	442	9.68	-20.76	1.40	-1.715	879.44

A *t*-test was conducted after matching to compare the 2017 PSSA scores across the matched Study Island user and Study Island non-user groups. Results from the analysis are shown in Table 12, where *N* reports the equal size of the matched groups.

Table 12. t-Test Comparisons of 2017 PSSA Scale Score Between Matched Study Island Users and Non-Users

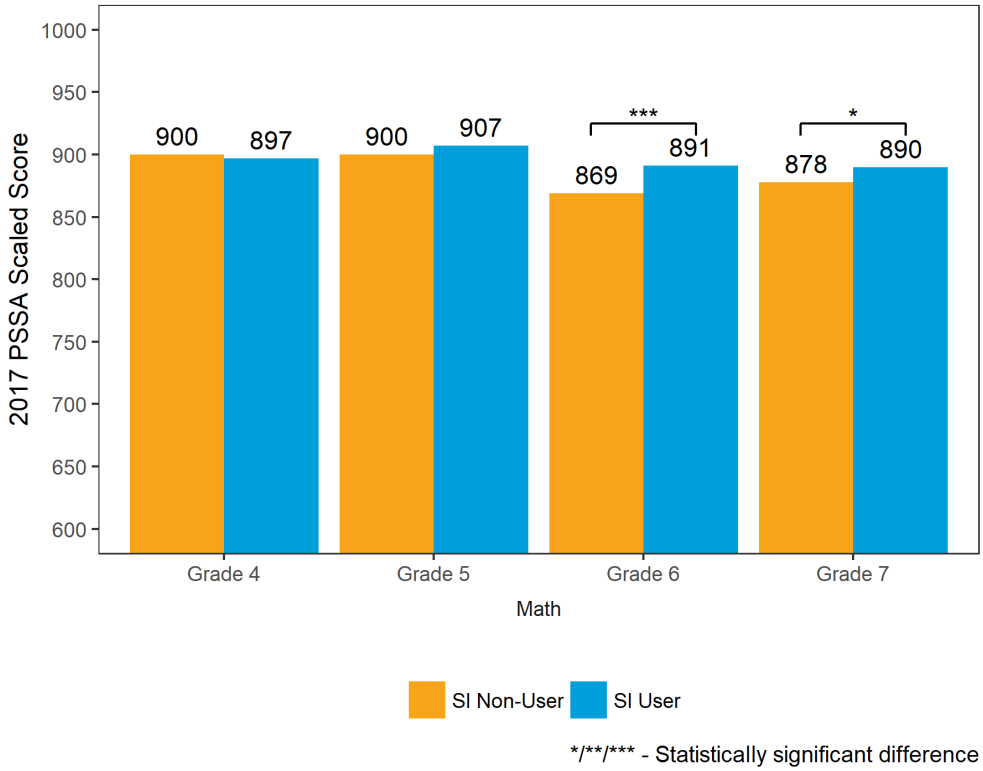
		SI User		SI Non-User		Matched	PSSA 2017				
Subject	**Grade**	**Mean**	**SD**	**Mean**	**SD**	**N**	**Mean Difference**	**95% CI**		***t***	**df**
ELA	4	944.78	92.67	938.00	89.36	285	6.77	-21.75	8.21	-0.888	567.25
	5	952.20	87.14	920.88	80.47	74	31.32	-58.58	-4.07	-2.272*	145.08
	6	944.52	82.76	943.34	76.46	492	1.18	-11.15	8.79	-0.233	975.89
	7	950.93	93.07	941.89	85.33	444	9.04	-20.80	2.72	-1.509	879.41
Math	4	897.64	75.46	900.22	83.42	229	-2.59	-12.02	17.19	0.348	451.49
	5	907.91	67.34	900.02	77.12	132	7.89	-25.44	9.65	-0.886	257.33
	6	891.79	78.88	869.42	60.95	346	22.37	-32.89	-11.85	-4.174***	648.70
	7	890.28	88.43	878.73	83.59	442	11.54	-22.90	-0.18	-1.994*	879.22

=significant at the .05 level **=significant at the .01 level *=significant at the .001 level*

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

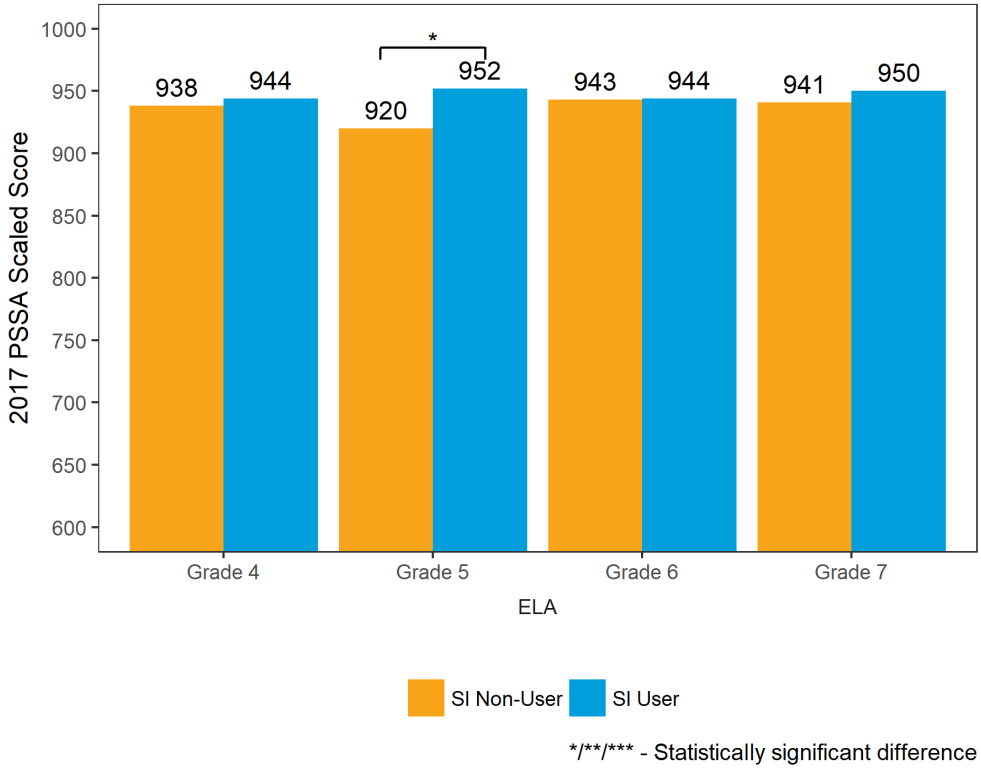
While the mean PSSA scale score for the Study Island user group is larger than for the matched non-user group in every category except for grade 4 math, only the mean scale score differences for grades 6 and 7 math and grade 5 ELA are statistically significant once the groups are matched based on 2016 PSSA score. Table 12 includes a column that shows the mean difference in 2017 PSSA scores between users and non-users. For example, in grade 6 math, the mean PSSA scores for the Study Island practice users is 22 points higher than for the matched sample of non-users, a statistically significant difference. Figures 5 and 6 show a visual representation of the mean scale scores by user group once they are matched based on 2016 ability.

Figure 5. Mean 2017 PSSA Scores by Grade and Usage Group, Math



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

Figure 6. Mean 2017 PSSA Scores by Grade and Usage Group, ELA



High Usage vs. Low Usage

Because of the high concentration of Study Island practice users in many grades within the district, we further examined the user group by breaking students into high- and low-usage groups relative to their grade and subject. For example, in grade 5 where Study Island practice usage is almost universal across the student population, we can compare student outcomes between students with high levels of usage and similar-ability students with low levels of usage. High usage is defined within grade and subject by taking users who fall in at least the 70th percentile for both active weeks *and* total time. Low usage is defined similarly by considering users who fall below the 70th percentile for both active weeks *and* total time, with some remainder of users who do not meet either criteria. Table 13 reports the frequency of students by high- and low-user group, along with the 70th percentiles for active weeks and total time that are used to define the high- and low-usage groups. Because the 70th percentile cut points were determined within grade and subject, different criteria are used to select students for the high-usage group, with grade 4 requiring less time and fewer active weeks in order to qualify for the high-usage group. Consistent with previous findings, grade 5 in both ELA and math has the highest usage required to be considered a high user, with at least 283 minutes.

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

Table 13. Study Island Practice Users by Usage Classification Level

		Total	70th Percentile		High Usage		Low Usage	
Subject	Grade	N	Active Weeks	Time Spent	N	%	N	%
ELA	4	1,130	8	134.76	195	17	538	48
	5	1,110	12	283.86	220	20	620	56
	6	1,072	14	231.11	99	9	297	28
	7	1,045	14	245.09	102	10	277	27
Math	4	1,145	8	141.11	218	19	588	51
	5	1,133	12	283.27	245	22	635	56
	6	1,093	12	236.65	162	15	473	43
	7	1,067	8	189.70	135	13	370	35

Similar to our earlier analysis, nearest neighbor propensity score matching was used to match high users to low users as before, creating matched pairs of students with equivalent 2016 PSSA scores. Table 14 reports the results from a *t*-test comparing mean 2016 PSSA scores of Study Island high users and low users in order to determine the statistical equivalence of the matched groups. Results show small mean differences, none of which are statistically significant, suggesting that the matching has successfully created user groups of equivalent ability.

Table 14. t-Test Comparisons of Propensity Score (2016 PSSA) Between Matched Study Island High-Usage and Low-Usage Users

		SI High Usage		SI Low Usage		Matched	PSSA 2016				
Subject	Grade	Mean	SD	Mean	SD	N	Mean Difference	95% CI		*t*	df
ELA	4	968.46	91.26	968.07	90.56	195	0.39	-18.50	17.71	-0.043	387.98
	5	934.63	90.29	934.44	89.98	220	0.20	-17.09	16.69	-0.023	437.99
	6	949.30	83.07	948.88	83.01	99	0.42	-23.70	22.85	-0.036	196.00
	7	962.84	79.28	962.39	78.79	100	0.45	-22.49	21.59	-0.040	197.99
Math	4	934.84	104.97	934.90	105.24	218	-0.06	-19.72	19.85	0.006	434.00
	5	895.06	85.51	895.39	85.58	245	-0.33	-14.86	15.51	0.042	488.00
	6	920.21	76.15	919.99	76.85	162	0.22	-16.94	16.51	-0.025	321.97
	7	886.49	77.65	886.70	77.21	135	-0.21	-18.34	18.77	0.023	267.99

In order to see when there were differences in PSSA achievement between high and low users, a *t*-test was conducted after matching to compare the mean 2017 PSSA scores between the matched groups. Results from the analysis are shown in Table 15, including the mean difference. Here, results show that while the mean 2017 PSSA score is higher for the high usage group in grades 5 through 7 in both math and ELA, the mean difference is statistically significant for math in grades 5 and 7. Given the concentration of users in grade 5, this comparison of outcomes between high and low users is a more appropriate analysis. No statistically significant differences in 2017 PSSA mean scores between high and low users were found in any grades in ELA. Figures 7 and 8 show a visual representation of the mean scale scores by high- and low-user group once they are matched based on 2016 ability.

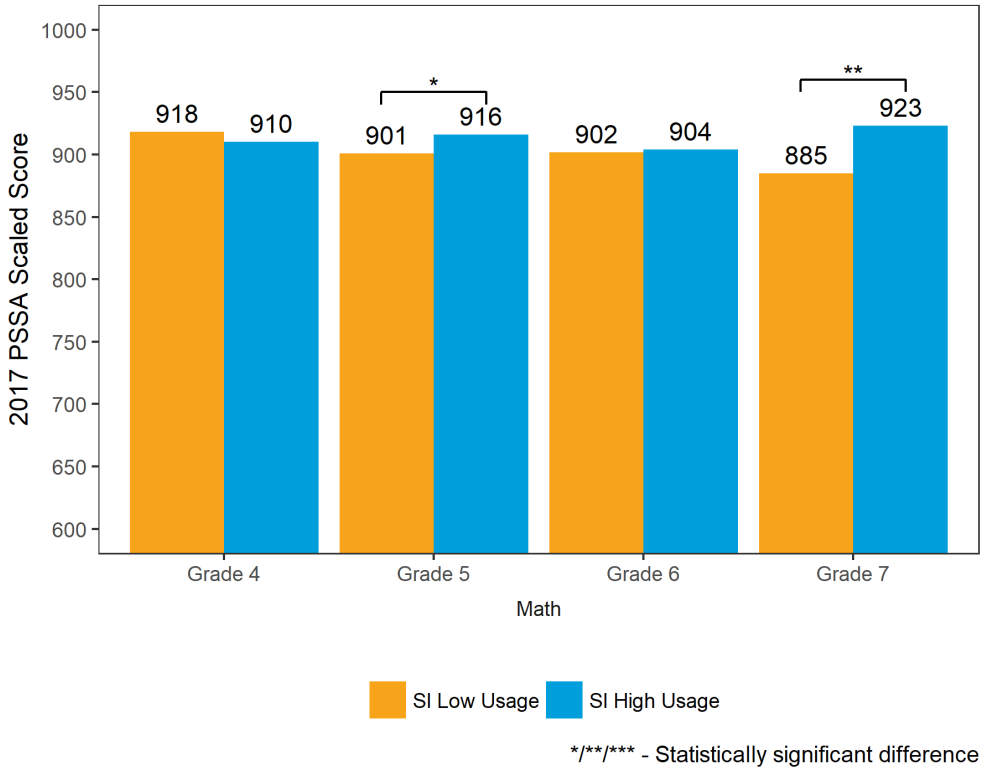
5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

