edmentum™

Research Report: Impacts of the Use of Study Island Practice and Benchmarks

Edmentum Research
April 12, 2018

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

Copyright © 2017 by Edmentum, Inc.

Executive Summary

Study Island is a practice and assessment tool that provides state-standards-aligned opportunities for students to practice their skills. Study Island is 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; Stiggins, 1999; Wolf, 2007).

The district of Allentown, Pennsylvania (PA) is a current Study Island partner. As a district in PA, Allentown 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, with the goal of closing the achievement gap for all students, and specifically for historically underperforming students. As part of their 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 (SC).

In support of Allentown School 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 Allentown 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, 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 Allentown varies by grade and content area. In Practice, students appear to be answering relatively few questions and spending minimal time over the course of the year. Where students spend more time, answer more questions, and spread their time over active weeks, positive differences are observed. This is evident in the grade 6 Math significant differences in mean scale scores and impact data. While not statistically significantly different, grade 6 ELA also shows some interesting differences in the method or approach to implementing Study Island Practice compared to other grades and content areas. In addition, when students are exposed to the Benchmarks – in this case limited to Grades 7 and 8 for ELA and Math, and grades 4 and 8 for SC – 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 6 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

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 support and empower educators to create successful student outcomes for the equitable benefit of individual students and societies, globally.

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 “involve 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, specifically for 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 & resources (as well as safe and supportive schools). As part of their accountability, the Pennsylvania System of School Assessment (PSSA) is administered annually to students in grades 3 – 8 for English Language Arts (ELA) and Math, as well as grades 4, 8, and 11 for Science (SC). 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.

The district of Allentown, Pennsylvania, is a current Study Island partner. In support of their partnership with Edmentum, this study is intended to provide a research basis for Study Island in terms of the research literature and analyses of Allentown students’ level of usage and performance data within Study Island compared to their 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; Stiggins, 1999; Wolf, 2007), particularly for students with lower achievement (Black & Wiliam, 1998), as well as in building student confidence (Stiggins, 1999). Monitoring student progress is at the heart of such programs as Curriculum Based Measurement (CBM) (Deno, 1985; Fuchs & Fuchs, 1999), Response to Intervention (RtI), and the more recent movement to consider RtI as part of a Multi-Tier System of Supports (MTSS) (Gresham, Reschly, & Shinn, 2010).

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

Although generally providing feedback to teachers and students regarding student performance can consistently enhance achievement (Baker, Gersten, & Lee, 2002), meta-analytic research indicates that it is the timeliness and

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

the type of feedback that 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. Such feedback was even more effective when it immediately followed each answer a student provided. Bangert-Drowns, Kulik, Kulik, and Morgan (1991) extended these findings by showing that timely feedback can correct future errors when it informs the learner of the correct answer, especially when students were confident in their answers (Kulhavy & Stock, 1989). Marzano, Pickering, and Pollock (2001) further concluded that feedback that also provided an explanation of the correct answer was the most effective. Through their meta-analysis, they additionally concluded that feedback is best when it encourages students to keep working on a task until they succeed and tells students where they stand relative to a target level of knowledge instead of how their performance ranks in comparison to other students.

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

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

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

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

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

Study Island uses technology both to provide students with remediation or practice at lower levels and to provide students with 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.

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

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

As students in Allentown engage in Study Island, and as teachers consider monitoring student progress throughout the year with the elements of the product, this study seeks to understand the relationship, if any, between students' use and their performance, both within the ongoing assessments in Study Island and on the state summative assessments. Early data from across the district suggests that Study Island may be a tool used in preparation for the end-of-year assessments. (See Figure 1, Appendix A, which shows higher usage across the district nearer the date of the state assessment.)

Specifically, this study seeks to answer the following research questions:

1. How were students in Allentown using Study Island practice and Study Island Benchmarks during the 2016-17 school year?
2. Is there a correlation between usage in Study Island Practice and academic performance in math, ELA, and science for students of similar ability, as measured on the summative, end-of-year PSSA state tests?
3. Is there a correlation between usage in Study Island Benchmark scores and academic performance in math, ELA, and science for students of similar ability, as measured on the summative, end-of-year PSSA state tests?

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. Resources offered in Study Island include assessments, practice tools, lessons, and instructional materials (games, flash cards, practice items, printables, etc.). The Study Island assessments are made up of formative, short-cycle "Practice" assessments and interim-like "Benchmark" assessments that include multiple-choice (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.

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 80% of the questions correctly. (Teachers can adjust the 80% threshold as appropriate for their students.) Students can also be assigned Benchmark assessments. These have been developed to mirror the content standards covered by the blueprint of the PSSA.

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

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

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 student participation in using 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 (Math) and English language arts (ELA) and students in grades 4 and 8 in science (SC). The assessment is a standards-based (criterion-referenced) test measuring Pennsylvania Core Standards of Math and ELA and the Pennsylvania Academic Standards of SC. 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 made up of various types of assessment 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 non-MC items are scored by independent raters. While Math includes 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 SC tests consists of standalone and scenario-based (Grade 8 only) MC items, and CR items scored on a 0-2-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 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 24 schools (14 elementary schools, four middle schools, three high schools, and three alternative schools) from the Allentown, PA, school district that were Study Island partners during the academic year of 2016-2017 (AY16-17). 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 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.

To evaluate just how much the district is using Study Island, "usage" is defined in terms of two participatory factors: Study Island Benchmarks (or Benchmarks) and Study Island Practice (or Practice).

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. Because of the close alignment between state tests and Benchmarks, the results of each Benchmark test should provide teachers with some indication of how prepared students could be for their state tests.

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

In terms of student usage of Benchmark assessments, Benchmark form administrations 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 1 shows the number of students responding to Benchmarks. Given the low volume of use in grades 3 through 6, *Study Island Benchmark analyses will necessarily be limited to grades 7 and 8, plus grade 4 SC; and the corresponding analyses treated separately.* The district is implementing a different district-enforced benchmark for the other grades and subjects.

Table 1 - Total Number of Students Using Study Island Benchmarks, 2016-17 School Year

Test Grade Level	ELA	Math	Science
3	23	24	
4	6	4	1406
5	4	4	
6	50	50	
7	1175	1184	
8	1123	1162	1153
Total	2381	2428	2559

Usage in Practice is defined by answering questions for a quiz or "session," in which a student answers questions associated with a ten-item practice quiz available for each topic. A topic in Practice is a grouping of conceptual material within in a subject and grade level that is associated to one or more state standards. The total number of topics available by grade and content area is provided in Table 2.

Table 2 - 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	

Table 3 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. For this study, these students are considered Study Island users **(SI Users**). All other students with no Practice questions answered are considered non-users (**SI Non-Users**).

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

Table 3 - Total Number and District Proportion of Students Using Study Island Practice

		ELA		Math		Science	
Grade	**District Total Enrolment***	**Study Island User (N)**	**Percent of District (%)**	**Study Island User (N)**	**Percent of District (%)**	**Study Island User (N)**	**Percent of District (%)**
3	1415	241	17	129	9.1		
4	1409	417	29.6	249	17.7	205	14.5
5	1306	351	26.9	125	9.6		
6	1158	197	17	217	18.7		
7	1140	330	28.9	471	41.3		
8	1123	259	23.1	232	20.7	336	29.9
Total	7551	1795	23.8	1423	18.8	541	7.2
**Total district enrollment counts from Pennsylvania Department of Education, Enrollment Reports and Projections*							

Proportionally, more students in the district are using Practice for ELA with 17 - 30% of students. In Math, the exception is 41.3% in Grade 7. Not surprisingly, SC items are mainly used by 4th and 8th graders, reflecting that the PSSA for SC is given for only these students. SC has a lower proportion of student users in grade 4 (14.5%) compared to grade 8 (29.9%).

As with any sample, it is important to understand how well the sample might generalize to other samples or the population overall. Table 4 provides the demographic make-up of the district overall with comparison to the state. The district has a much higher percentage of students qualifying for free or reduced lunch, a higher proportion of Hispanic students, and far fewer white students than the state. Table 5 provides the demographic make-up of the sample for this study. It appears the students using Study Island in the sample is comparable to the district as a whole.

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

Table 4 – District Demographics Compared to State Average

	District (%)*	State Average (%)*	Difference (District vs. State)
Individualized Education Program (IEP)	19.2	17.6	+1.6
Free and Reduced Lunch	64.1	46.7	+17.4
Hispanic	68.3	10.4	+57.9
Black	15.5	14.8	+0.7
White	11.5	67.5	-56.0
Two or More Races	3.2	3.4	-0.2
Asian or Asian/Pacific Islander	1.4	3.7	-2.3
American Indian/Alaska Native	0.1	0.2	-0.1
Hawaiian/Pacific Islander	0.1	0.1	0.0
**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.			

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

Table 5 – Sample Demographics of Study Island Practice Use (Study Island Users)

		Complete 2017 District Sample						Sample of Study Island Users					
Variable	**Category**	**ELA**		**Math**		**Science**		**ELA**		**Math**		**Science**	
		N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Race/ Ethnicity Category	American Indian/ Alaskan Native	8	0.1	9	0.1	2	0.1	4	0.2	3	0.2	0	0
	Black/African American	1041	14.4	1077	14.3	377	14.9	235	13.1	214	15	74	13.7
	Hispanic	4986	69	5208	69.2	1755	69.2	1222	68.1	933	65.6	362	66.9
	White	751	10.4	787	10.5	279	11	233	13	205	14.4	85	15.7
	Multi-Racial	333	4.6	342	4.5	89	3.5	67	3.7	45	3.2	14	2.6
	Asian	89	1.2	91	1.2	27	1.1	28	1.6	19	1.3	6	1.1
	Native Hawaiian or Other Pacific Islander	16	0.2	16	0.2	6	0.2	6	0.3	4	0.3	0	0
	Total	7224	100	7530	100	2535	100	1795	100	1423	100	541	100
Gender	Female	3405	47.1	3509	46.6	1218	48	850	47.4	684	48.1	273	50.5
	Male	3819	52.9	4021	53.4	1317	52	945	52.6	739	51.9	268	49.5
	Total	7224	100	7530	100	2535	100	1795	100	1423	100	541	100
Special Education	No	911	12.6	969	12.9	345	13.6	227	12.6	204	14.3	87	16.1
	Yes	6313	87.4	6561	87.1	2190	86.4	1568	87.4	1219	85.7	454	83.9
	Total	7224	100	7530	100	2535	100	1795	100	1423	100	541	100
Economically Disadvan-taged	No	6067	84	6142	81.6	2076	81.9	1531	85.3	1211	85.1	473	87.4
	Yes	1157	16	1388	18.4	459	18.1	264	14.7	212	14.9	68	12.6
	Total	7224	100	7530	100	2535	100	1795	100	1423	100	541	100
Title I	No	350	4.8	398	5.3	141	5.6	44	2.5	65	4.6	37	6.8
	Yes	6874	95.2	7132	94.7	2394	94.4	1751	97.5	1358	95.4	504	93.2
	Total	7224	100	7530	100	2535	100	1795	100	1423	100	541	100

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

Analyses - Study Island Practice

How were students in Allentown using Study Island Practice and Study Island Benchmarks during the 2016-17 school year?