Table 15. t-Test Comparisons of 2017 PSSA Scale Score Between Matched Study Island High-Usage and Low-Usage Users

		SI High Usage		SI Low Usage		Matched	PSSA 2017				
Subject	**Grade**	**Mean**	**SD**	**Mean**	**SD**	**N**	**Mean Difference**	**95% CI**		***t***	**df**
ELA	4	952.59	88.81	954.76	97.52	195	-2.17	-16.40	20.74	0.230	384.65
	5	962.38	80.51	949.88	84.93	220	12.50	-28.01	3.01	-1.584	436.75
	6	951.69	68.84	941.52	79.45	99	10.17	-31.01	10.67	-0.963	192.11
	7	983.87	81.82	962.49	87.12	100	21.38	-44.95	2.19	-1.789	197.23
Math	4	910.06	89.40	918.27	84.24	218	-8.21	-8.15	24.56	0.986	432.48
	5	916.98	74.80	901.07	75.49	245	15.91	-29.25	-2.57	-2.344*	487.96
	6	904.90	71.84	902.46	77.18	162	2.44	-18.74	13.86	-0.294	320.36
	7	923.38	99.80	885.81	87.17	135	37.56	-60.02	-15.11	-3.294**	263.24

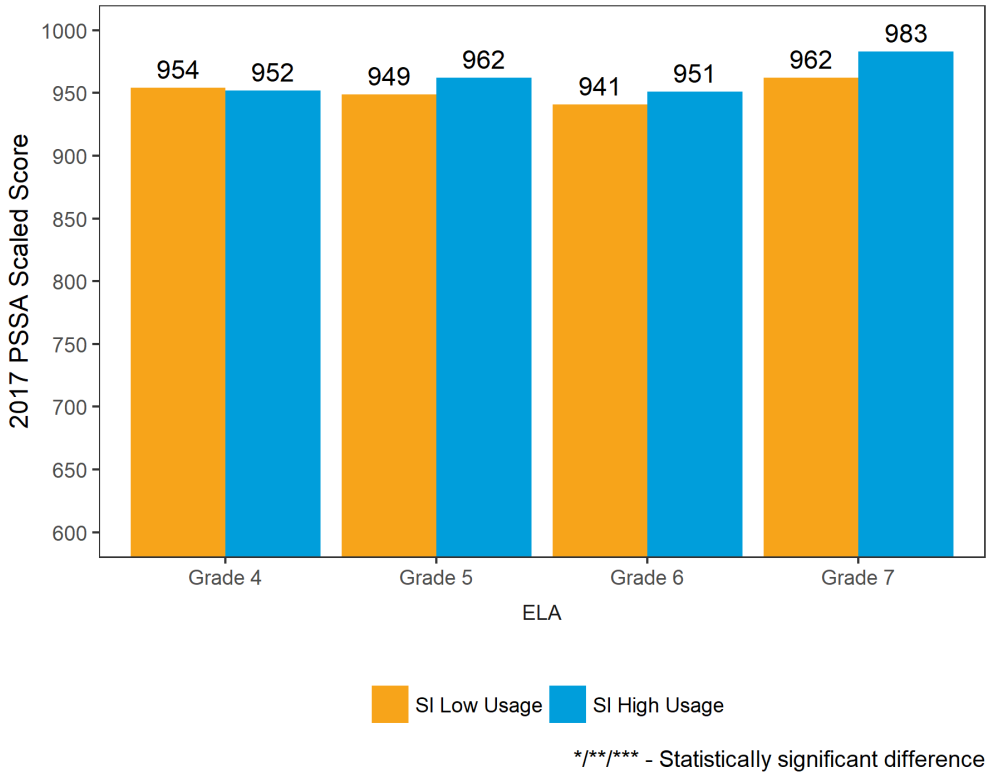
=significant at the .05 level **=significant at the .01 level *=significant at the .001 level*

Figure 7. Mean 2017 PSSA Scores by Grade and Usage Group, Math



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

Figure 8. Mean 2017 PSSA Scores by Grade and Usage Group, ELA



Research Question 3: Is there a significant relationship between PSSA performance-level classification and use of Study Island practice?

To answer this question, we specifically consider practice in two ways: 1) use vs. no use and 2) among users, high use vs. low. Because the performance level is a key variable in accountability, Table 16 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, RSD, and the state. For total N counts by group on which the percentages are based, see Table 10. Similar differences in overall proficiency are seen with mean scale score differences: in most groups, Study Island users tend to have higher proficiency percentages than Study Island non-users, Study Island users are more similar to the district overall, and the district has far lower proficiency rates than the state.

Table 16. Percentage of Students in Grades 3–8 Scoring Proficient or Advanced on the 2017 PSSA, District vs State

		SI User		SI Non-User		District		State	
Subject	**Grade**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
ELA	3	303	30	81	23	384	28	80,700	65
	4	235	23	84	24	319	23	76,246	61
	5	296	24	20	18	316	24	74,013	60
	6	162	26	160	23	322	25	78,336	64
	7	145	27	146	21	291	23	74,817	60
Math	3	236	22	64	18	300	21	68,236	54
	4	143	13	60	17	203	14	58,517	47
	5	133	11	23	12	156	11	54,489	44
	6	85	9	20	4	105	8	49,614	40
	7	85	12	66	12	151	12	47,470	38

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

By using the same matched groups from the PSM that were used earlier to explore mean score difference (see Table 11 for cell counts), we ran a chi-square test to discern the relationship in performance-level classification between Study Island users and non-users; findings are reported in Table 17. Two groups, grade 6 and grade 7 math, show a statistically significant relationship in the student performance-level classification. Both grades also had statistically different mean PSSA scale scores as shown earlier. Both of these Study Island user groups have smaller proportions of students scoring at the Below Basic proficiency level and larger proportions scoring at the Basic level and higher.

Table 17. Chi-Square Test Comparison of 2017 PSSA Performance-Level Classification Between Matched Study Island Users and Non-Users

		ELA			Math		
Grade	**Performance Level**	**User (%)**	**Non-User (%)**	**Chi-Sq.**	**User (%)**	**Non-User (%)**	**Chi-Sq.**
4	Below Basic	30	28	6.137	63	54	4.187
	Basic	45	47		26	34	
	Proficient	18	22		9	10	
	Advanced	6	2		1	1	
5	Below Basic	24	38	5.066	48	58	5.444
	Basic	47	43		44	33	
	Proficient	26	19		8	8	
	Advanced	3	0		0	2	
6	Below Basic	22	21	6.808	62	71	17.673***
	Basic	53	53		29	27	
	Proficient	23	25		8	2	
	Advanced	3	1		2	0	
7	Below Basic	11	11	6.035	63	71	8.097*
	Basic	59	64		25	19	
	Proficient	26	23		10	8	
	Advanced	3	1		2	2	

=significant at the .05 level **=significant at the .01 level *=significant at the .001 level*

To examine proficiency for grades where the use of Study Island practice is highly concentrated, we use the same matched groups as before to compare the PSSA proficiency classification of high users to low users (see Table 14 for cell counts; see Table 13 for operationalization of high and low usage). A similar chi-square test was run to discern the relationship in proficiency between the matched groups of high and low users; findings are reported in Table 18. Here, both grade 5 and grade 7 math show statistically significant relationships in the proficiency classification. Given the high concentration of Study Island users within grade 5, this result is particularly relevant, showing that 5th graders with high levels of usage in Study Island math are less likely to score at the Below Basic level than similar-ability students with low usage.

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

Table 18. Chi-Square Test Comparison of 2017 PSSA Performance-Level Classification Between Matched Study Island High-Usage and Low-Usage Users

		ELA			Math		
Grade	**Performance Level**	**High Usage (%)**	**Low Usage (%)**	**Chi-Sq.**	**High Usage (%)**	**Low Usage (%)**	**Chi-Sq.**
4	Below Basic	25	26	0.914	55	48	4.040
	Basic	45	44		28	36	
	Proficient	25	24		14	14	
	Advanced	5	7		4	2	
5	Below Basic	22	25	6.366	43	57	10.129*
	Basic	45	48		43	33	
	Proficient	33	25		13	9	
	Advanced	0	2		1	1	
6	Below Basic	14	24	7.649	52	59	3.128
	Basic	67	48		38	31	
	Proficient	16	25		8	7	
	Advanced	3	2		2	4	
7	Below Basic	2	6	5.352	51	62	9.345*
	Basic	56	65		30	27	
	Proficient	38	25		14	10	
	Advanced	4	4		5	0	

=significant at the .05 level **=significant at the .01 level *=significant at the .001 level*

Analysis: Study Island Benchmarks and the PSSA

Research Question 4: How did students in RSD use and perform on Study Island Benchmarks during the 2016-17 school year?

As discussed earlier, Study Island Benchmarks were administered to almost all RSD students. Appendix B shows the volume of usage of Study Island Benchmarks by grade and subject over the course of the school year. To evaluate the scores on Study Island Benchmarks, student data include only the MC item responses. CR items are not always assigned or graded by the teacher, nor can Edmentum guarantee that scoring rubrics are applied with fidelity or consistency. Thus, using the MC items alone, the maximum score for the Pennsylvania Study Island Benchmarks is 28 for ELA and math in all grades 3–8. Table 19 reports descriptive statistics for student performance on Study Island Benchmark MC items for fall (Benchmark 1), winter (Benchmark 2 in December and Benchmark 3 in February-March), and spring (Benchmark 4) administrations. Table 19 shows that most students in all grades and both subjects take forms 1, 2, and 4, with a substantial decrease in the counts of students taking Benchmark form 3. For the subsequent analysis, we do not include the results from Benchmark form 3, and our sample is based on students with reported scores for forms 1, 2, and 4, in addition to the 2016 and 2017 PSSA.

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

Table 19. Benchmark Raw Scores Descriptive Statistics

Subject	Grade	Benchmark	Max Score Possible	N	Min.	Max.	Mean	SD
ELA	3	Benchmark 1	28	1,203	0	24	8.90	3.57
		Benchmark 2	28	1,249	0	26	10.88	4.51
		Benchmark 3	28	555	1	27	12.42	5.02
		Benchmark 4	28	1,270	0	26	11.27	4.95
	4	Benchmark 1	28	1,199	0	26	10.67	4.90
		Benchmark 2	28	1,208	0	27	11.63	5.51
		Benchmark 3	28	454	0	26	12.45	5.46
		Benchmark 4	28	1,281	0	28	11.37	5.36
	5	Benchmark 1	28	1,174	1	28	11.89	5.18
		Benchmark 2	28	1,228	0	27	12.57	5.34
		Benchmark 3	28	708	1	26	13.70	5.24
		Benchmark 4	28	1,249	0	28	12.44	5.16
	6	Benchmark 1	28	1,185	0	26	11.00	4.26
		Benchmark 2	28	1,221	1	27	12.83	5.23
		Benchmark 3	28	923	0	25	13.15	4.85
		Benchmark 4	28	1,137	0	28	14.59	6.28
	7	Benchmark 1	28	1,127	2	27	13.24	5.46
		Benchmark 2	28	1,138	1	27	12.72	5.05
		Benchmark 3	28	918	0	27	14.39	5.39
		Benchmark 4	28	1,105	0	26	11.44	5.12
Math	3	Benchmark 1	28	1,238	0	26	9.77	3.55
		Benchmark 2	28	1,262	0	27	12.87	4.53
		Benchmark 3	28	588	0	27	14.07	4.91
		Benchmark 4	28	1,282	0	28	16.32	5.70
	4	Benchmark 1	28	1,224	1	26	10.39	3.41
		Benchmark 2	28	1,247	0	26	11.46	4.00
		Benchmark 3	28	461	1	25	11.57	4.50
		Benchmark 4	28	1,300	0	28	12.95	4.87
	5	Benchmark 1	28	1,192	1	26	10.71	3.86
		Benchmark 2	28	1,253	0	26	12.98	4.03
		Benchmark 3	28	710	0	25	13.34	4.32
		Benchmark 4	28	1,266	0	27	13.84	5.10
	6	Benchmark 1	28	1,213	1	23	9.85	3.27
		Benchmark 2	28	1,218	0	24	10.99	3.87
		Benchmark 3	28	909	1	28	12.93	4.49
		Benchmark 4	28	1,210	1	26	11.46	4.42
	7	Benchmark 1	28	1,124	0	22	9.71	3.77
		Benchmark 2	28	1,076	0	25	10.85	3.85
		Benchmark 3	28	908	0	27	12.36	4.74
		Benchmark 4	28	1,155	0	27	11.19	4.94

In general, the mean Benchmark raw scores are low, with only grade 6 ELA students (form 4), grade 7 ELA students (form 3), and grade 3 math students (forms 3 and 4) having a mean score that is greater than 50% correct. For the most part, the mean raw scores do increase somewhat from Benchmark 1 to Benchmark 4.

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

However, it is important to keep in mind that while the Benchmarks have been designed to be comparable in content, item type, and standards coverage across forms, they have not been statistically equated and thus may vary in difficulty from form to form. Additionally, they have not been statistically linked to or evaluated against state summative test results in terms of scores or levels of proficiency.

To address the potential variability in Benchmark form difficulty, Benchmark Z-scores were calculated from raw scores using only the MC dichotomously scored (0 or 1) questions. These are provided, along with the final sample sizes for this study, in Table 20. The final sample for this analysis included only those students for whom there was complete data: PSSA results for both 2016 and 2017 as well as scores for the fall, first winter, and spring Benchmarks for ELA and math. The data show an increase in average Benchmark Z-scores from fall to winter for all subjects. Because the PSSA scores are not vertically scaled, and each grade's Proficient cut point is fixed at 1000, it is not appropriate to compare PSSA scores from year to year.

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

Table 20. Sample Sizes for Benchmark Data Analysis

Subject	Grade	Assessment	N	Min.	Max.	Mean	SD
ELA	4	Benchmark 1	950	-2.18	3.13	0.05	1.00
		Benchmark 2	950	-2.11	2.79	0.04	1.01
		Benchmark 4	950	-1.75	3.10	0.06	1.00
		PSSA Scaled Score (2016)	950	728.00	1233.00	944.55	89.02
		PSSA Scaled Score (2017)	950	630.00	1225.00	942.47	87.78
	5	Benchmark 1	961	-2.10	3.11	0.05	1.00
		Benchmark 2	961	-2.35	2.70	0.06	1.00
		Benchmark 4	961	-2.41	3.01	0.05	1.01
		PSSA Scaled Score (2016)	961	720.00	1214.00	928.08	93.38
		PSSA Scaled Score (2017)	961	718.00	1237.00	947.20	84.17
	6	Benchmark 1	867	-2.58	3.53	0.06	1.02
		Benchmark 2	867	-2.26	2.71	0.07	0.99
		Benchmark 4	867	-2.16	2.14	0.03	0.99
		PSSA Scaled Score (2016)	867	702.00	1299.00	946.45	88.06
		PSSA Scaled Score (2017)	867	751.00	1215.00	946.09	76.74
	7	Benchmark 1	854	-2.06	2.52	0.10	0.97
		Benchmark 2	854	-2.12	2.83	0.10	0.98
		Benchmark 4	854	-2.23	2.84	0.06	1.02
		PSSA Scaled Score (2016)	854	695.00	1211.00	941.37	87.50
		PSSA Scaled Score (2017)	854	687.00	1249.00	947.66	88.28
Math	4	Benchmark 1	1,005	-2.75	4.58	0.06	1.01
		Benchmark 2	1,005	-2.87	3.64	0.05	1.00
		Benchmark 4	1,005	-2.45	3.09	0.07	1.00
		PSSA Scaled Score (2016)	1,005	682.00	1278.00	924.25	101.63
		PSSA Scaled Score (2017)	1,005	734.00	1265.00	908.76	83.87
	5	Benchmark 1	1,009	-2.51	3.96	0.03	1.00
		Benchmark 2	1,009	-2.72	3.23	0.05	0.99
		Benchmark 4	1,009	-2.71	2.58	0.08	0.99
		PSSA Scaled Score (2016)	1,009	647.00	1248.00	897.37	86.49
		PSSA Scaled Score (2017)	1,009	721.00	1240.00	909.41	75.60
	6	Benchmark 1	950	-2.71	4.03	0.04	1.00
		Benchmark 2	950	-2.84	3.10	0.01	0.98
		Benchmark 4	950	-2.37	3.29	-0.00	1.01
		PSSA Scaled Score (2016)	950	730.00	1258.00	909.01	79.14
		PSSA Scaled Score (2017)	950	715.00	1215.00	887.55	75.99
	7	Benchmark 1	870	-2.58	3.26	0.03	1.00
		Benchmark 2	870	-2.56	3.67	0.04	1.00
		Benchmark 4	870	-2.27	3.20	0.06	1.01
		PSSA Scaled Score (2016)	870	675.00	1207.00	868.46	83.68
		PSSA Scaled Score (2017)	870	694.00	1316.00	887.51	88.14

Note. Study Island Benchmark scores were transformed to Z score scale for comparison.

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

Research Question 5: Is there a significant relationship between student scores on Study Island Benchmarks and their scores on summative, end-of-year PSSA state tests? If so, does the significant relationship between Study Island Benchmark scores and PSSA scores remain after accounting for a student's previous PSSA performance?

When the alignment of learning standards and assessments is sound, then there is a greater likelihood that one test score may predict another. The relationship between the two test scores can be called *predictive* or *criterion validity*. *Predictive validity* can be investigated by calculating the correlation coefficient between the results of the assessment and the subsequent targeted outcome. The stronger the correlation between the assessment data and the targeted outcome, the greater the degree of predictive validity the assessment possesses. Furthermore, when a correlation is statistically significant at the .05 level or lower, the probability of obtaining such a correlation coefficient by chance would occur fewer than five times out of 100, giving us confidence that a relationship between the two test scores does exist.

The correlations between the Benchmark test scores and the PSSA scores provide evidence of the predictive validity of Study Island Benchmarks to the PSSA scores. Correlation coefficients range from 0 to +/-1 and are interpreted such that the larger the correlation coefficient, the stronger the association between the two assessments. The interpretation is that the highly correlated assessments are likely measuring similar constructs or have what Messick (1989) referred to as convergent validity; one may predict the other.

As with any statistic, there are assumptions about the data to consider before trusting the correlations. Specifically, the data should be normally distributed, linear, and homoscedastic (the errors are random, and variances are similar across variables). When these assumptions are violated, the correlation may become inadequate to explain a given relationship. In this study, only the 2016 PSSA scores for grade 7 ELA were normally distributed (see Appendix D for a table displaying the results of all tests for normal distributions of the PSSA and Study Island Benchmark scores as well as histograms for visual representation). Therefore, the Spearman rank correlation coefficients are provided. The Spearman rho is a nonparametric statistic that does not require normally distributed data and is interpreted in a similar way to other types of correlations.

Table 21 provides the Spearman rho correlations between the Study Island Benchmark Z scores and the PSSA test scores by grade level. (Scatterplots of these correlations are provided in Appendix E.) All correlations are statistically significant at the 0.001 level. This indicates that there is a strong enough association that one can infer that the two assessments are measuring similar constructs and that performance on one can be predictive of performance on the other.

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

Table 21. Correlation Between Scores on PSSA and Study Island Benchmarks by Grade and Subject

Subject	Grade	Score	Benchmark 1	Benchmark 2	Benchmark 4	PSSA 2016	PSSA 2017
ELA	4	Benchmark 1	1				
		Benchmark 2	0.630***	1			
		Benchmark 4	0.626***	0.654***	1		
		PSSA 2016	0.683***	0.692***	0.697***	1	
		PSSA 2017	0.704***	0.746***	0.734***	0.795***	1
	5	Benchmark 1	1				
		Benchmark 2	0.710***	1			
		Benchmark 4	0.645***	0.666***	1		
		PSSA 2016	0.757***	0.730***	0.690***	1	
		PSSA 2017	0.738***	0.751***	0.726***	0.813***	1
	6	Benchmark 1	1				
		Benchmark 2	0.602***	1			
		Benchmark 4	0.603***	0.701***	1		
		PSSA 2016	0.642***	0.724***	0.694***	1	
		PSSA 2017	0.663***	0.718***	0.743***	0.807***	1
	7	Benchmark 1	1				
		Benchmark 2	0.704***	1			
		Benchmark 4	0.623***	0.601***	1		
		PSSA 2016	0.767***	0.747***	0.627***	1	
		PSSA 2017	0.753***	0.742***	0.651***	0.836***	1
Math	4	Benchmark 1	1				
		Benchmark 2	0.498***	1			
		Benchmark 4	0.492***	0.587***	1		
		PSSA 2016	0.571***	0.653***	0.641***	1	
		PSSA 2017	0.531***	0.673***	0.738***	0.783***	1
	5	Benchmark 1	1				
		Benchmark 2	0.574***	1			
		Benchmark 4	0.501***	0.602***	1		
		PSSA 2016	0.567***	0.568***	0.636***	1	
		PSSA 2017	0.532***	0.584***	0.705***	0.762***	1
	6	Benchmark 1	1				
		Benchmark 2	0.425***	1			
		Benchmark 4	0.420***	0.544***	1		
		PSSA 2016	0.502***	0.566***	0.587***	1	
		PSSA 2017	0.496***	0.568***	0.687***	0.713***	1
	7	Benchmark 1	1				
		Benchmark 2	0.532***	1			
		Benchmark 4	0.521***	0.577***	1		
		PSSA 2016	0.593***	0.642***	0.672***	1	
		PSSA 2017	0.579***	0.644***	0.728***	0.780***	1

=significant at the .05 level **=significant at the .01 level *=significant at the .001 level*

To understand the magnitude of the association, Cohen, Cohen, West, and Aiken (2003) provided a standard or rule of thumb for interpreting the strength of the relationship. Correlation coefficients between 0.10 and 0.29

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

represent a small association, coefficients between 0.30 and 0.49 represent a medium association, and coefficients of 0.50 and above represent a large association or relationship. As Table 21 shows, there is a large, positive, and significant correlation between students' performance on Study Island Benchmarks and their performance on the PSSA in all grades and subjects.

As with the investigation into the impact of Study Island practice, where differences in scores were evaluated after controlling for ability via propensity score matching, it is important to similarly control for ability when evaluating the strength of these score correlations. In the practice analyses, categorical variables were used (Study Island User and Study Island Non-User, Study Island High User and Study Island Low User), allowing for the comparison of treatment (Study Island User and Study Island High User) and pseudo-control groups (Study Island Non-User and Study Island Low User). Given the continuous nature of the Benchmark assessments, partial correlations were used to determine if Benchmark scores are correlated with the 2017 PSSA scores. The partial correlation method allows for the removal of the 2016 PSSA scores' influence on the correlation between scores– in other words, teasing out ability. The 2016 PSSA scores were treated as the mediating or controlling variable to investigate the bivariate correlations between the two benchmark scores and the 2017 PSSA score.

After controlling for prior ability with the partial correlations, significant small and medium-sized correlations remain between use of Study Island Benchmarks and 2017 PSSA scores, as shown in Table 22. All values are significant at the .001 level. This indicates that using Benchmarks is significantly related to PSSA scores above and beyond student ability, suggesting that the opportunity to practice items on the Benchmarks that are similar to what students see on the PSSA assists students in PSSA performance.

Table 22. Correlations Between Scores on PSSA 2017 and Study Island Benchmarks After Accounting for PSSA 2016

Test Name	Benchmark 1	Benchmark 2	Benchmark 4
2017 PSSA ELA Grade 4	0.364***	0.449***	0.414***
2017 PSSA ELA Grade 5	0.323***	0.396***	0.391***
2017 PSSA ELA Grade 6	0.318***	0.328***	0.429***
2017 PSSA ELA Grade 7	0.317***	0.322***	0.298***
2017 PSSA Math Grade 4	0.164***	0.342***	0.494***
2017 PSSA Math Grade 5	0.187***	0.284***	0.442***
2017 PSSA Math Grade 6	0.228***	0.286***	0.473***
2017 PSSA Math Grade 7	0.230***	0.297***	0.441***
=significant at the .05 level **=significant at the .01 level *=significant at the .001 level*			

Conclusions

The findings in this study suggest there are discernable and statistically significant positive impacts on PSSA scores for students participating in Study Island practice and Benchmarks. Generally, implementation and use of Study Island practice and Benchmarks in RSD vary by grade and content area, with grade 5 exhibiting the strongest usage. Some groups of students appear to be answering relatively few practice questions and spending minimal time over the course of the year, while other groups have a stronger implementation. Where students spend more time, answer more questions, and spread their time over active weeks, positive differences are observed. This is particularly evident in math in grades 6 and 7 between users and non-users and in grades 5 and 7 between high and low users, with significant differences in mean scale scores and proficiency classification. Grade 5 ELA, used among large swaths of the student population at high levels, also shows significant differences in mean PSSA scores when comparing users to non-users. In addition, when students are exposed to the Benchmarks, there is a strong and significant association between scores on the Benchmarks and scores on the PSSA. These statistically significant observations remain even after controlling for student ability, based on their prior-year PSSA scores.

These analyses are clearly impacted by the quality and approach by which schools use Study Island practice or Benchmarks. 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.

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

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.

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.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

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.

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

Jinkins, 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.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education/Macmillan.

National Center for Education Statistics. (n.d.). *Common core of data*. Retrieved from https://nces.ed.gov/ccd/

Organisation for Economic Cooperation and Development. (2012). *Equity and quality in education: Supporting disadvantaged students and schools*. Retrieved from http://dx.doi.org/10.1787/9789264130852-en

Pennsylvania Department of Education. (n.d.-a). *Enrollment reports and projections*. Retrieved from http://www.education.pa.gov/Data-and-Statistics/Pages/Enrollment%20Reports%20and%20Projections.aspx

Pennsylvania Department of Education. (n.d.-b). *Reading SD*. Retrieved from http://paschoolperformance.org/Profile/378

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.