To gauge student usage, 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. On average, students attempted the most questions in grade 6 Math (612.40), followed by grade 3 ELA (336.44) and grade 4 SC (305.34). The proportion of items students answered correctly hovers around 50% across the board, ranging from an average of 47% in 4th grade ELA to 61% in 3rd grade 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	241	2	263.0	2595	336.44	324.96	0	0.52	1	0.51	0.18
	4	417	2	72.0	1518	143.65	182.89	0	0.48	1	0.47	0.18
	5	351	1	64.0	2119	155.57	269.10	0	0.57	1	0.54	0.18
	6	197	1	99.0	1470	216.73	285.73	0	0.49	1	0.48	0.21
	7	330	1	59.0	856	80.27	88.62	0	0.55	1	0.52	0.21
	8	259	1	31.0	293	55.83	59.26	0	0.51	1	0.48	0.23
	Total	1795	1	71.0	2595	155.56	232.79	0	0.52	1	0.50	0.20
Math	3	129	2	50.0	1182	142.67	194.48	0	0.62	1	0.61	0.19
	4	249	1	41.0	470	60.98	61.29	0	0.55	1	0.53	0.21
	5	125	1	52.0	956	106.26	139.74	0	0.55	1	0.52	0.25
	6	217	1	113.0	6110	612.40	910.44	0	0.59	1	0.58	0.19
	7	471	1	112.0	1309	165.26	177.92	0	0.48	1	0.48	0.21
	8	232	1	29.0	1001	119.34	209.43	0	0.53	1	0.51	0.23
	Total	1423	1	69.0	6110	200.48	425.59	0	0.54	1	0.52	0.22
Science	4	205	1	222.0	1431	305.34	343.95	0	0.62	1	0.56	0.21
	8	336	1	135.5	765	165.10	147.20	0	0.52	1	0.51	0.17
	Total	541	1	138.0	1431	218.24	250.54	0	0.56	1	0.53	0.19

To understand how much time Study Island Users spent answering these questions, Table 7 provides descriptive data on the amount of time spent by grade and content area. Eighth graders spend, on average, the least amount of time and answer the fewest items with a median of about 25 minutes and 29 and 31 items answered in Math and ELA, respectively (Table 6). Grade 6 Math students spent the most amount of time overall – about 562 minutes, or 9½ hours, answering just over 600 items, on average. Sixth graders also spend more time in ELA – about 241 minutes, or 4 hours, answering just over 200 items. The students who answer the most questions in ELA are 3rd graders, with an average of 337 items attempted in an average of just over 200 minutes or about 3 ½ hours. By looking at Figure 1, we can see how many students are distributed across the amount of time spent. For example, there are many Math student users in Grade 4, but they are spending much less time using Study Island than the fewer users spending more time in Grade 6.

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 most frequent number of weeks with usage are in Grade 6 Math and Grade 3 ELA at about 10 weeks. These data are comparable to the frequency rates of number of items and time provided above. Figure 2 shows the distribution of active weeks for each grade and subject. It shows that, generally, ELA has more active weeks for grades 3 – 5, while grades 6 – 8 Math have more active weeks. These views help to illustrate how the Practice items are used across grades and across subject areas. It helps to show that, for example, there are many grade 8 students

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

using, but not in as many active weeks as grade 6. Also, while there are fewer grade 3 and 6 students, their use spans far greater active weeks in ELA (grade 3) and Math (grade 6).

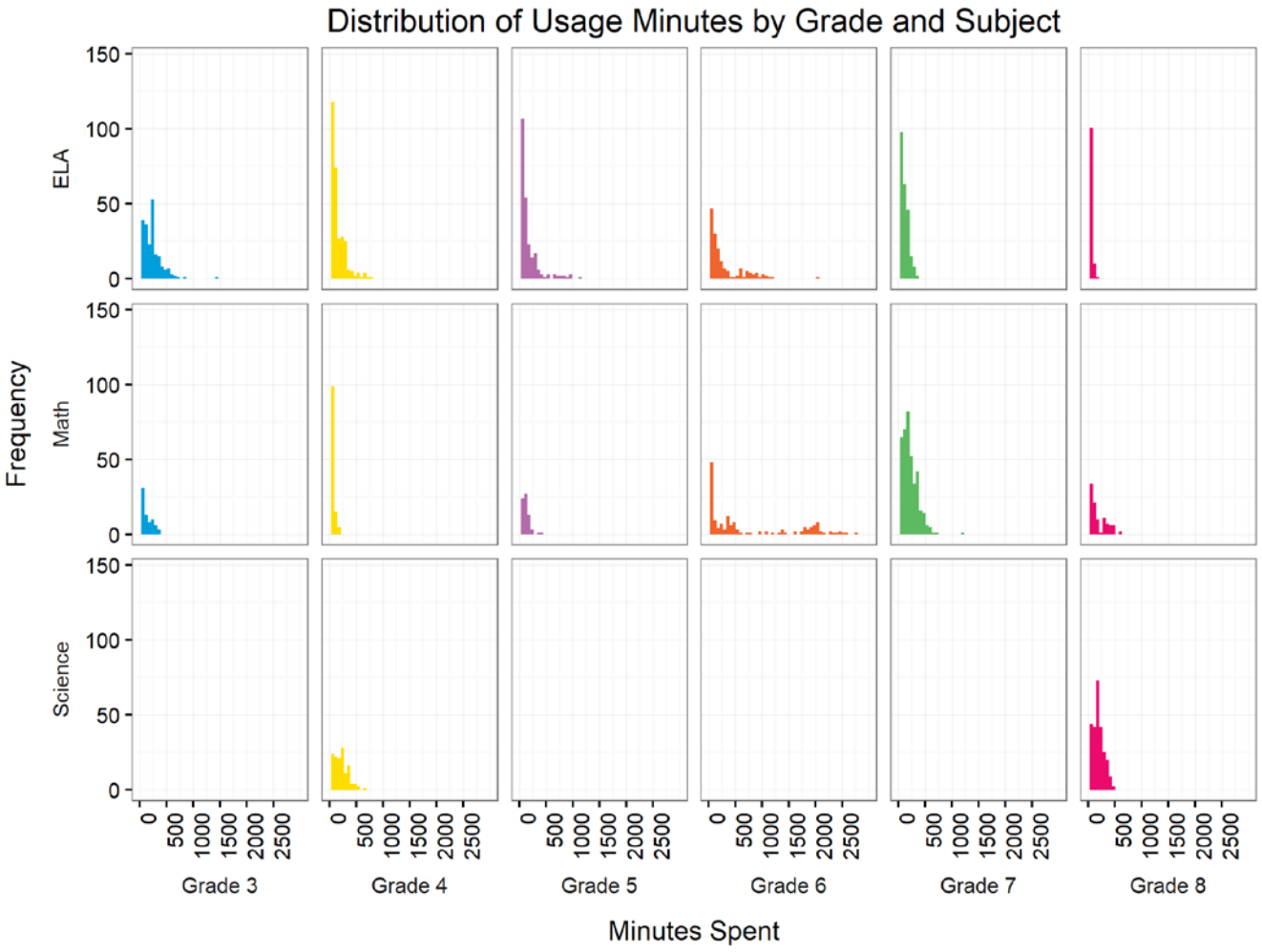
Finally, to see just how much of the time occurs within each active week, Table 9 provides the amount of time per week as a result of calculating the total time spent in Practice divided by the number of active weeks. The average amount of time per active week ranges from about 15 minutes in 8th grade ELA to 35 minutes in 6th grade Math. Sixth grade ELA is the highest for ELA with 31 minutes per week. (Third grade ELA spent about 20 minutes per active week.)

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	241	6.33	190.17	1418.28	206.53	176.04
	4	417	0.68	71.48	765.57	114.98	130.37
	5	351	0.42	55.93	1121.40	122.18	176.62
	6	197	0.57	113.77	2067.85	240.56	315.17
	7	330	0.22	68.52	353.37	89.35	75.92
	8	259	0.08	25.30	150.98	34.30	27.90
	Total	1795	0.08	64.45	2067.85	126.11	173.20
Math	3	129	0.55	34.37	388.22	82.59	97.42
	4	249	0.42	28.08	198.55	39.31	36.35
	5	125	0.68	47.67	433.30	70.05	74.65
	6	217	0.40	89.92	2789.23	561.67	791.82
	7	471	0.12	162.03	1221.57	186.72	152.80
	8	232	0.03	25.10	574.83	89.21	133.50
	Total	1423	0.03	59.95	2789.23	182.52	369.31
Science	4	205	0.18	112.22	645.67	143.38	142.39
	8	336	0.27	153.00	461.10	152.36	116.47
	Total	541	0.18	139.10	645.67	148.96	126.86

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

Figure 1. Distribution of time in minutes by grade and content area for Study Island Practice Users in the 2016-17 School Year.



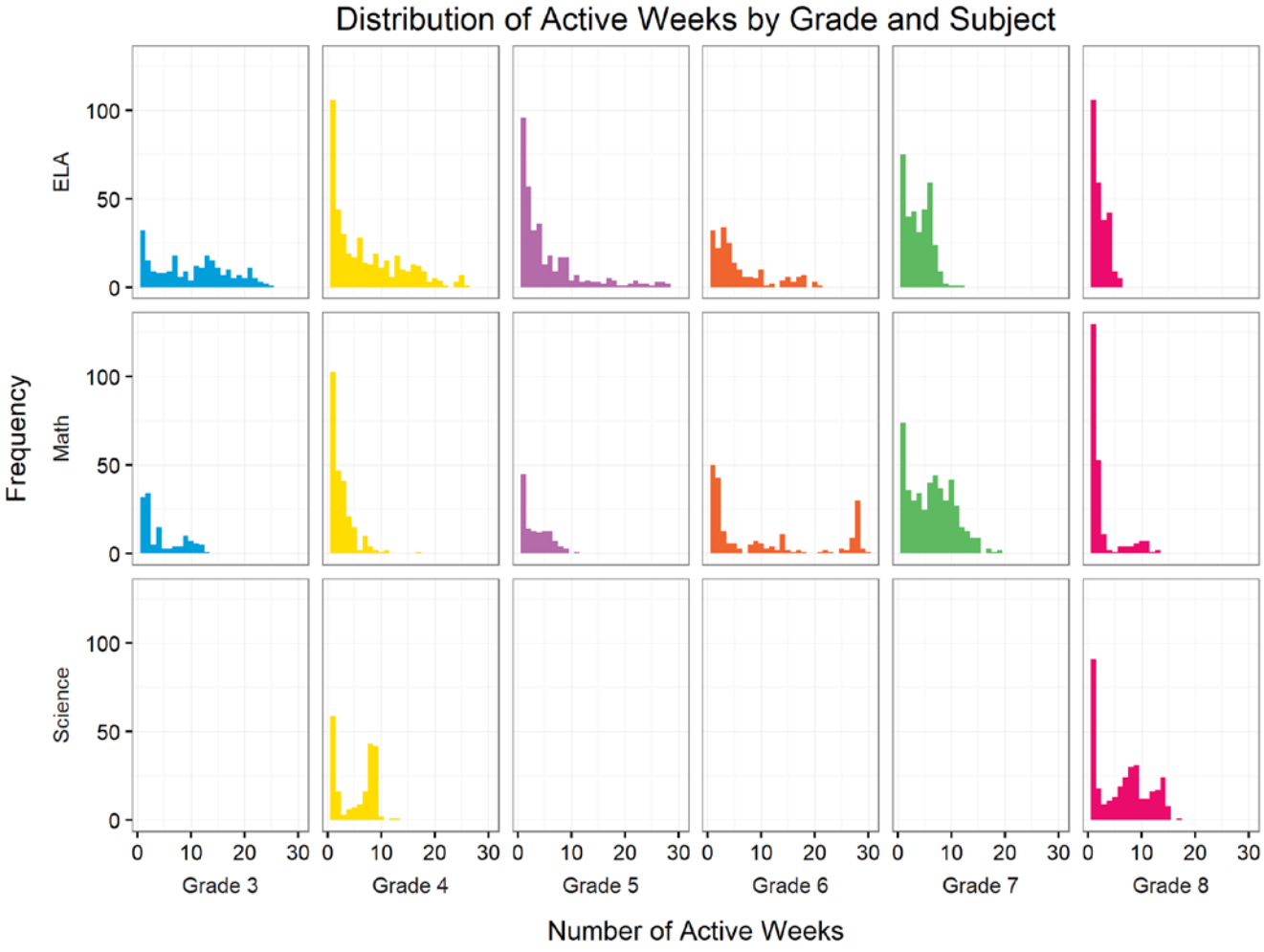
5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

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

Subject	Grade	N	Min.	Med.	Max.	Mean	SD
ELA	3	241	1	11	25	10.25	6.72
	4	417	1	5	26	7.08	6.41
	5	351	1	3	28	5.58	6.14
	6	197	1	4	21	5.99	5.29
	7	330	1	4	12	3.86	2.28
	8	259	1	2	6	2.24	1.33
	Total	1795	1	4	28	5.80	5.72
Math	3	129	1	2	13	4.40	3.67
	4	249	1	2	17	2.65	2.23
	5	125	1	3	11	3.38	2.45
	6	217	1	4	30	10.06	10.52
	7	471	1	6	19	6.46	4.10
	8	232	1	1	13	2.61	2.97
	Total	1423	1	3	30	5.26	5.75
Science	4	205	1	7	13	5.33	3.42
	8	336	1	7	17	6.55	4.64
	Total	541	1	7	17	6.09	4.26

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

Figure 2. Distribution of active weeks by grade and content area for Study Island Practice Users in the 2016-17 School Year.



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.	Med.	Max.	Mean	SD
ELA	3	241	4.48	18.84	64.47	20.53	9.98
	4	417	0.68	13.79	127.88	18.29	17.82
	5	351	0.42	18.76	73.67	20.20	11.72
	6	197	0.57	28.08	165.07	31.44	20.85
	7	330	0.22	20.01	185.88	21.40	13.89
	8	259	0.08	13.79	144.65	15.81	13.24
	Total	1795	0.08	17.48	185.88	20.62	15.47
Math	3	129	0.55	15.75	46.53	16.05	9.11
	4	249	0.42	13.60	198.55	15.87	15.46
	5	125	0.68	16.75	89.86	18.00	14.28
	6	217	0.40	28.62	99.62	35.16	24.57
	7	471	0.12	25.71	255.70	27.62	19.46
	8	232	0.03	22.27	161.36	26.22	21.65
	Total	1423	0.03	20.53	255.70	24.59	20.10
Science	4	205	0.18	19.08	71.74	21.57	15.57
	8	336	0.27	22.61	90.45	23.17	11.66
	Total	541	0.18	22.05	90.45	22.56	13.29

Is there a relationship between usage in Study Island Practice and the summative, end-of-year PSSA state tests? If there is a relationship, how strong is it?

PSSA Performance and Study Island Practice Use

Table 10 compares performance on all content areas of the PSSA in terms of scale scores for both Study Island User and SI Non-User groups, compared to the district and state. This table shows that Allentown has lower mean scale scores compared to the state. Study Island Non-Users have similar scores to Allentown, and those are lower than Users. Specifically, Study Island Users outperform the district and Study Island Non-Users in all content areas and grades, except grade 8, which sees Study Island Users scoring less than Study Island Non-Users. Mean scores differ as much as 48 points in Math Grade 3 and 43 points in ELA Grade 4. PSSA standard errors are only about 4 points.

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

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

		Study Island Practice User			Study Island Practice Non-User			District			State		
Subject	**Grade**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
ELA	3	241	990.59	93.82	1055	972.62	103.30	1296	975.96	101.81	124923	1039.30	111.21
	4	417	997.74	98.24	883	954.82	99.03	1300	968.59	100.76	125200	1030.55	112.72
	5	351	989.78	89.09	867	950.01	98.58	1218	961.47	97.58	124183	1029.58	112.26
	6	197	981.34	88.79	867	961.09	85.97	1064	964.84	86.82	123170	1035.08	106.23
	7	330	961.35	94.03	724	952.65	100.16	1054	955.37	98.32	125744	1031.71	113.46
	8	259	947.64	95.26	778	953.12	95.18	1037	951.75	95.18	123653	1025.03	108.86
Math	3	129	997.33	115.63	1199	948.51	115.29	1328	953.26	116.18	125205	1019.85	129.66
	4	249	959.03	114.99	1091	922.74	98.51	1340	929.49	102.70	125575	993.58	118.67
	5	125	932.54	86.93	1147	922.00	89.55	1272	923.04	89.31	124405	991.82	119.70
	6	217	935.34	91.11	884	906.15	92.58	1101	911.90	92.98	123112	976.25	115.64
	7	471	888.32	91.53	631	876.63	90.59	1102	881.63	91.14	125584	968.65	126.69
	8	232	858.94	79.55	862	875.00	80.71	1094	871.59	80.70	123271	953.46	118.27
Science	4	205	1324.75	153.54	1129	1286.37	136.98	1334	1292.26	140.27	125488	1406.07	170.94
	8	336	1154.41	144.29	758	1150.72	136.22	1094	1151.85	138.69	122716	1299.33	183.99

To discern whether or not these differences are significant, we must take into account the differences in student ability across the user groups. That is to say, if students using Study Island are generally higher-ability students, whether or not they are users may be meaningless. To understand the impact of the treatment – in this case Study Island Practice use – 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.

A propensity score matching (PSM) (Rosenbaum & Rubin, 1983) was conducted to align students in the user group to the students in the non-user group by ability – in this case the 2016 PSSA scores – so that statistical analyses of the 2017 PSSA mean score differences can be conducted, while assuring any discernable differences reflect a difference in the impact of use rather than an inherent difference in ability from the start.

Only grades 4 – 8 for ELA and Math could be included in the analysis because third graders do not have a prior PSSA score and because the PSSA in SC is only given to 4th and 8th graders. Some other users within these grades in Math and ELA were also eliminated from the sample because they did not have a PSSA 2016 score. The total resulting N is included in Table 11. (Please 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.)

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

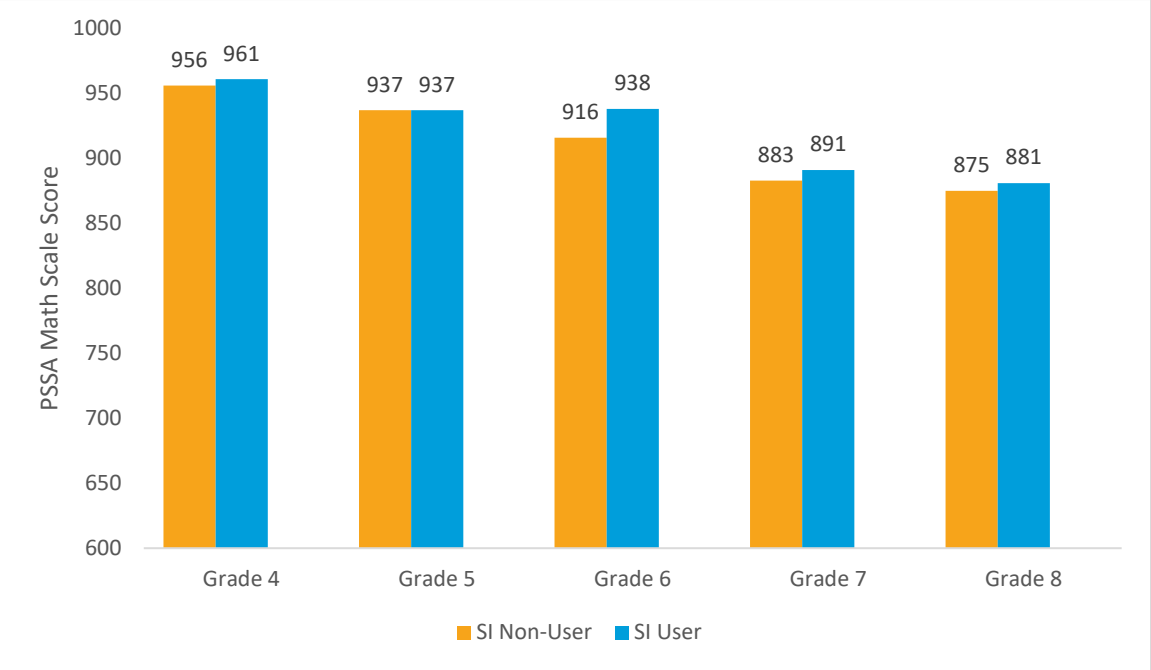
Table 11 - *T*-Test Comparisons of PSSA Scale Score between Matched Study Island Users and Study Island Non-Users

		Study Island User		Study Island Non-User		Matched	PSSA 2017					
Subject	**Grade**	**Mean**	**SD**	**Mean**	**SD**	**N**	**Mean Difference**	**95% CI**		***t***		**df**
ELA	4	988.15	94.65	986.84	96.21	357	1.31	-15.33	12.72	-0.183		712
	5	990.52	87.15	986.04	94.68	315	4.48	-18.72	9.76	-0.618		628
	6	987.02	90.02	981.66	88.53	166	5.36	-24.63	13.92	-0.546		330
	7	967.88	91.10	972.66	103.33	261	-4.78	-11.97	21.53	0.561		520
	8	953.84	90.63	952.81	91.21	219	1.03	-18.11	16.05	-0.119		436
Math	4	960.86	114.69	956.43	107.64	228	4.43	-24.90	16.05	-0.425		454
	5	936.98	85.34	936.54	87.49	112	0.44	-23.21	22.31	-0.039		222
	6	938.18	90.37	915.86	92.32	187	**22.32**	-40.90	-3.75	-2.363	**	372
	7	890.63	90.41	882.51	90.29	336	8.12	-21.80	5.57	-1.164		670
	8	881.16	81.52	874.95	76.73	175	6.21	-22.86	10.43	-0.734		348

Note: SC Grades 4 and 8, as well as Grade 3 ELA and Math were not included in the PSM matching, and thus not included in this analysis.