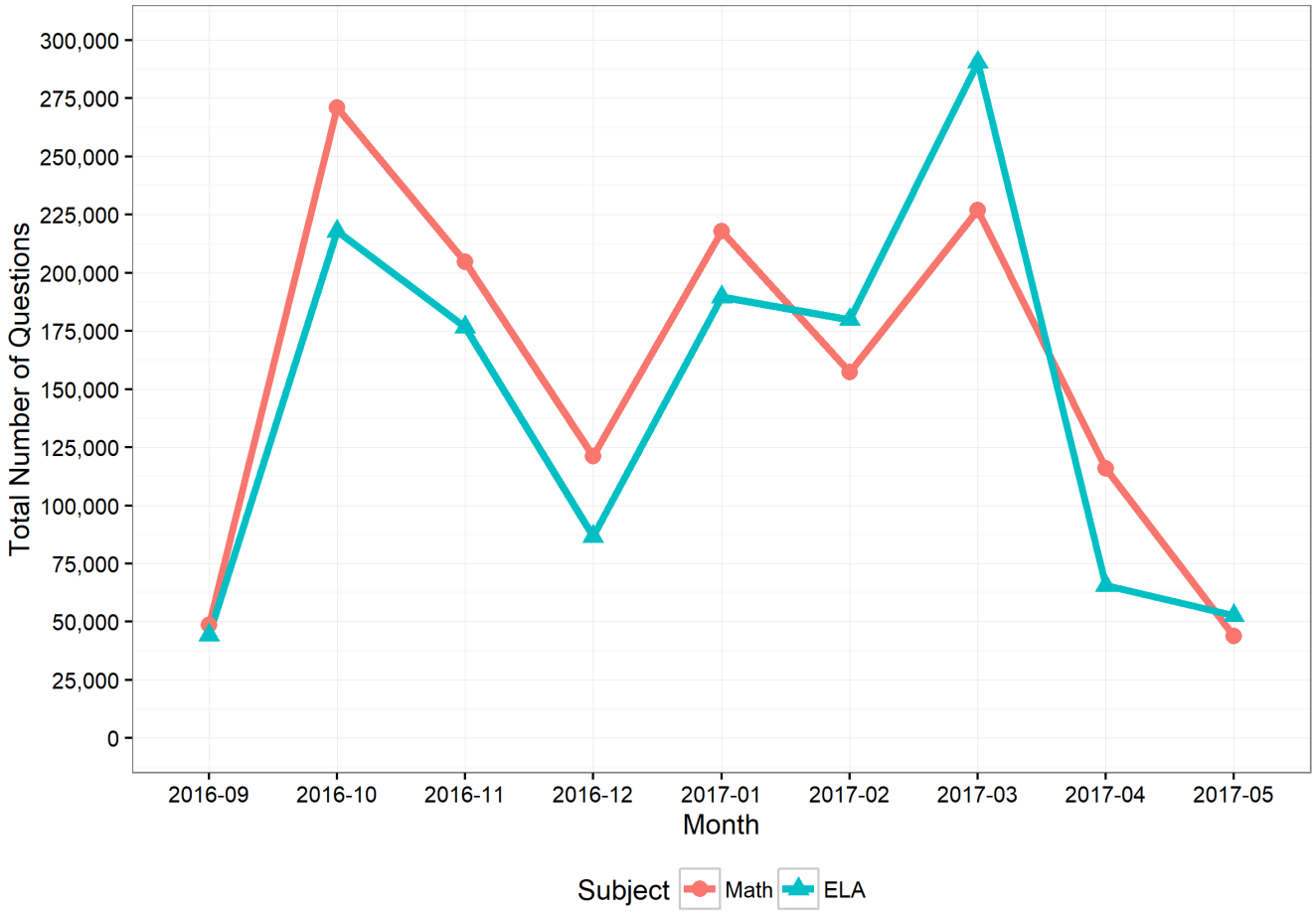
Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21,* 361–375.

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.

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

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



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

Appendix B: Volume of Benchmark Test Use

Table B1. Volume and Date of Benchmark Test Use

Subject	Grade	Form	N	First Test Date	Last Test Date
ELA	3	1	1,203	2016-09-01	2016-10-04
		2	1,249	2016-12-01	2017-01-18
		3	555	2017-02-02	2017-03-31
		4	1,270	2017-05-02	2017-05-31
	4	1	1,199	2016-09-06	2016-10-07
		2	1,208	2016-12-01	2017-01-13
		3	454	2017-02-02	2017-03-31
		4	1,281	2017-05-05	2017-05-31
	5	1	1,174	2016-09-01	2016-10-07
		2	1,228	2016-12-01	2017-01-11
		3	708	2017-02-02	2017-03-31
		4	1,249	2017-05-02	2017-05-31
	6	1	1,185	2016-09-01	2016-10-07
		2	1,221	2016-12-02	2016-12-23
		3	923	2017-02-27	2017-03-30
		4	1,137	2017-05-02	2017-05-30
	7	1	1,127	2016-09-06	2016-10-06
		2	1,138	2016-12-01	2016-12-22
		3	918	2017-02-24	2017-03-31
		4	1,105	2017-05-03	2017-05-30
Math	3	1	1,238	2016-09-01	2016-09-30
		2	1,262	2016-12-01	2017-01-13
		3	588	2017-02-02	2017-03-31
		4	1,282	2017-05-02	2017-05-31
	4	1	1,224	2016-09-06	2016-10-20
		2	1,247	2016-12-01	2017-01-13
		3	461	2017-02-03	2017-03-31
		4	1,300	2017-05-02	2017-05-30
	5	1	1,192	2016-09-01	2016-10-06
		2	1,253	2016-12-01	2017-01-09
		3	710	2017-02-02	2017-03-31
		4	1,266	2017-05-02	2017-05-31
	6	1	1,213	2016-09-06	2016-10-07
		2	1,218	2016-12-01	2016-12-23
		3	909	2017-02-28	2017-03-31
		4	1,210	2017-05-02	2017-05-30
	7	1	1,124	2016-09-09	2016-10-07
		2	1,076	2016-12-05	2016-12-22
		3	908	2017-02-28	2017-03-31
		4	1,155	2017-05-03	2017-05-31

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

Appendix C: Propensity Score Matching

Figure C1. ELA Grade 4 User vs Non-User

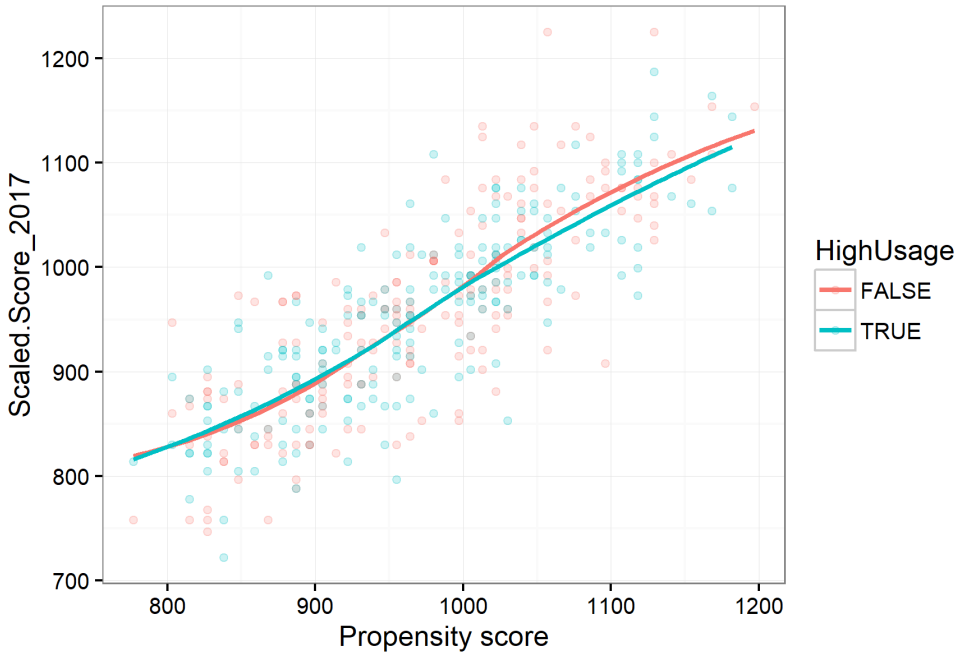
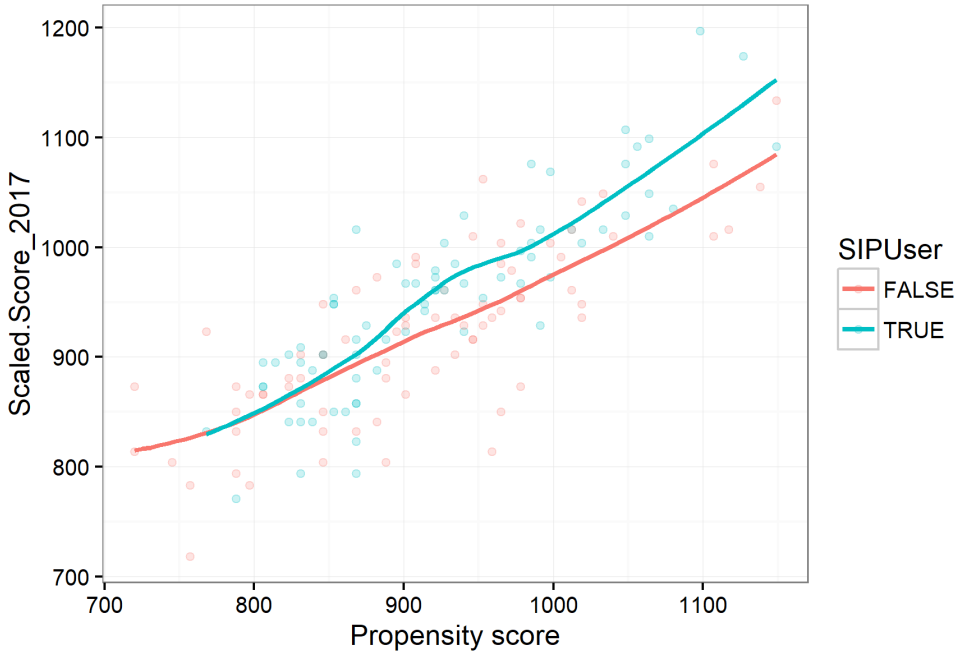


Figure C2. ELA Grade 5 User vs Non-User



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

Figure C3. ELA Grade 6 User vs Non-User

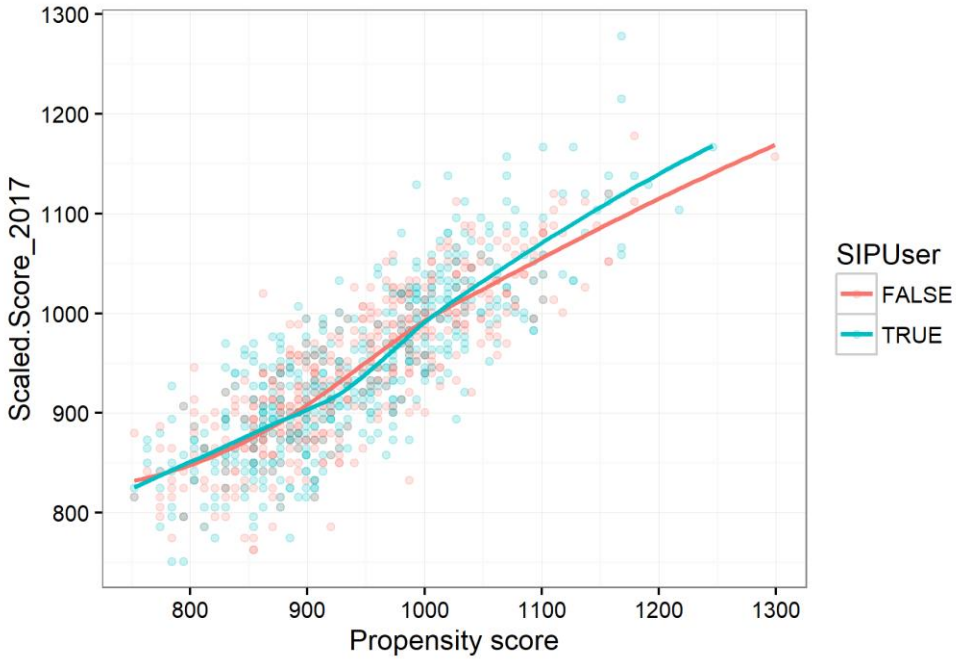
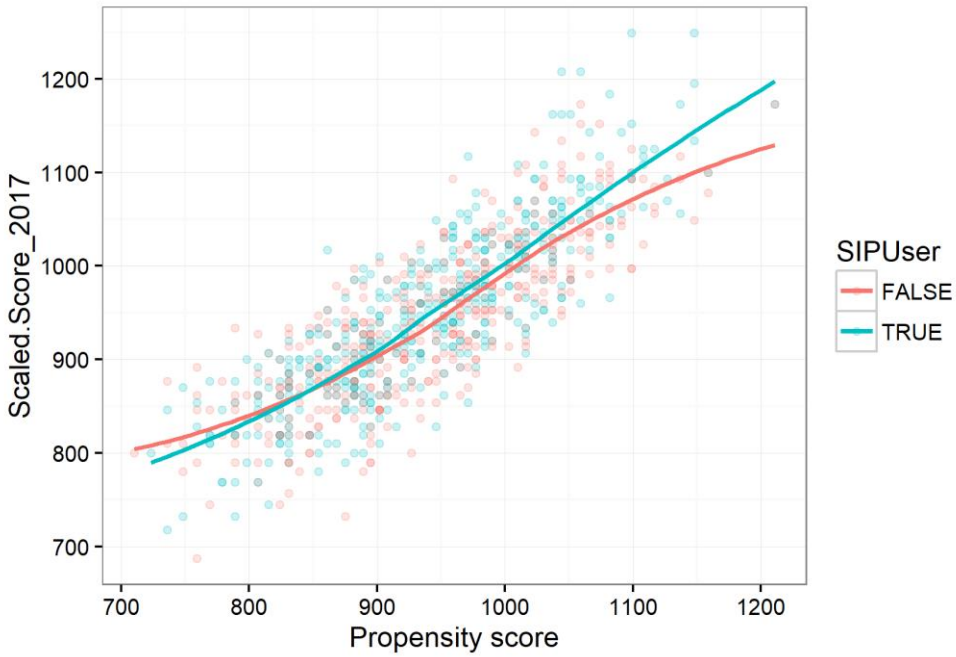


Figure C4. ELA Grade 7 User vs Non-User



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

Figure C5. Math Grade 4 User vs Non-User

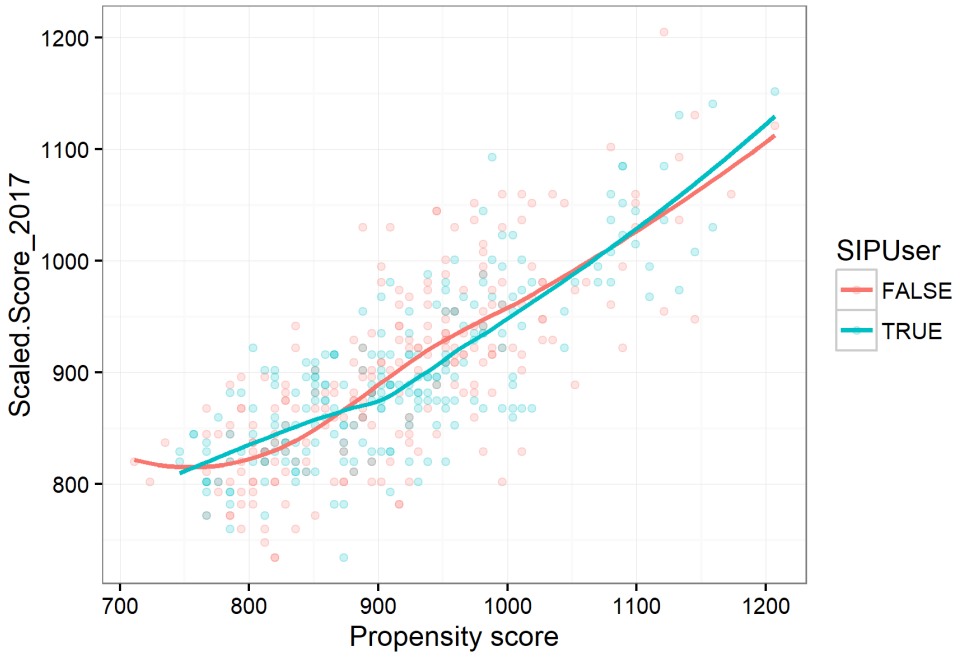
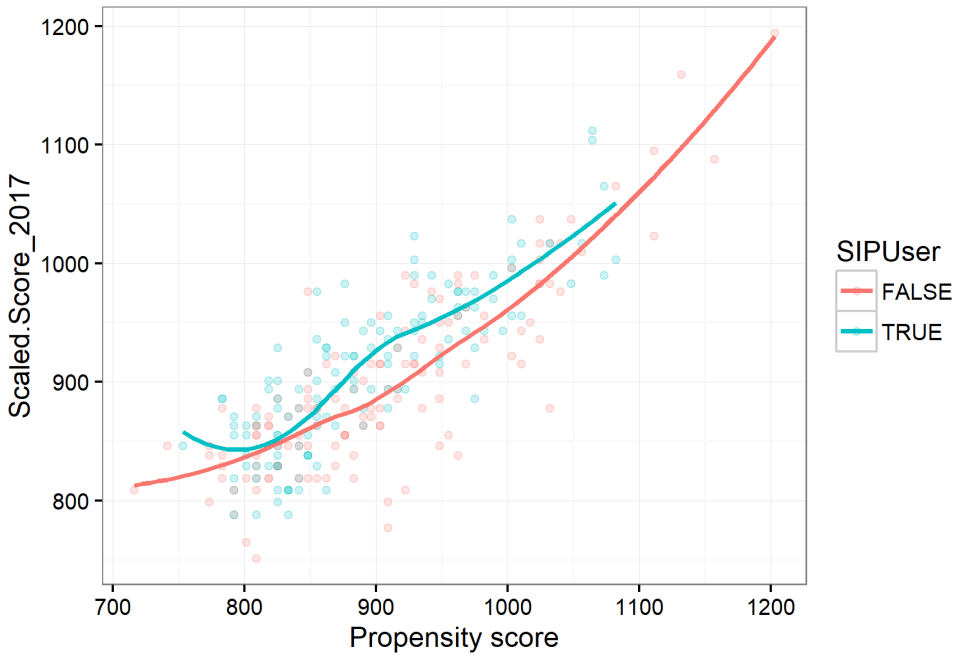


Figure C6. Math Grade 5 User vs Non-User



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

Figure C7. Math Grade 6 User vs Non-User

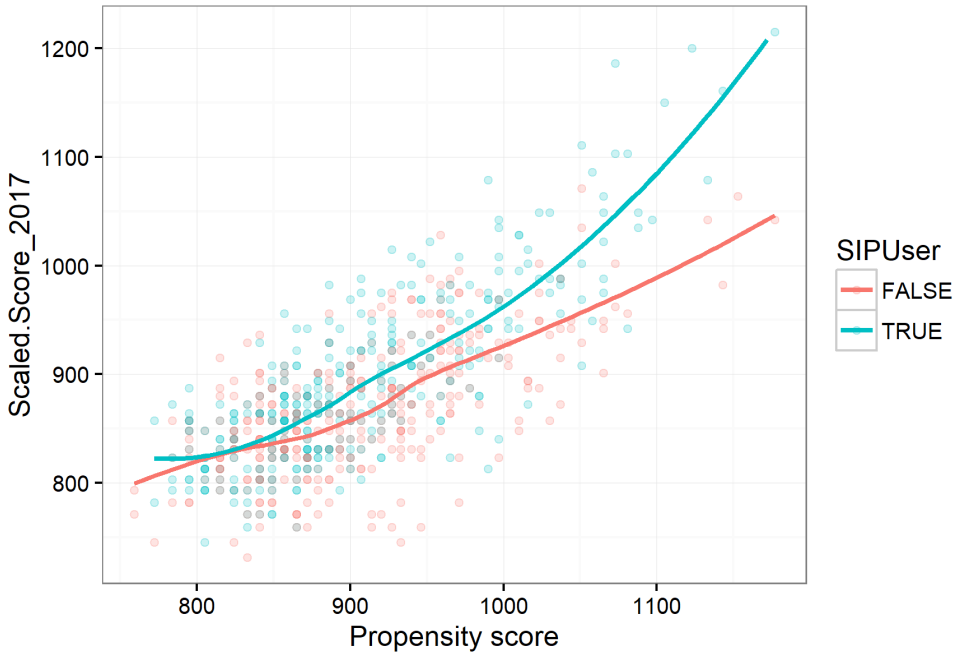
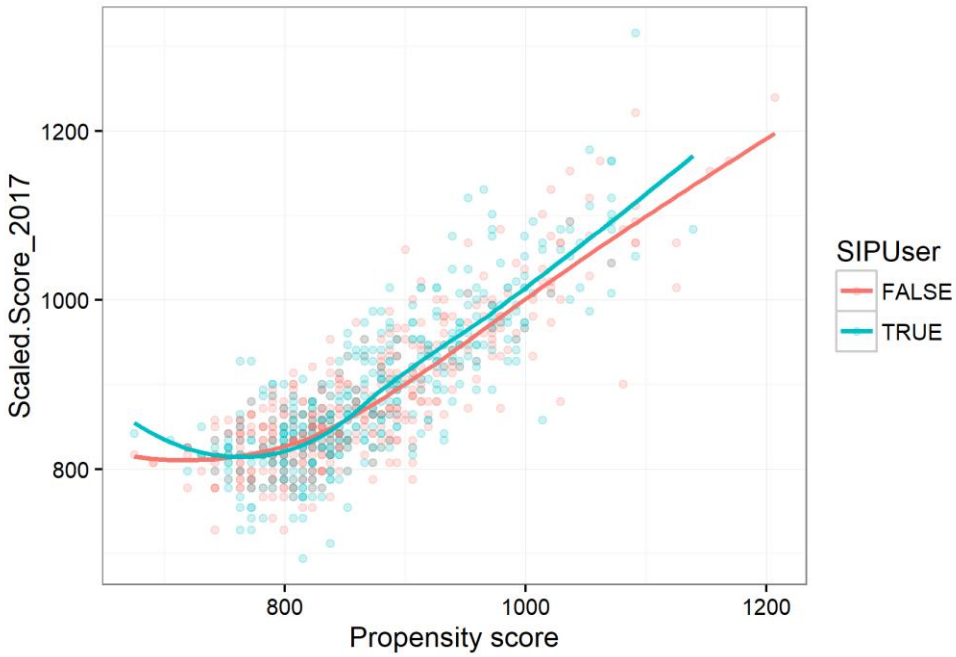


Figure C8. Math Grade 7 User vs Non-User



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

Figure C9. ELA Grade 4 High Usage vs Low Usage

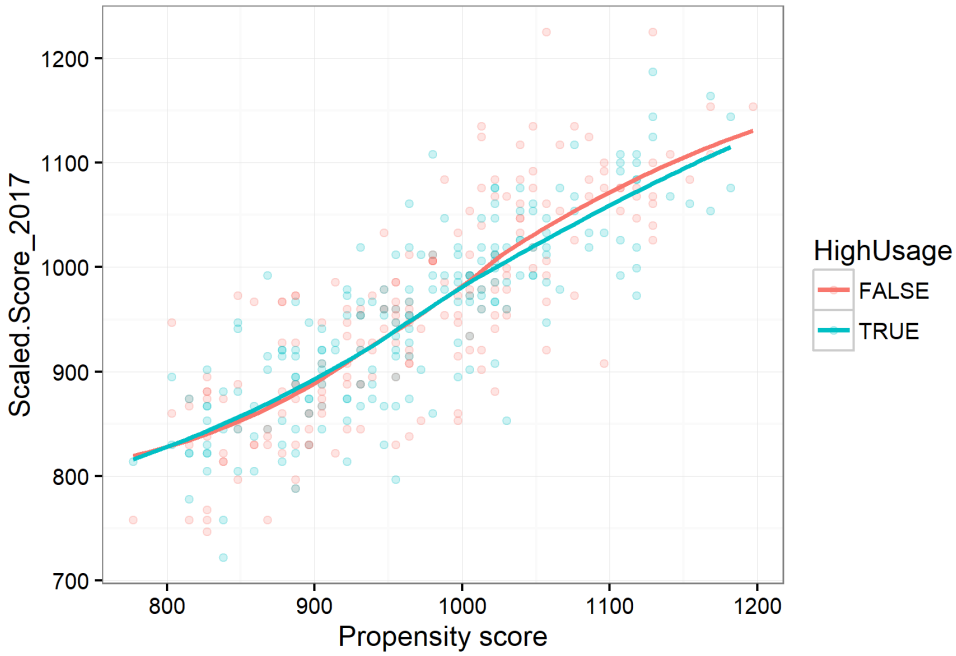
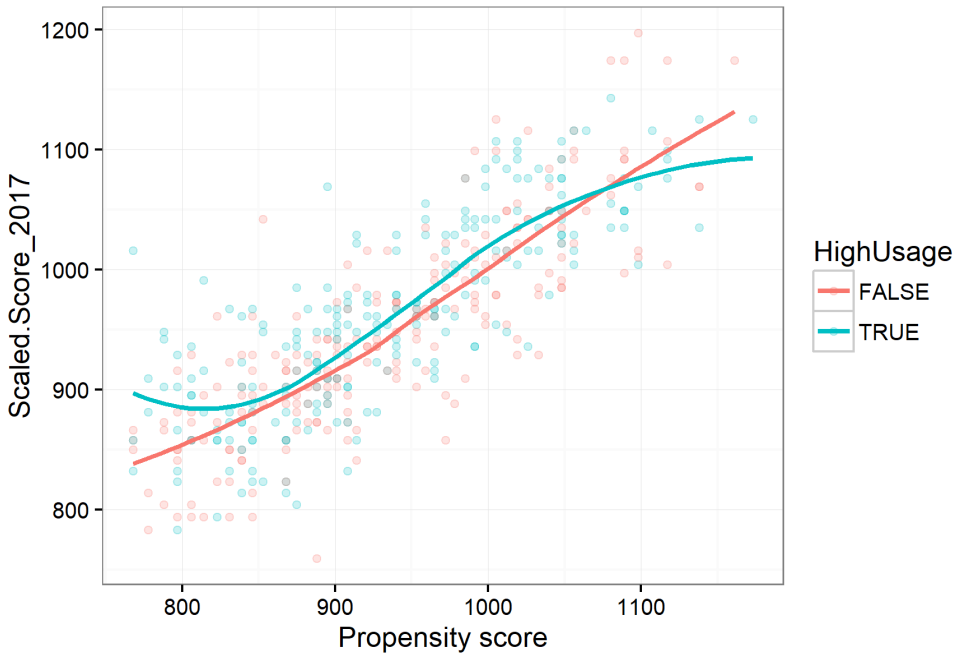