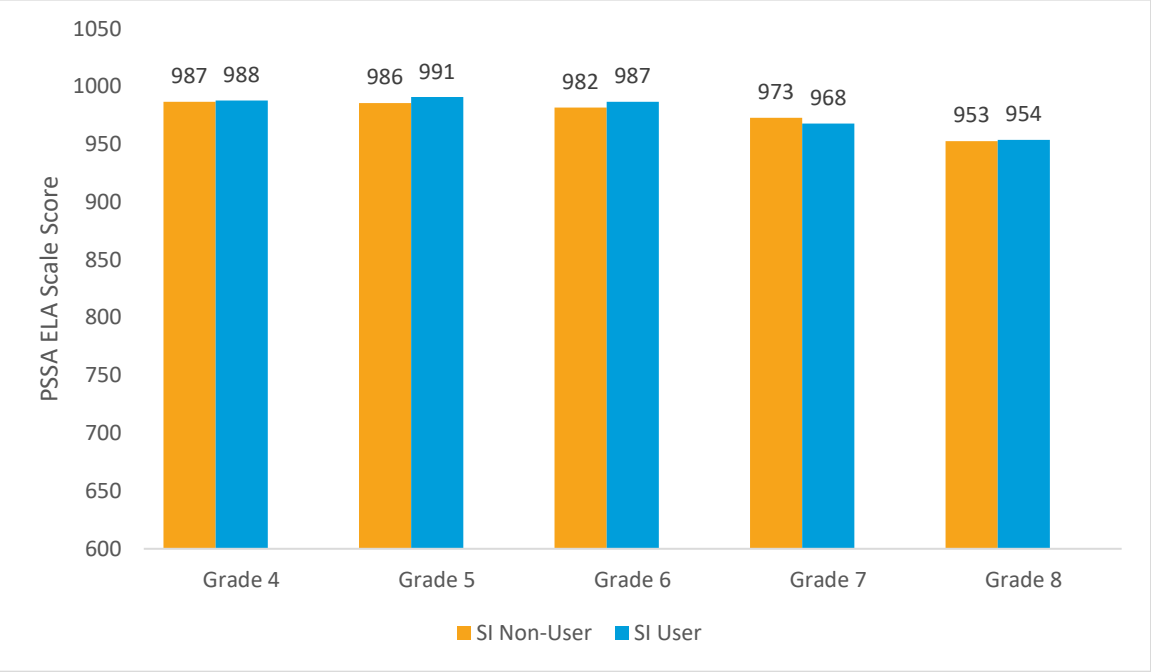
A *t*-test was conducted after matching to compare the 2017 PSSA scores across the matched SI User and SI Non-User groups. Results from the analysis are shown in Table 11 where *N* reports the equal size of the matched groups. Figures 3 and 4 display the mean differences in PSSA scale score between SI User and Non-User groups after propensity score matching. 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 5th grade Math and 7th grade ELA, only the mean scale score difference for 6th grade math is statistically significant. This is not surprising given the various differences observed in the number of items answered, the amount of time, and active weeks where grade 6 was clearly using differently.

Figure 3. Adjusted 2017 PSSA Math mean scale scores for Study Island (SI) Users and Study Island Non-Users



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

Figure 4. Adjusted 2017 PSSA ELA mean scale scores for Study Island (SI) Users and Study Island Non-Users



Because the proficiency level is a key variable in accountability, Table 12 provides the impact data of the number and percentage of students performing in the top two proficiency categories as “overall proficiency” across the unmatched user groups, Allentown, and the state. Similar differences in overall proficiency are seen as with mean scale score differences: SI Users tend to have higher percentages than SI Non-Users, SI Non-Users are more similar to the district overall, and the district has far fewer than the state.

By using the matched scores from the PSM, a chi-square test was run to discern differences in proficiency levels across groups, findings for which are reported in Table 13. The only group that has statistically significant differences in the proficiency level student categorization is within 6th grade math, the same group where mean PSSA scale score differences were found.

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

Table 12 - Percentage of Students in Grades 3-8 Scoring Proficient or Advanced on the 2017 PSSA Allentown Compared to Pennsylvania

		Study Island Practice User		Study Island Practice Non-User		District		State	
Subject	**Grade**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
ELA	3	114	47.3	429	40.7	543	41.9	80825	64.7
	4	197	47.2	273	30.9	470	36.2	76372	61.0
	5	155	44.2	237	27.3	392	32.2	74013	59.6
	6	80	40.6	289	33.3	369	34.7	78336	63.6
	7	104	31.5	213	29.4	317	30.1	74692	59.4
	8	72	27.8	252	32.4	324	31.2	72708	58.8
Math	**3**	57	44.2	366	30.5	441	32.5	68112	54.4
	4	92	36.9	229	21.0	352	25.5	58518	46.6
	5	27	21.6	220	19.2	264	20.2	54365	43.7
	6	53	24.4	135	15.3	211	18.4	49491	40.2
	7	60	12.7	63	10.0	146	12.8	47471	37.8
	8	14	6.0	66	7.7	94	8.3	40063	32.5
SC	**4**	116	56.6	534	47.3	668	48.7	93614	74.6
	8	63	18.8	149	19.7	229	20.3	64671	52.7

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

Table 13 - Chi-Square Test Comparison of 2017 PSSA Proficiency Level Categorization between Matched *Study Island* Users and *Study Island* Non-Users

		ELA			Math		
Grade	**Performance Level**	**User (%)**	**Non-User (%)**	**Chi-Sq.**	**User (%)**	**Non-User (%)**	**Chi-Sq.**
4	Below Basic	11.17	10.41	2.443	36.24	38.43	5.8
	Basic	40.86	44.92		25.76	32.75	
	Proficient	32.74	32.74		23.58	15.72	
	Advanced	15.23	11.93		14.41	13.1	
	Total	100	100		100	100	
5	Below Basic	13.75	12.5	6.085	37.5	34.82	1.455
	Basic	40.94	46.88		40.18	41.07	
	Proficient	40.94	33.44		19.64	23.21	
	Advanced	4.38	7.19		2.68	0.89	
	Total	100	100		100	100	
6	Below Basic	10.71	13.69	2.543	36.17	49.47	7.361*
	Basic	44.64	47.62		39.89	34.04	
	Proficient	36.31	33.93		19.68	13.3	
	Advanced	8.33	4.76		4.26	3.19	
	Total	100	100		100	100	
7	Below Basic	7.27	10.55	5.036	61.84	66.05	1.948
	Basic	59.27	50.18		23.95	22.11	
	Proficient	27.27	32.36		10.53	8.16	
	Advanced	6.18	6.91		3.68	3.68	
	Total	100	100		100	100	
8	Below Basic	21.86	23.26	0.423	71.69	72.89	0.46
	Basic	49.3	47.91		19.88	20.48	
	Proficient	25.58	24.65		7.23	5.42	
	Advanced	3.26	4.19		1.2	1.2	
	Total	100	100		100	100	

Note: SC Grades 4 and 8, as well as Grade 3 ELA and Math were not included in the PSM matching and thus not included in this analysis.

Study Island Benchmarks and the PSSA

Is there a relationship between Study Island Benchmark scores and the summative, end-of-year PSSA state tests? If there is a relationship, how strong is it?

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. To evaluate the scores on Study Island Benchmarks, student data include only the MC item responses. In addition, 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, and 35 for SC grades 4 and 8. Table 14 reports descriptive statistics for student performance of Study Island Benchmark MC items for fall (Benchmark 1) and winter (Benchmark 2) administrations.

In general, the mean benchmark raw scores are low, with only 7th grade ELA students having a mean score that is greater than 50% correct. The mean raw scores do increase very slightly from Benchmark 1 to Benchmark 2. 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

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

vary in difficulty from form to form. Additionally, they have not been statistically linked to or evaluated against state summative test results, of scores or levels of proficiency.

Table 14 - Benchmark Raw Scores Descriptive Statistics

Subject	Grade	Benchmark	Max Score Possible	N	Min.	Max.	Mean	SD
ELA	7	Benchmark 1	28	1,004	0	27	13.79	5.70
		Benchmark 2	28	1,091	0	27	14.04	5.39
	8	Benchmark 1	28	985	0	27	13.23	5.93
		Benchmark 2	28	1,035	0	27	12.79	5.40
Math	7	Benchmark 1	28	1,008	0	24	10.04	4.75
		Benchmark 2	28	1,088	0	26	11.24	4.39
	8	Benchmark 1	28	1,016	0	25	10.02	4.47
		Benchmark 2	28	1,066	0	27	11.08	4.13
SC	4	Benchmark 1	35	1,193	0	30	15.39	5.93
		Benchmark 2	35	1,266	0	34	16.45	6.26
	8	Benchmark 1	35	931	0	32	13.45	5.91
		Benchmark 2	35	1,028	0	33	14.18	6.25

To address the potential variability in difficulty, Benchmark Z-scores 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 15. 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 and winter Benchmarks for ELA and Math, *or* PSSA 2017 scores and fall and winter Benchmarks for SC considering that the PSSA is given only in grades 4 and 8 (meaning no prior-year scores are available).