Figure C10. ELA Grade 5 High Usage vs Low Usage



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

Figure C11. ELA Grade 6 High Usage vs Low Usage

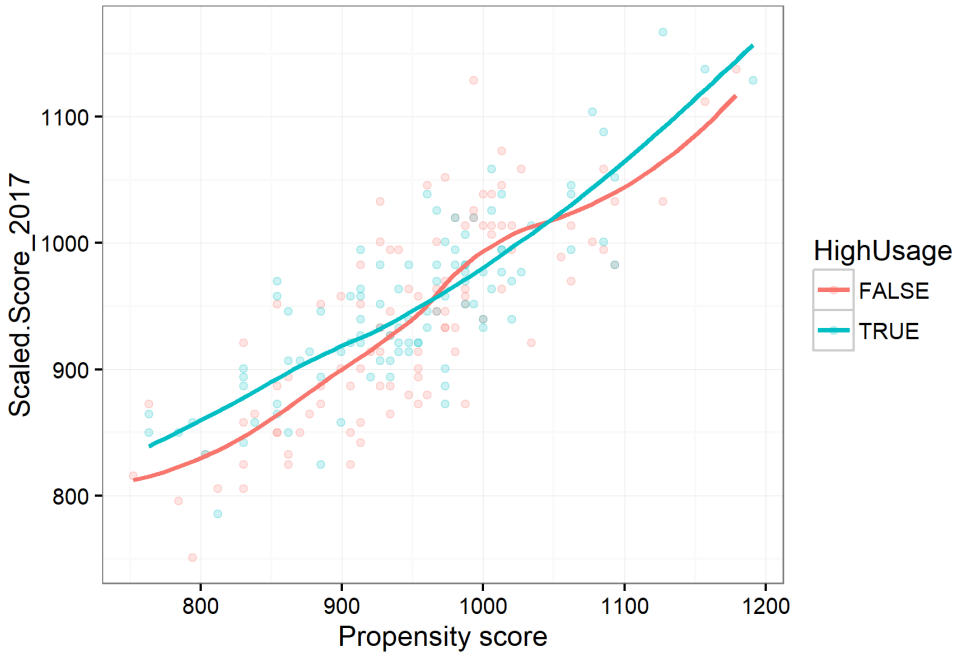
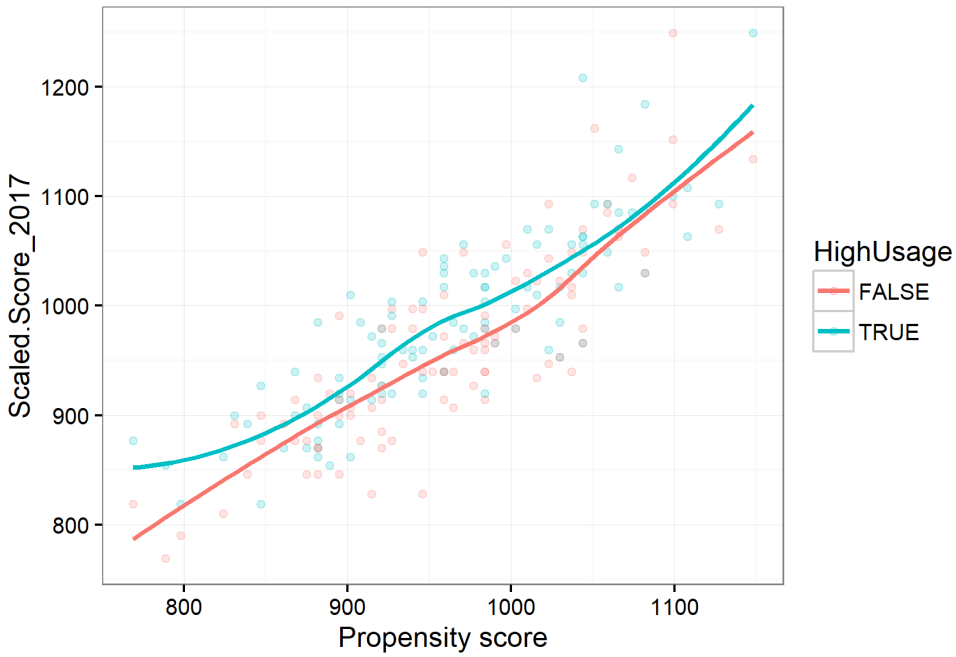


Figure C12. ELA Grade 7 High Usage vs Low Usage



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

Figure C13. Math Grade 4 High Usage vs Low Usage

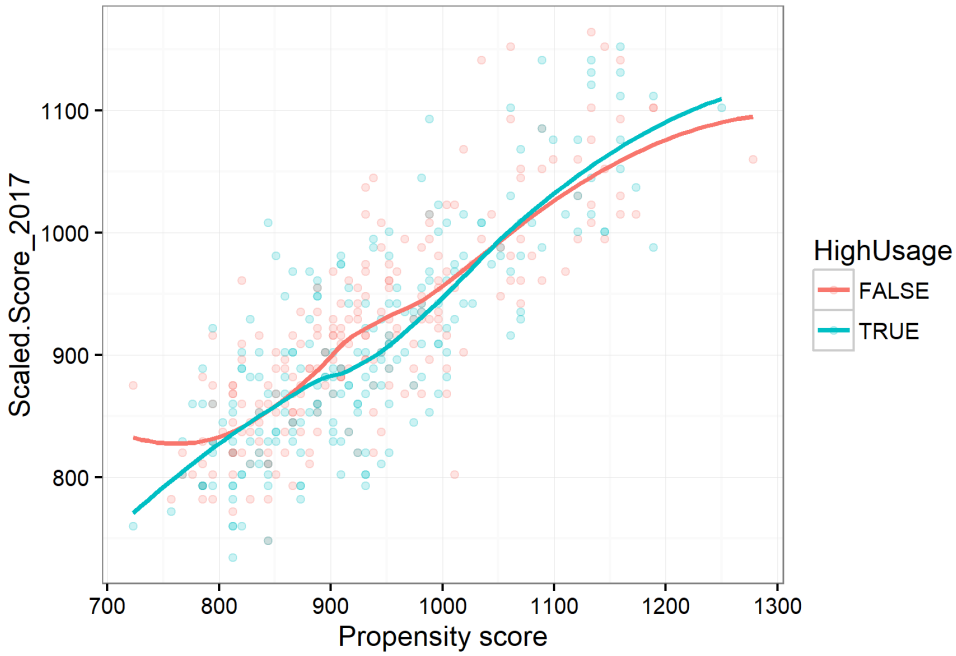
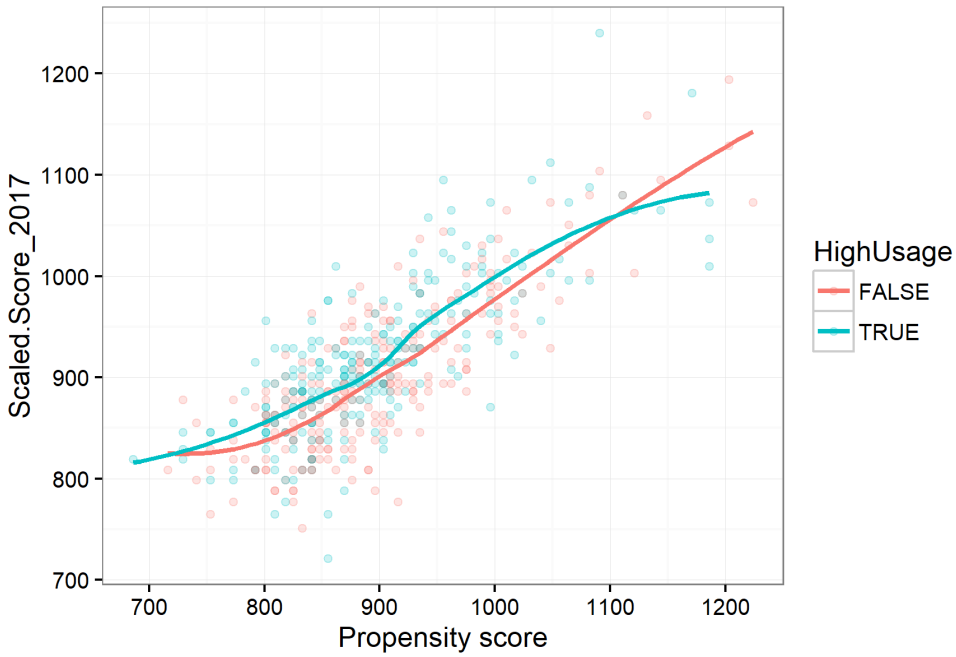


Figure C14. Math Grade 5 High Usage vs Low Usage



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

Figure C15. Math Grade 6 High Usage vs Low Usage

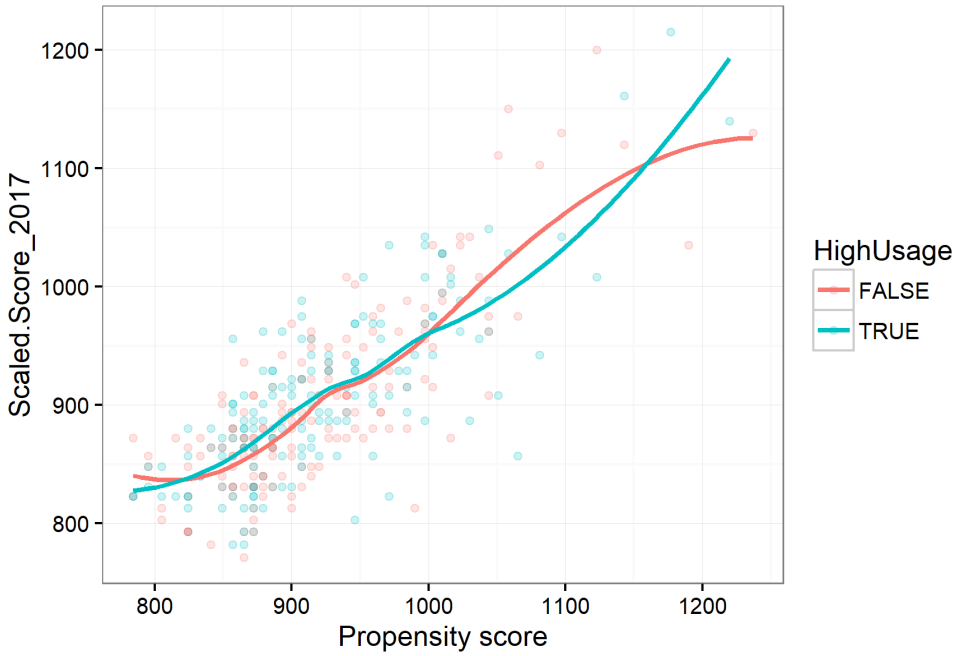
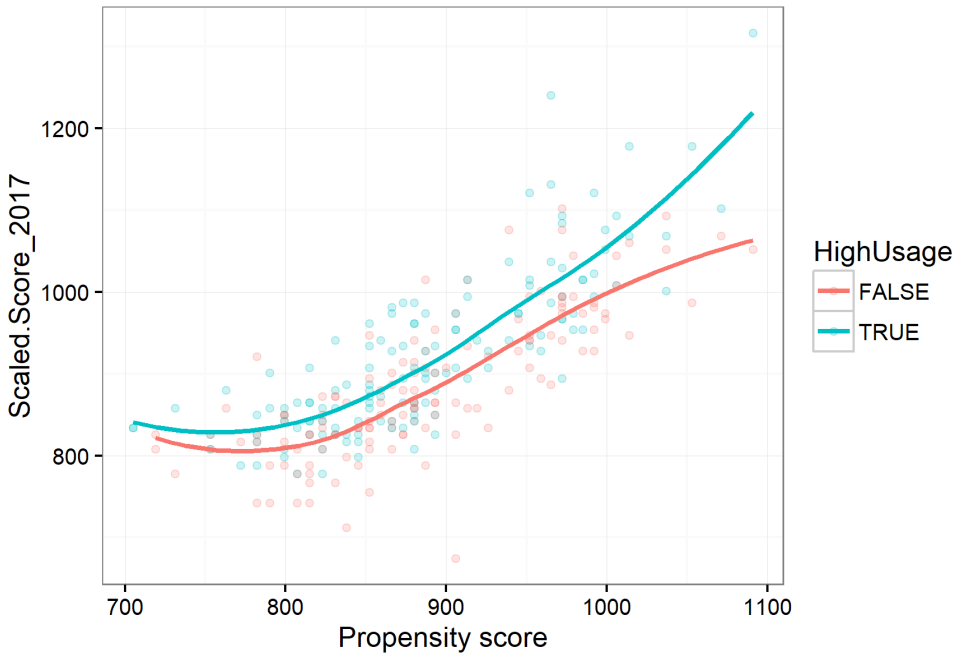


Figure C16. Math Grade 7 High Usage vs Low Usage



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

Appendix D: Test for Normal Distribution of Scores

Table D1. Normal Distribution

			Shapiro-Wilk	
Subject	**Grade**	**Assessment**	**Statistic**	**Sig.**
ELA	4	PSSA Scaled Score 2016	0.993	0.000
		PSSA Scaled Score 2017	0.995	0.003
		Z Score: PSSA ELA Benchmark 1	0.953	0.000
		Z Score: PSSA ELA Benchmark 2	0.957	0.000
		Z Score: PSSA ELA Benchmark 4	0.961	0.000
	5	PSSA Scaled Score 2016	0.986	0.000
		PSSA Scaled Score 2017	0.989	0.000
		Z Score: PSSA ELA Benchmark 1	0.969	0.000
		Z Score: PSSA ELA Benchmark 2	0.974	0.000
		Z Score: PSSA ELA Benchmark 4	0.967	0.000
	6	PSSA Scaled Score 2016	0.990	0.000
		PSSA Scaled Score 2017	0.994	0.002
		Z Score: PSSA ELA Benchmark 1	0.977	0.000
		Z Score: PSSA ELA Benchmark 2	0.978	0.000
		Z Score: PSSA ELA Benchmark 4	0.972	0.000
	7	PSSA Scaled Score 2016	0.997	0.149
		PSSA Scaled Score 2017	0.995	0.009
		Z Score: PSSA ELA Benchmark 1	0.982	0.000
		Z Score: PSSA ELA Benchmark 2	0.982	0.000
		Z Score: PSSA ELA Benchmark 4	0.964	0.000
Math	4	PSSA Scaled Score 2016	0.974	0.000
		PSSA Scaled Score 2017	0.972	0.000
		Z Score: PSSA Math Benchmark 1	0.981	0.000
		Z Score: PSSA Math Benchmark 2	0.988	0.000
		Z Score: PSSA Math Benchmark 4	0.982	0.000
	5	PSSA Scaled Score 2016	0.962	0.000
		PSSA Scaled Score 2017	0.960	0.000
		Z Score: PSSA Math Benchmark 1	0.983	0.000
		Z Score: PSSA Math Benchmark 2	0.991	0.000
		Z Score: PSSA Math Benchmark 4	0.988	0.000
	6	PSSA Scaled Score 2016	0.960	0.000
		PSSA Scaled Score 2017	0.949	0.000
		Z Score: PSSA Math Benchmark 1	0.986	0.000
		Z Score: PSSA Math Benchmark 2	0.990	0.000
		Z Score: PSSA Math Benchmark 4	0.976	0.000
	7	PSSA Scaled Score 2016	0.963	0.000
		PSSA Scaled Score 2017	0.926	0.000
		Z Score: PSSA Math Benchmark 1	0.988	0.000
		Z Score: PSSA Math Benchmark 2	0.977	0.000
		Z Score: PSSA Math Benchmark 4	0.961	0.000

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

Figure D1. ELA and Math, Study Island Benchmark 1

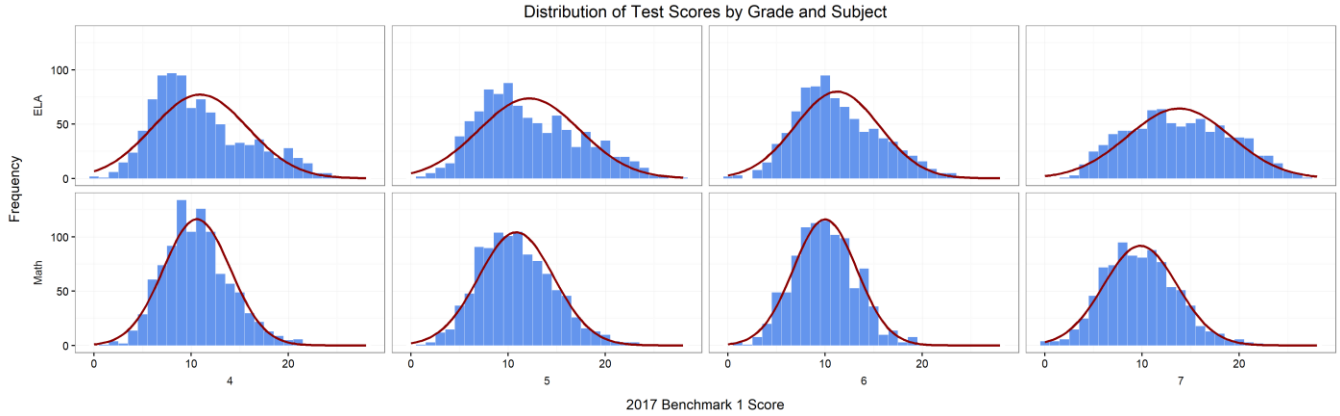


Figure D2. ELA and Math, Study Island Benchmark 2

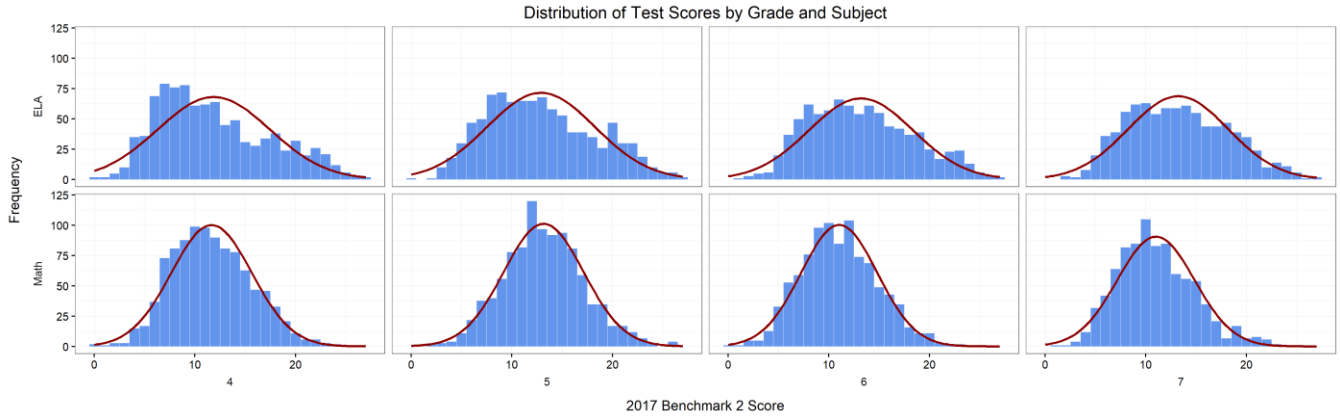
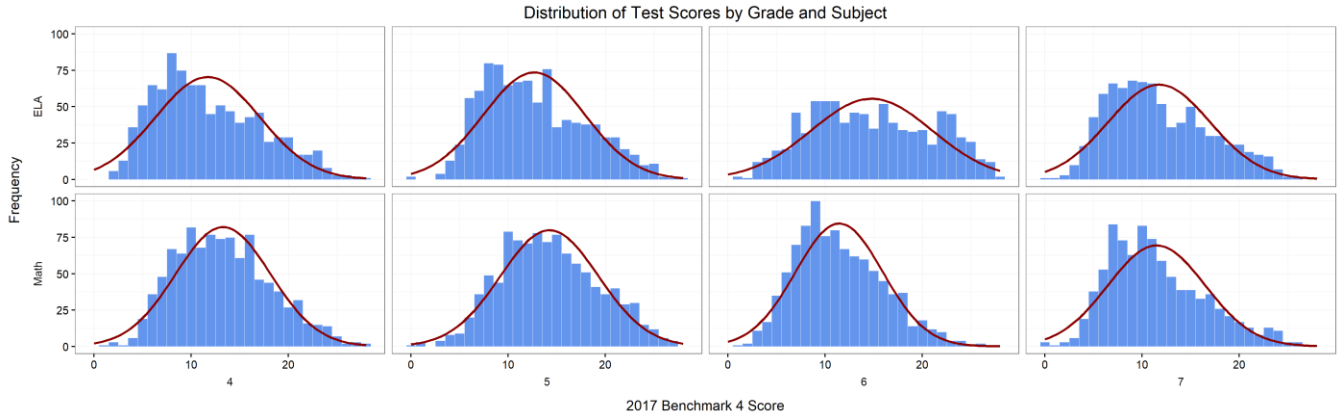


Figure D3. ELA and Math, Study Island Benchmark 4



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

Figure D4. ELA and Math, PSSA 2016

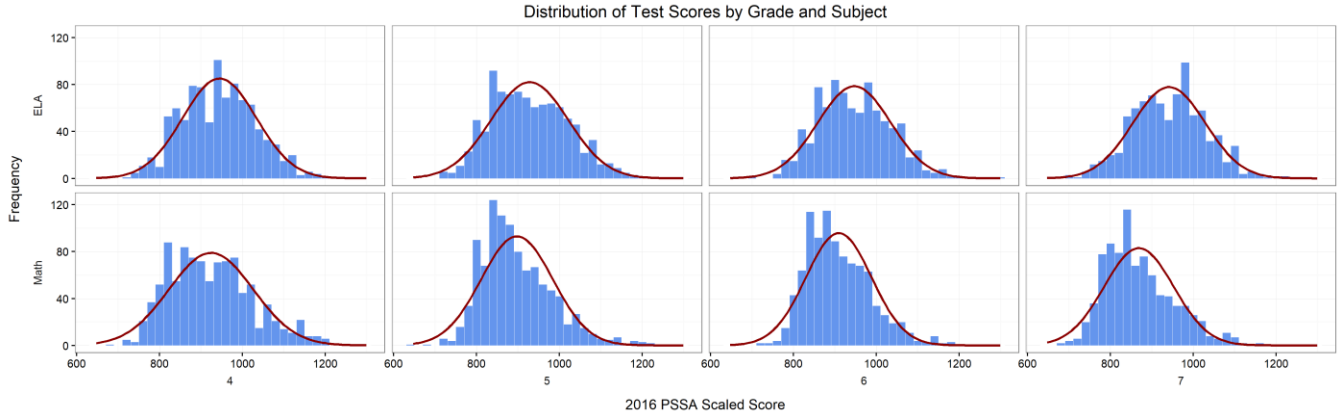
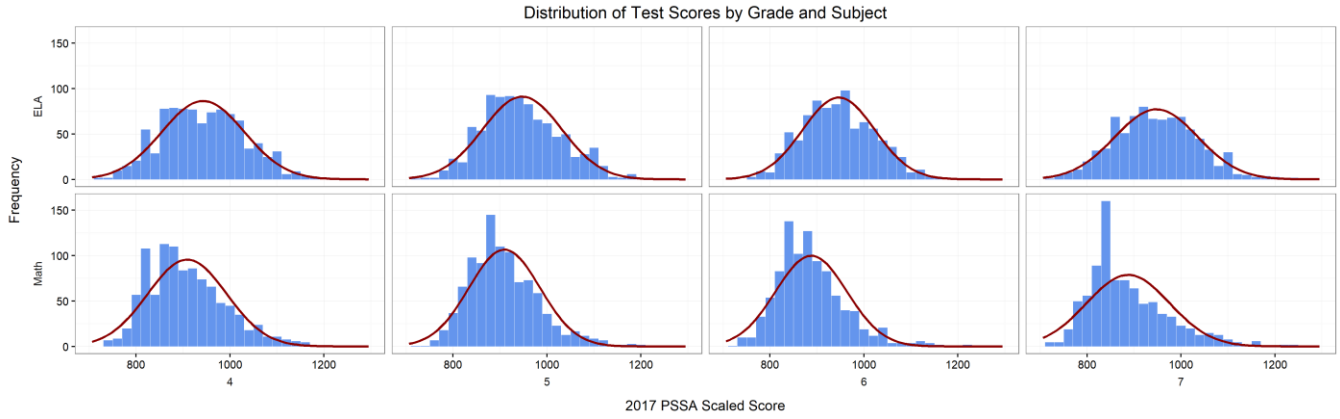


Figure D5. ELA and Math, PSSA 2017



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

Appendix E: Scatterplots Showing Correlations Between Benchmark Scores and PSSA Scores

Figure E1. Relationship Between Benchmark 1 and 2016 PSSA

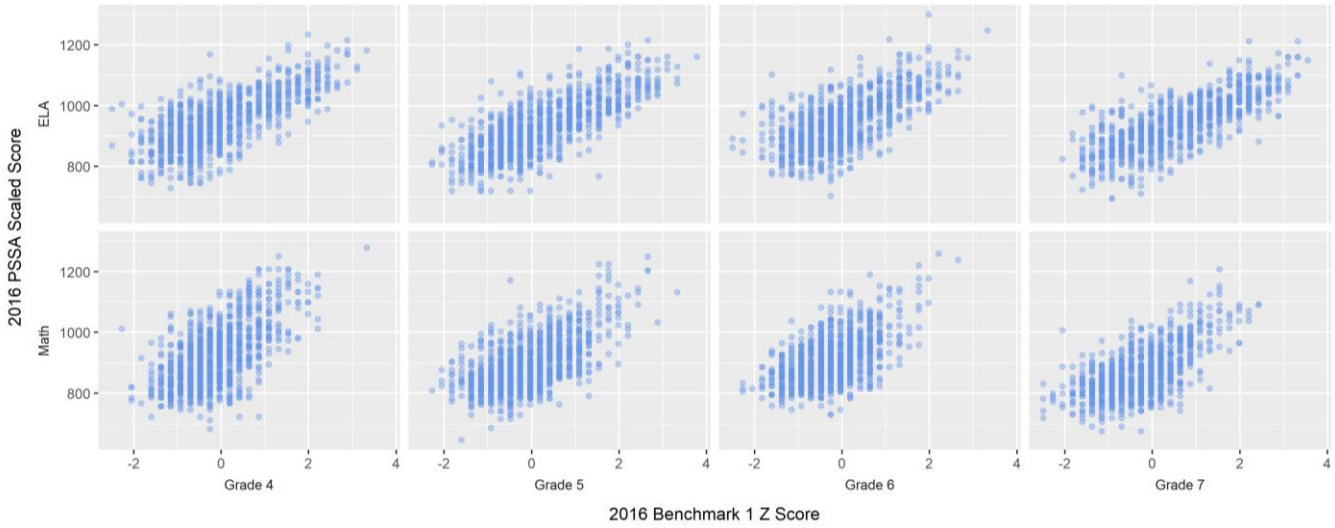
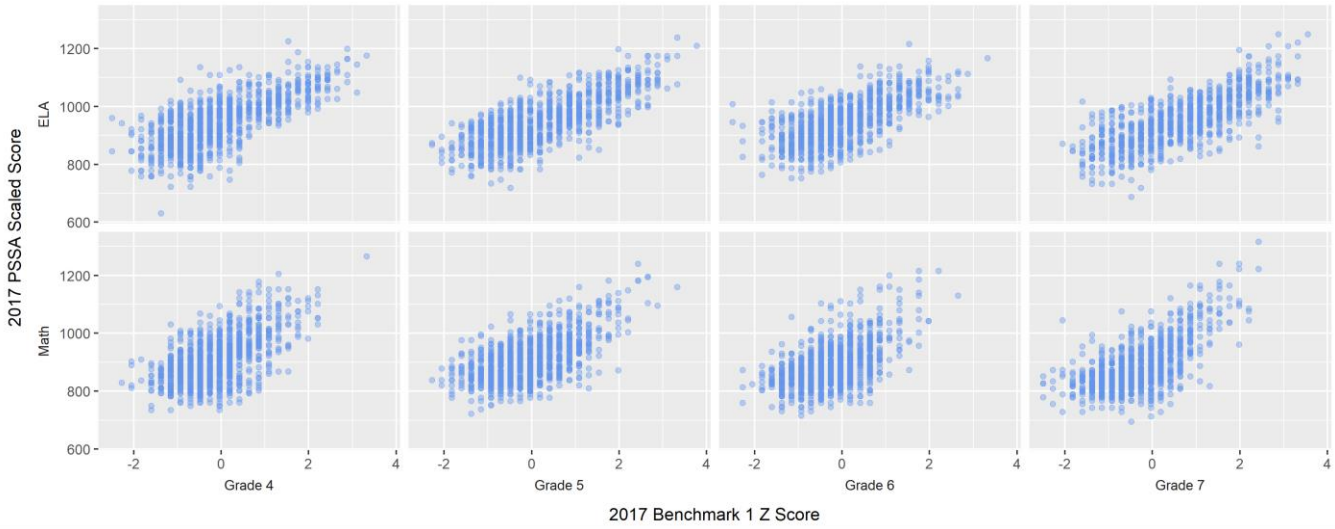