The data shows 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 15 - Sample Sizes for Benchmark Data Analysis

Subject	Grade	Score*	N	Min.	Max.	Mean	SD
ELA	7	PSSA Scaled Score (2016)	765	710	1282	958.03	100.01
		PSSA Scaled Score (2017)	765	718	1265	969.21	97.23
		Study Island Benchmark 1	765	-2.42	2.32	0.07	0.99
		Study Island Benchmark 2	765	-2.6	2.4	0.09	0.99
	8	PSSA Scaled Score (2016)	741	724	1239	963.26	91.71
		PSSA Scaled Score (2017)	741	722	1282	967.10	90.59
		Study Island Benchmark 1	741	-2.23	2.32	0.10	0.99
		Study Island Benchmark 2	741	-2.37	2.63	0.11	0.97
Math	7	PSSA Scaled Score (2016)	749	691	1289	899.60	102.11
		PSSA Scaled Score (2017)	749	728	1285	888.77	91.84
		Study Island Benchmark 1	749	-2.12	2.94	0.05	1.00
		Study Island Benchmark 2	749	-2.56	3.36	0.08	0.98
	8	PSSA Scaled Score (2016)	761	698	1270	887.05	82.73
		PSSA Scaled Score (2017)	761	716	1259	883.06	83.59
		Study Island Benchmark 1	761	-2.24	3.35	0.06	1.02
		Study Island Benchmark 2	761	-2.68	3.85	0.09	1.04
SC	4	PSSA Scaled Score (2017)	1,150	1050	1873	1300.27	139.45
		Study Island Benchmark 1	1,150	-2.43	2.46	0.02	1.00
		Study Island Benchmark 2	1,150	-2.63	2.64	0.04	0.99
	8	PSSA Scaled Score (2017)	893	925	1623	1163.73	140.16
		Study Island Benchmark 1	893	-1.94	2.97	0.02	1.00
		Study Island Benchmark 2	893	-2.27	3.01	0.08	0.98

** Study Island Benchmark scores were transformed to Z score scale for comparison.*
Note that the PSSA is not administered in SC in prior grades 3 or 7, and thus no PSSA 2016 scores are provided.

Analyses – Study Island Benchmarks

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.

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 and may predict one from 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). In situations where assumptions are violated, the correlation may become inadequate to explain a given relationship. In this study, only the PSSA 8th grade ELA scores 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

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

correlation coefficients are provided. The Spearman rho is a nonparametric statistic that does not require normally distributed data and is interpreted in similar fashion to other types of correlations.

Tables 16 and 17 provide 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 0.01 level. This indicates that there is a strong enough association that one can infer that the two assessments are measuring similar constructs and performance on one can be predictive of performance on the other.

To understand the magnitude of the association, Cohen, Cohen, West, and Aiken (2003) provide a standard or rule of thumb for interpreting the strength of the relationship, or the effect size. Correlation coefficients between 0.10 and 0.29 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 16 shows, there is a large, positive correlation between students' performance on *Study Island* Benchmarks and their performance on the PSSA in all grades and subjects.

Table 16 - Correlation between Scores on PSSA and *Study Island* Benchmarks by Grade and Subject

Subject	Grade	Score	PSSA 2016	Benchmark 1	Benchmark 2	PSSA 2017
ELA	7	PSSA 2016	1.000			
		Benchmark 1	.763**	1.000		
		Benchmark 2	.772**	.730**	1.000	
		PSSA 2017	.840**	.773**	.783**	1.000
	8	PSSA 2016	1.000			
		Benchmark 1	.736**	1.000		
		Benchmark 2	.708**	.728**	1.000	
		PSSA 2017	.852**	.755**	.755**	1.000
Math	7	PSSA 2016	1.000			
		Benchmark 1	.650**	1.000		
		Benchmark 2	.711**	.568**	1.000	
		PSSA 2017	.804**	.585**	.717**	1.000
	8	PSSA 2016	1.000			
		Benchmark 1	.587**	1.000		
		Benchmark 2	.617**	.599**	1.000	
		PSSA 2017	.743**	.610**	.683**	1.000
Science	4	Benchmark 1		1.000		
		Benchmark 2		.723**	1.000	
		PSSA 2017		.750**	.774**	1.000
	8	Benchmark 1		1.000		
		Benchmark 2		.632**	1.000	
		PSSA 2017		.644**	.683**	1.000

Does the relationship between Study Island Benchmark scores and PSSA scores remain after accounting for a student's previous PSSA performance?

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 (SI User and SI Non-User) and allowed for the comparison of treatment (SI User) and a pseudo-control group (SI Non-User). Given the continuous nature of the Benchmark assessments, partial correlations were used to determine if Benchmark scores are correlated with the PSSA 2017 scores. The partial correlation method allows for the

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

removal of the prior PSSA 2016 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 in order 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 medium-sized correlations remain between use of Study Island Benchmarks and 2017 PSSA scores. All values are significant at the 0.01 level. This indicates that Benchmark scores do influence the PSSA scores, suggesting that students who participate in the Benchmarks have a positive and significantly different outcome on their PSSA scores.

Table 17 - Correlations between Scores on PSSA 2017 and Study Island Benchmarks after Accounting for PSSA 2016

Subject	Grade	Score*	Benchmark 1	Benchmark 2	PSSA 2017
ELA	7	Benchmark 1	1		
		Benchmark 2	.343**	1	
		PSSA 2017	.377**	.391**	1
	8	Benchmark 1	1		
		Benchmark 2	.432**	1	
		PSSA 2017	.362**	.409**	1
Math	7	Benchmark 1	1		
		Benchmark 2	.198**	1	
		PSSA 2017	.138**	.349**	1
	8	Benchmark 1	1		
		Benchmark 2	.371**	1	
		PSSA 2017	.320**	.426**	1

** PSSA 2016 is the PSSA scaled score in 2016, Benchmark 1 is the Study Island Benchmark 1 Z score, Benchmark 2 is the Study Island Benchmark 2 Z score, and PSSA 2017 is the PSSA scaled score in 2017.*

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 Allentown vary by grade and content area. In Practice, students appear to be answering relatively few questions and spending minimal time over the course of the year. Where students spend more time, answer more questions, and spread their time over active weeks, positive differences are observed. This is evident in the grade 6 Math significant differences in mean scale scores and impact data. While not statistically significantly different, grade 6 ELA also shows some interesting differences in the method or approach to implementing Study Island Practice compared to other grades and content areas. In addition, when students are exposed to the Benchmarks – in this case limited to Grades 7 and 8 for ELA and Math, and grades 4 and 8 for SC – 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 6 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

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.

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), 219–32.

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), 659–71.

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).

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.

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

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership, 63(3),* 19–24.

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.). *Enrollment Reports and Projections*. Retrieved from http://www.education.pa.gov/Data-and-Statistics/Pages/Enrollment%20Reports%20and%20Projections.aspx

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.

Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan, 81*(3), 191–198.

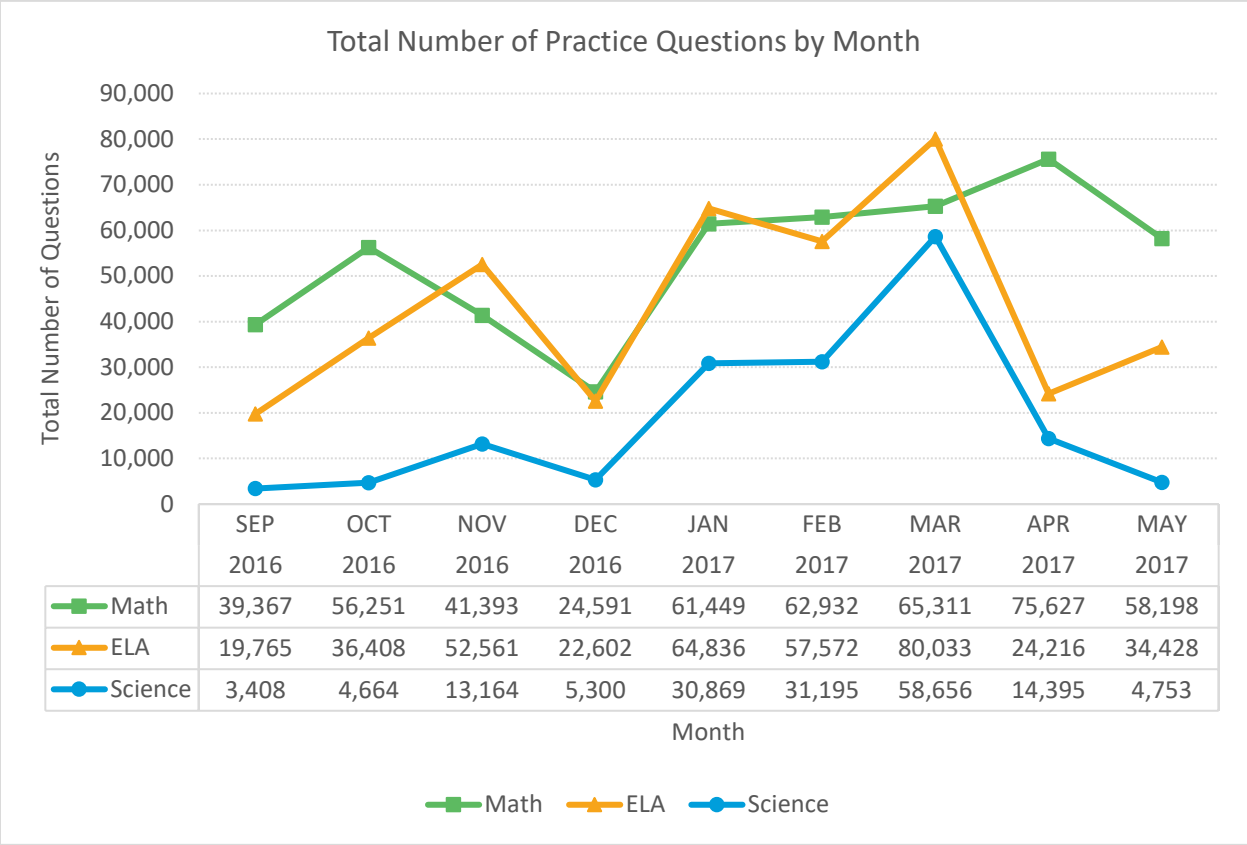
Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21,* 361–375.

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



	SEP 2016	OCT 2016	NOV 2016	DEC 2016	JAN 2017	FEB 2017	MAR 2017	APR 2017	MAY 2017
Math	39,367	56,251	41,393	24,591	61,449	62,932	65,311	75,627	58,198
ELA	19,765	36,408	52,561	22,602	64,836	57,572	80,033	24,216	34,428
Science	3,408	4,664	13,164	5,300	30,869	31,195	58,656	14,395	4,753

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

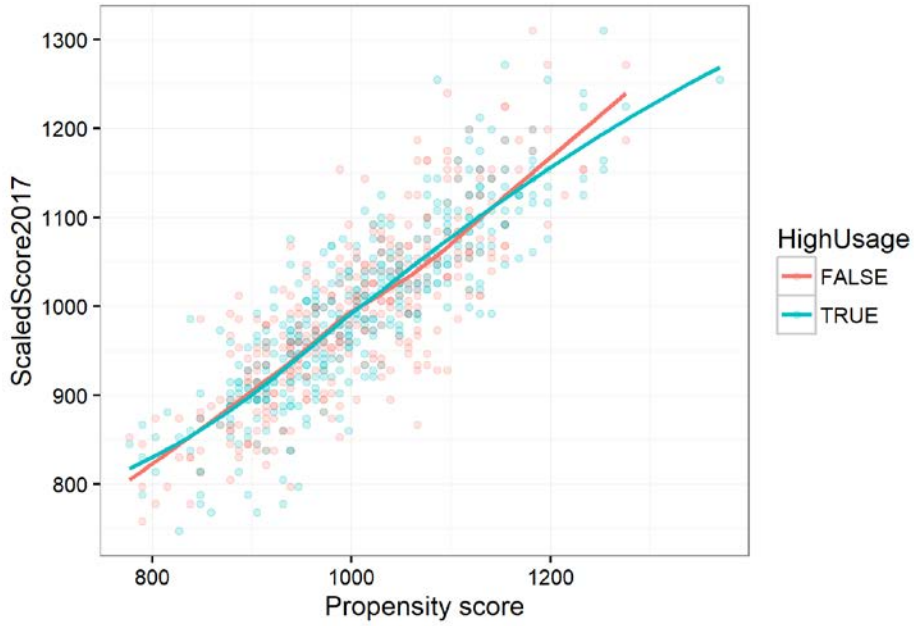
Appendix B: Volume of Benchmark Test Use

Subject	Grade	Form	N	First Test Date	Last Test Date
ELA	3	1	17	10/14/2016	10/14/2016
	3	2	21	12/7/2016	12/8/2016
	3	3	20	3/1/2017	3/3/2017
	3	4	21	5/18/2017	5/18/2017
	4	1	6	3/23/2017	3/30/2017
	5	1	4	3/21/2017	3/29/2017
	6	1	49	11/15/2016	11/18/2016
	6	2	49	2/10/2017	2/21/2017
	7	1	1,008	9/7/2016	10/26/2016
	7	2	1,101	11/15/2016	2/3/2017
	7	3	48	2/13/2017	2/16/2017
	8	1	985	9/6/2016	10/21/2016
	8	2	1,039	11/11/2016	2/3/2017
	8	3	38	2/13/2017	2/15/2017
Math	3	1	21	10/13/2016	10/17/2016
	3	2	21	12/8/2016	3/2/2017
	3	3	20	3/2/2017	3/3/2017
	3	4	20	5/19/2017	5/19/2017
	4	1	4	3/6/2017	4/21/2017
	5	1	4	4/10/2017	4/12/2017
	6	1	50	11/11/2016	11/21/2016
	6	2	50	2/8/2017	2/16/2017
	7	1	1,020	9/8/2016	11/14/2016
	7	2	1,098	12/20/2016	2/21/2017
	8	1	1,021	9/8/2016	11/14/2016
	8	2	1,069	1/6/2017	2/21/2017
Science	4	1	1,282	10/11/2016	10/31/2016
	4	2	1,306	2/6/2017	3/9/2017
	4	3	45	5/30/2017	6/1/2017
	8	1	1,023	10/11/2016	11/18/2016
	8	2	1,089	2/6/2017	3/27/2017

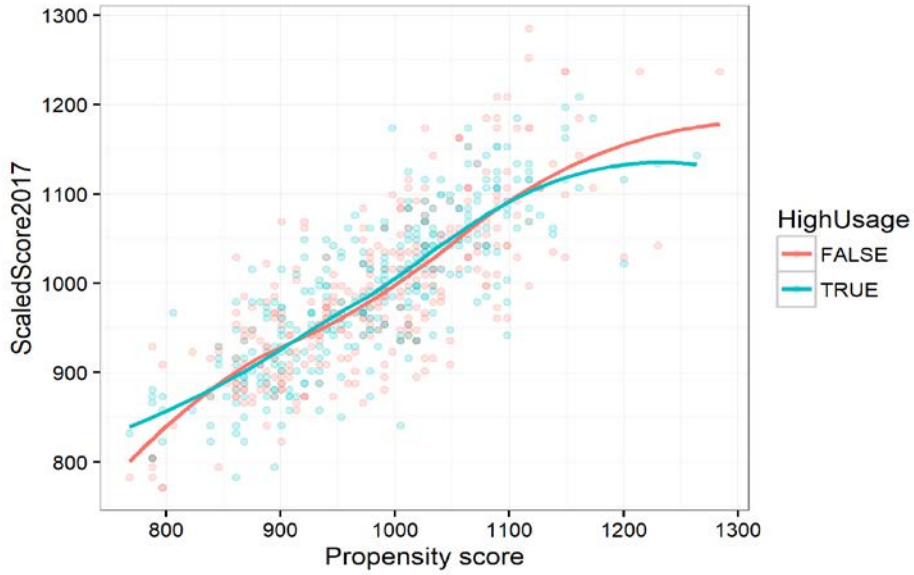
5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