Figure E2. Relationship Between Benchmark 1 and 2017 PSSA



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

Figure E3. Relationship Between Benchmark 2 and 2016 PSSA

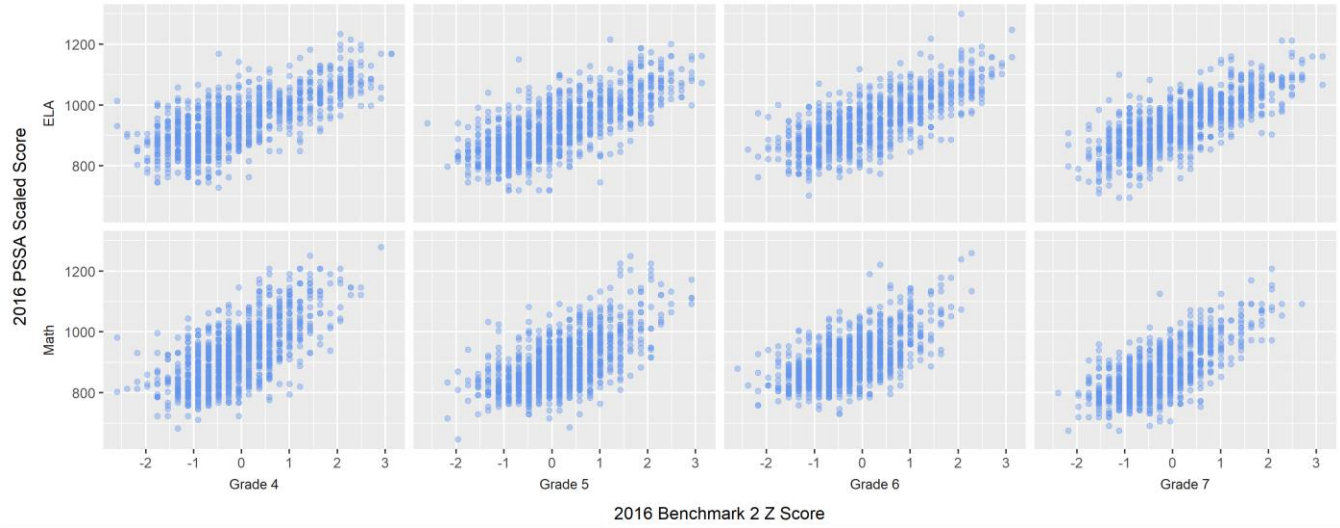
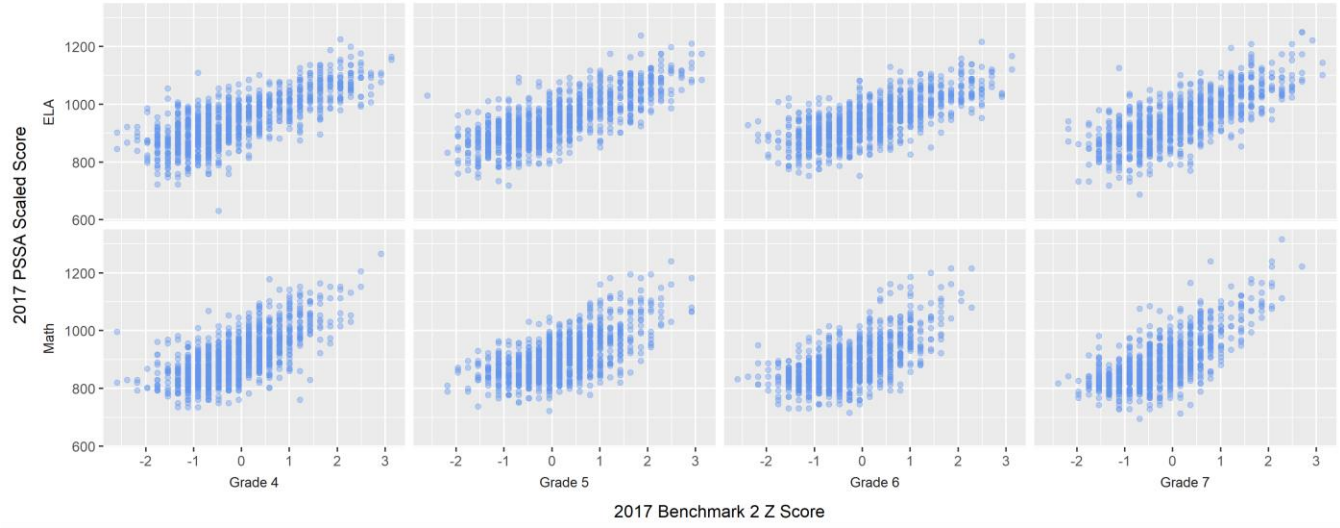


Figure E4. Relationship Between Benchmark 2 and 2017 PSSA



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

Figure E5. Relationship Between Benchmark 4 and 2016 PSSA

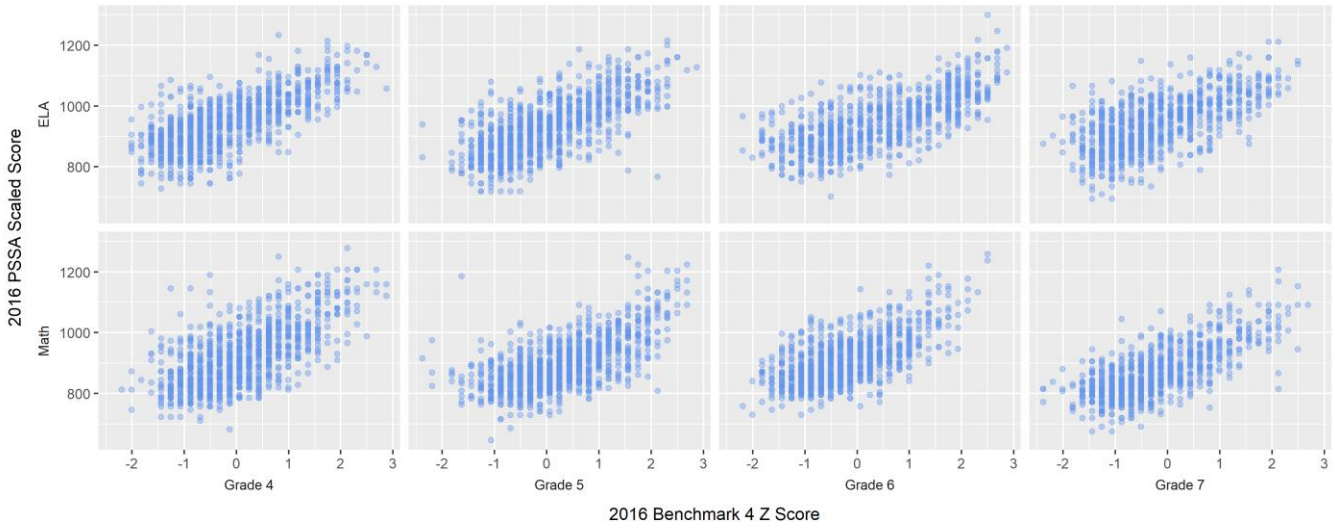
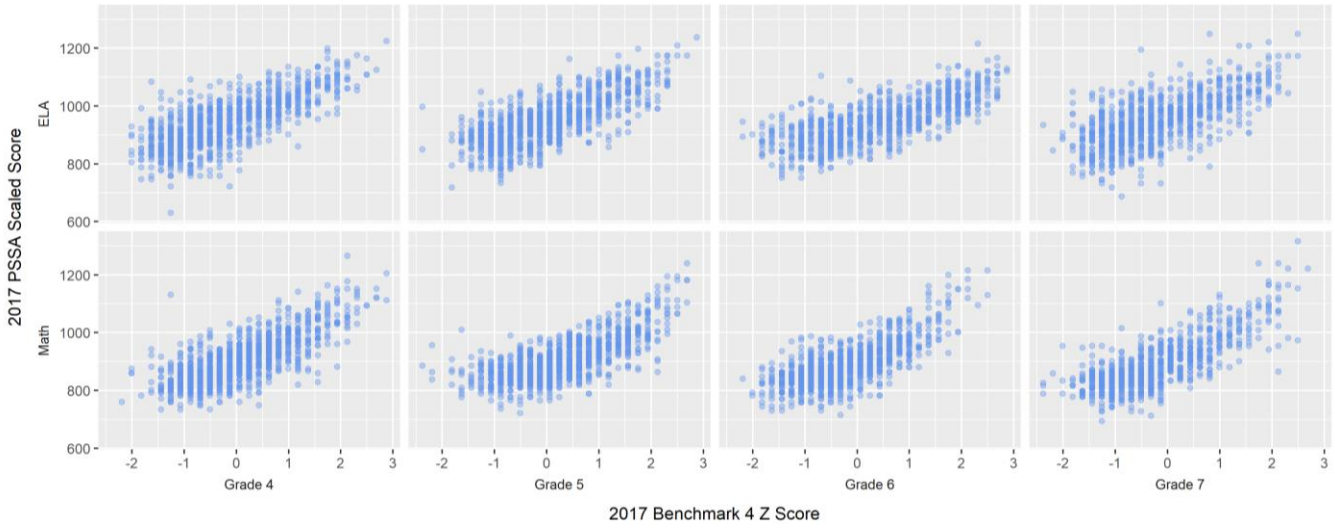
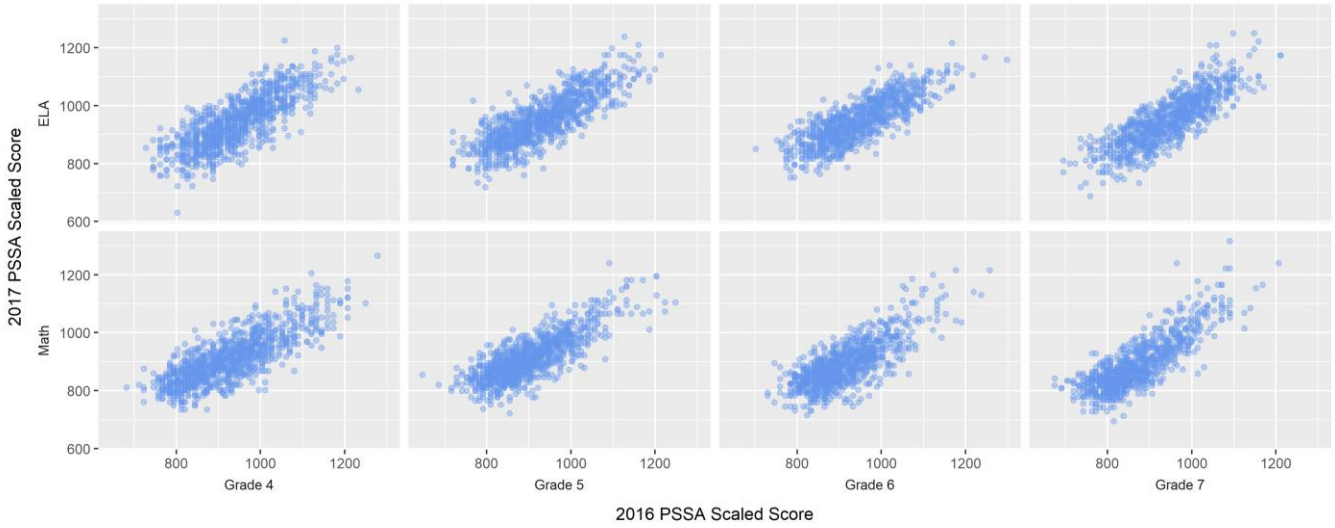


Figure E6. Relationship Between Benchmark 4 and 2017 PSSA



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

Figure E7. Relationship Between 2016 PSSA and 2017 PSSA



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