Appendix C: Propensity Score Matching

ELA Grade 4

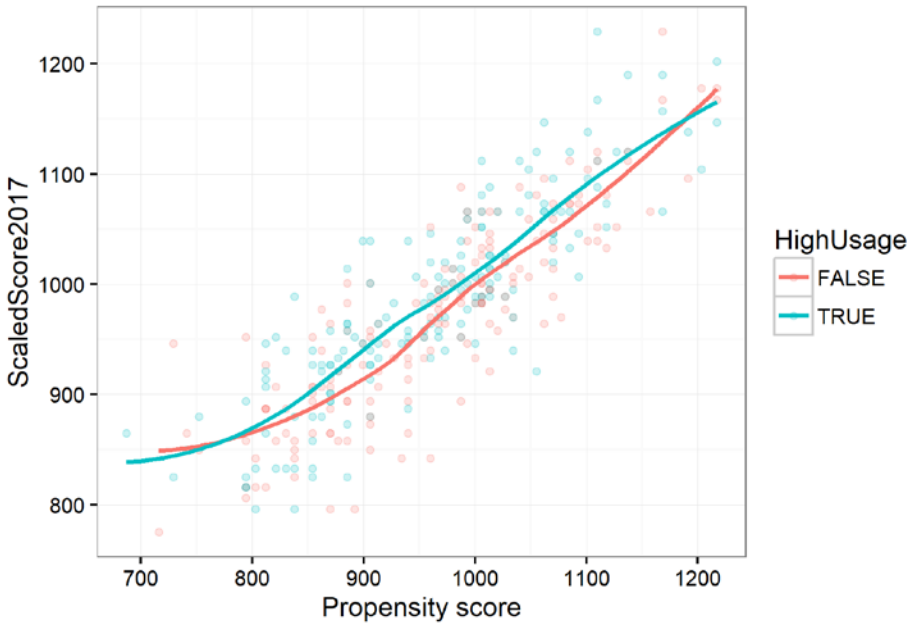


ELA Grade 5

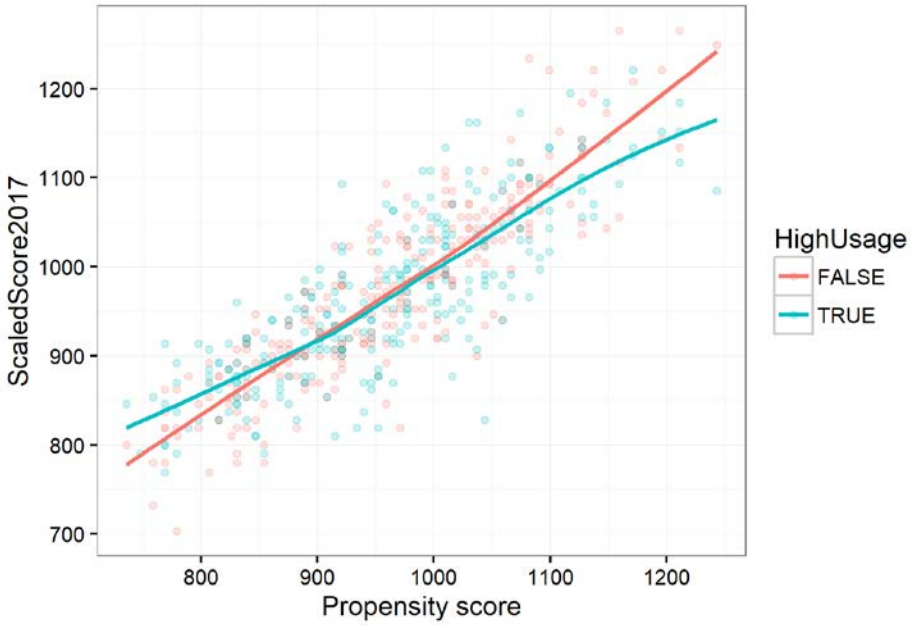


5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

ELA Grade 6

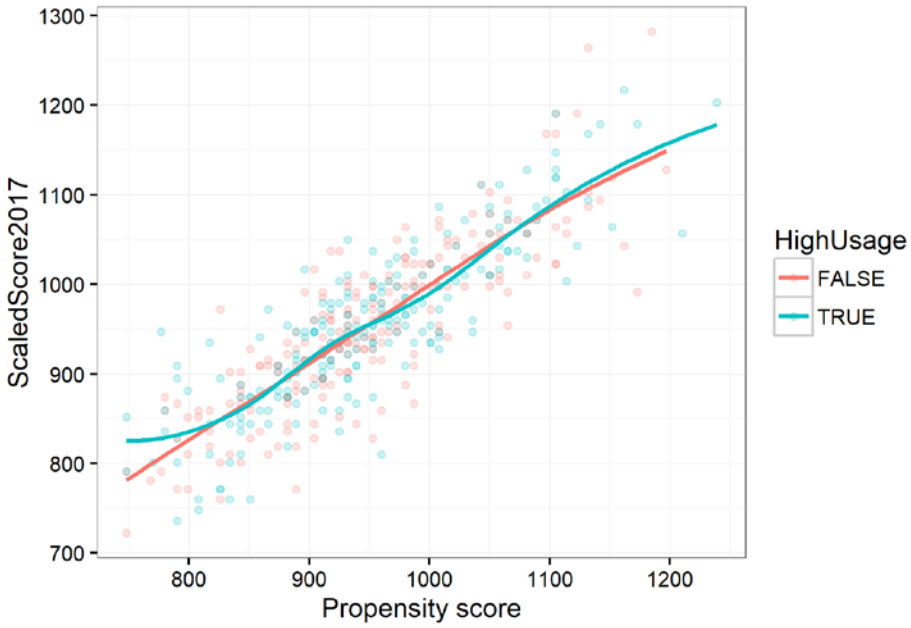


ELA Grade 7

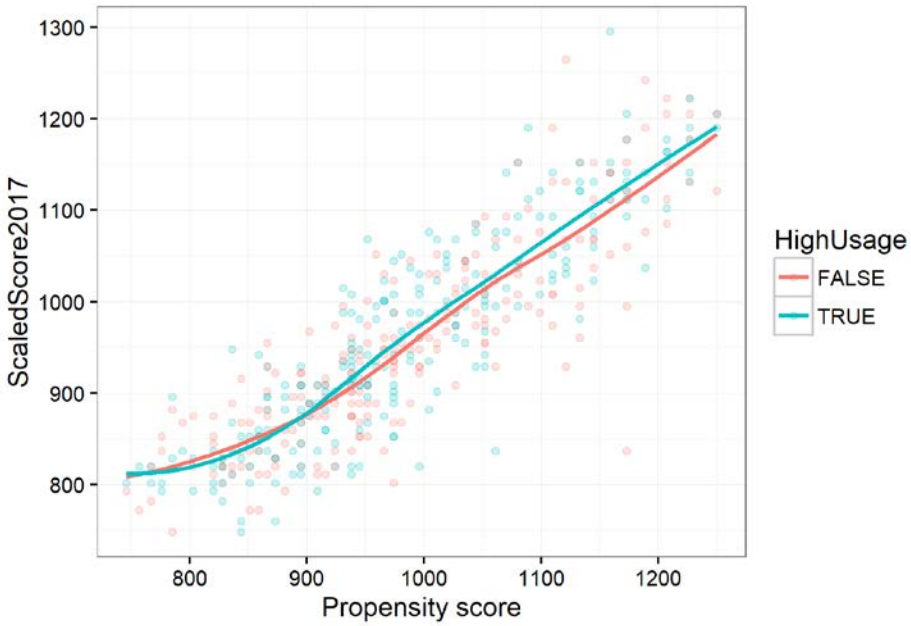


5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

ELA Grade 8

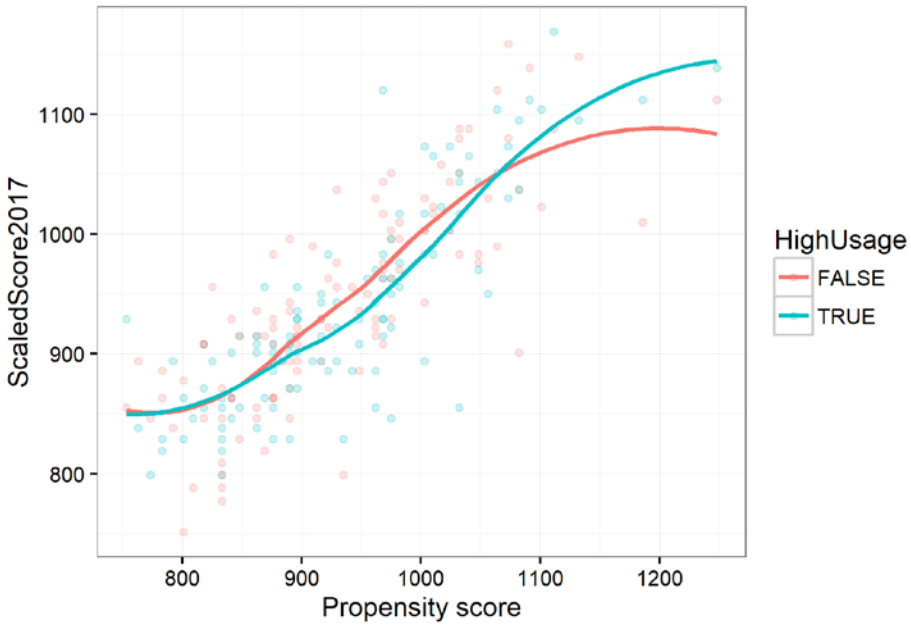


Math Grade 4

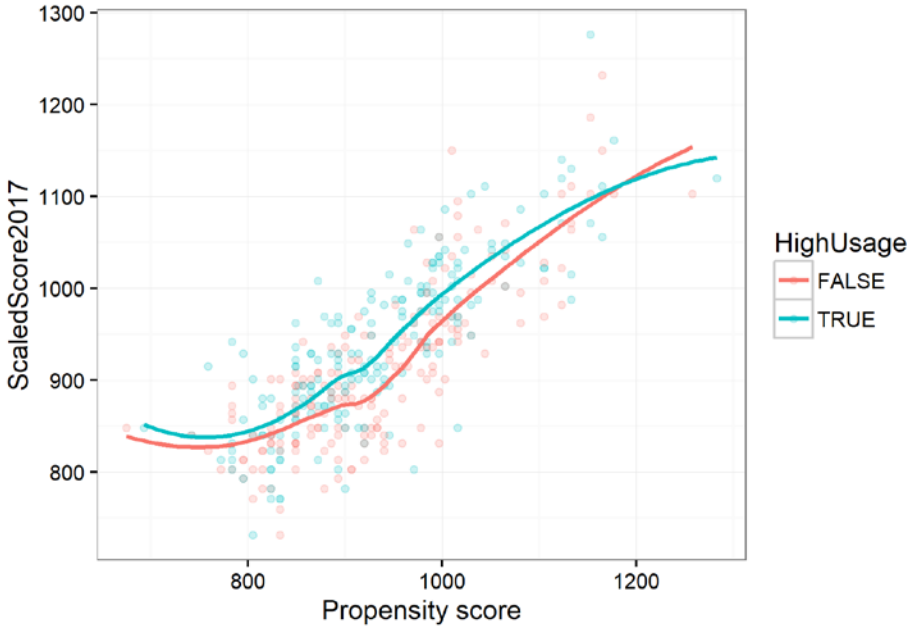


5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

Math Grade 5

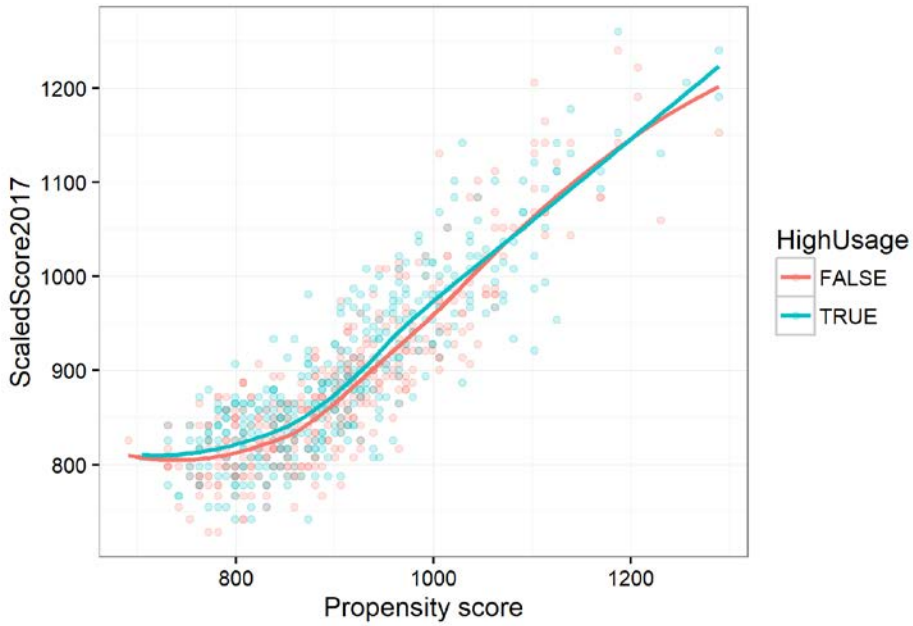


Math Grade 6

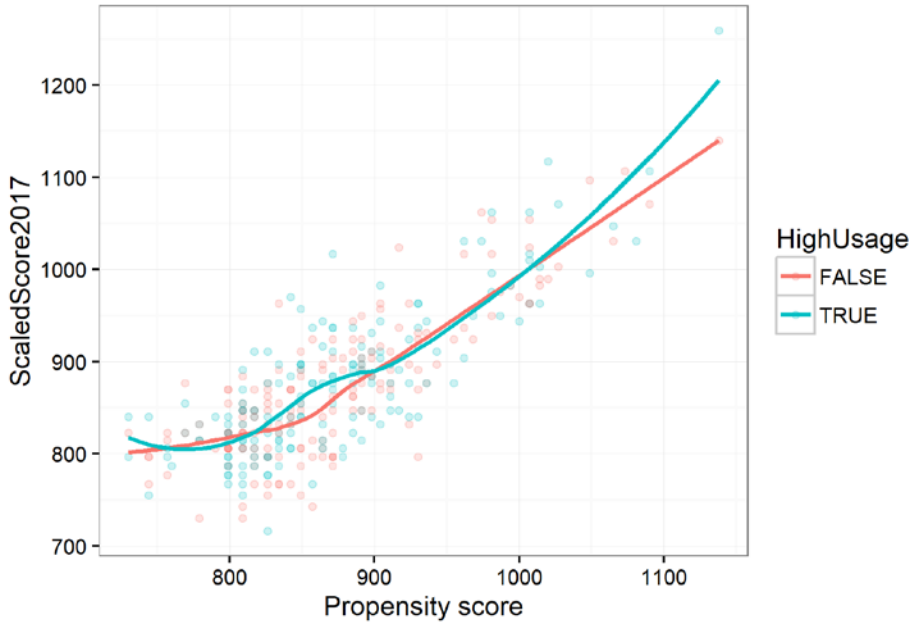


5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

Math Grade 7



Math Grade 8



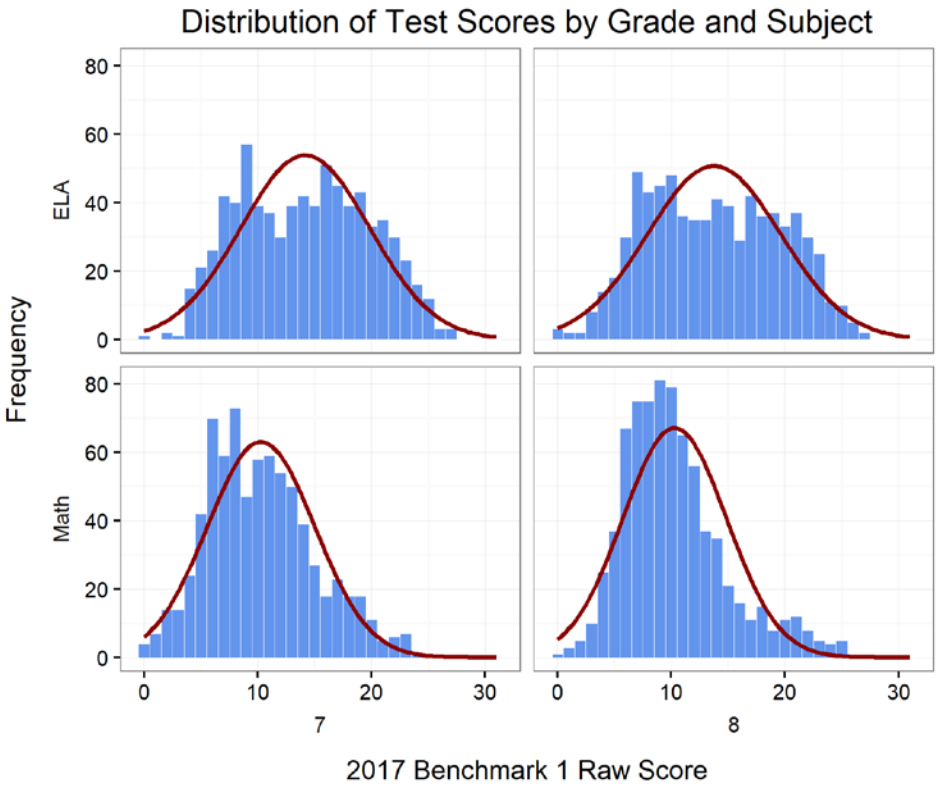
5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

Appendix D: Test for Normal Distribution of Scores

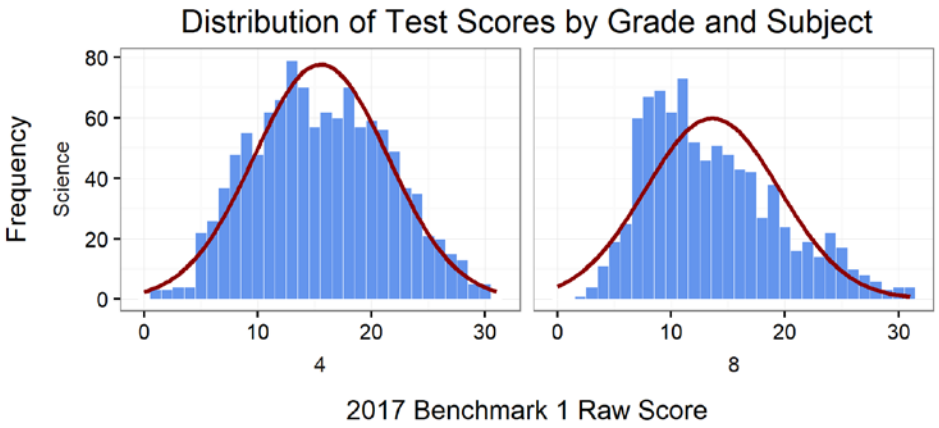
			Shapiro-Wilk	
Subject	**Grade**	**Assessment**	**Statistic**	**Sig.**
ELA	7	PSSA Scaled Score 2016	0.990	0.000
		PSSA Scaled Score 2017	0.987	0.000
		Z Score: PSSA ELA Benchmark 1	0.975	0.000
		Z Score: PSSA ELA Benchmark 2	0.982	0.000
	8	PSSA Scaled Score 2016	0.995	0.009
		PSSA Scaled Score 2017	0.998	0.399
		Z Score: PSSA ELA Benchmark 1	0.975	0.000
		Z Score: PSSA ELA Benchmark 2	0.983	0.000
Math	7	PSSA Scaled Score 2016	0.952	0.000
		PSSA Scaled Score 2017	0.915	0.000
		Z Score: PSSA Math Benchmark 1	0.978	0.000
		Z Score: PSSA Math Benchmark 2	0.980	0.000
	8	PSSA Scaled Score 2016	0.924	0.000
		PSSA Scaled Score 2017	0.942	0.000
		Z Score: PSSA Math Benchmark 1	0.941	0.000
		Z Score: PSSA Math Benchmark 2	0.968	0.000
Science	4	PSSA Scaled Score 2017	0.968	0.000
		Z Score: PSSA Science Benchmark 1	0.988	0.000
		Z Score: PSSA Science Benchmark 2	0.990	0.000
	8	PSSA Scaled Score 2017	0.961	0.000
		Z Score: PSSA Science Benchmark 1	0.955	0.000
		Z Score: PSSA Science Benchmark 2	0.972	0.000

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

ELA and Math Grade 7 & 8, Benchmark 1

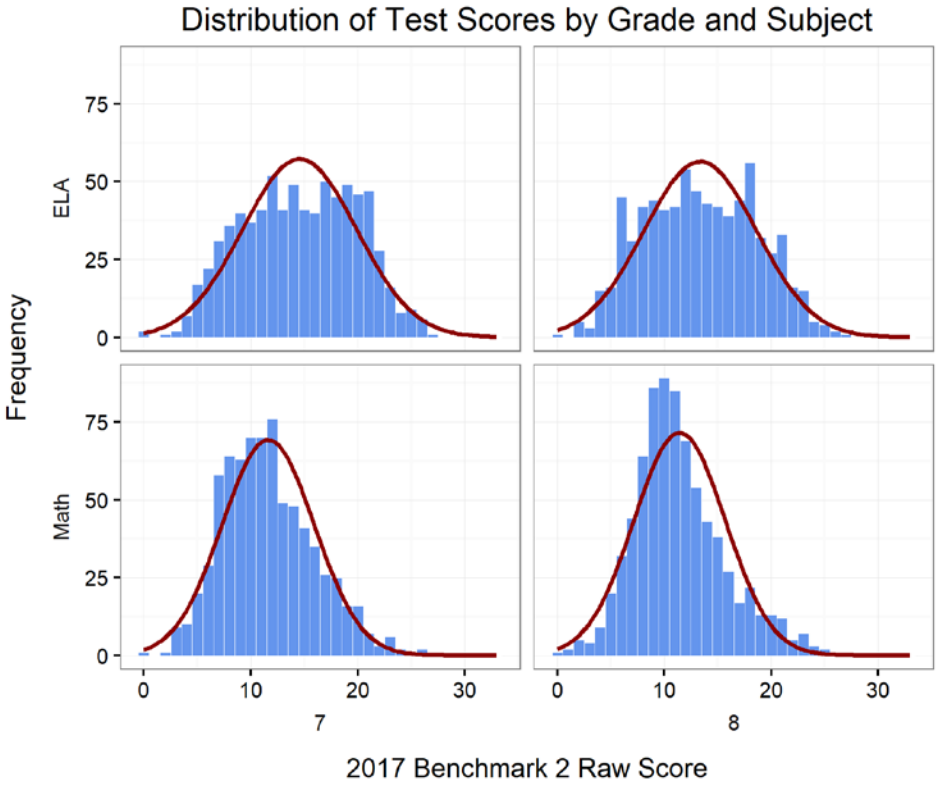


Science Grade 4 & 8, Benchmark 1

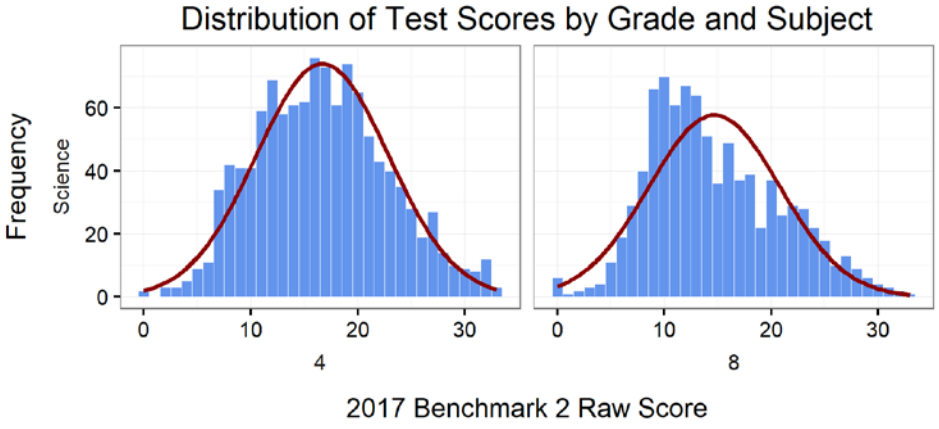


5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

ELA and Math Grade 7 & 8, Benchmark 2

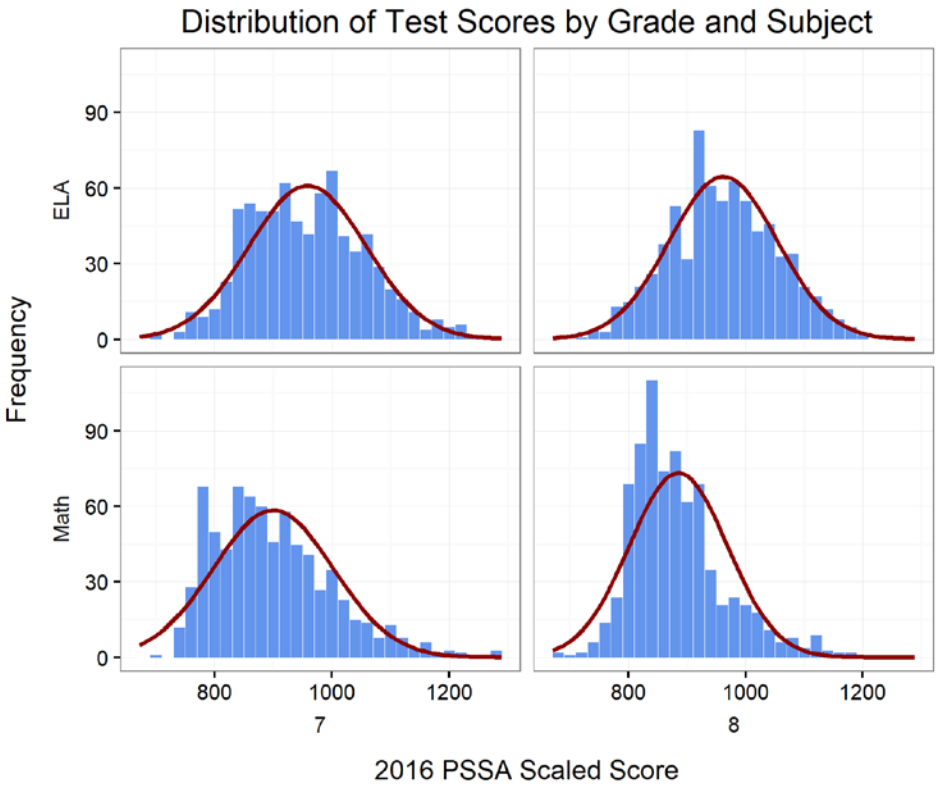


Science Grade 4 & 8, Benchmark 2



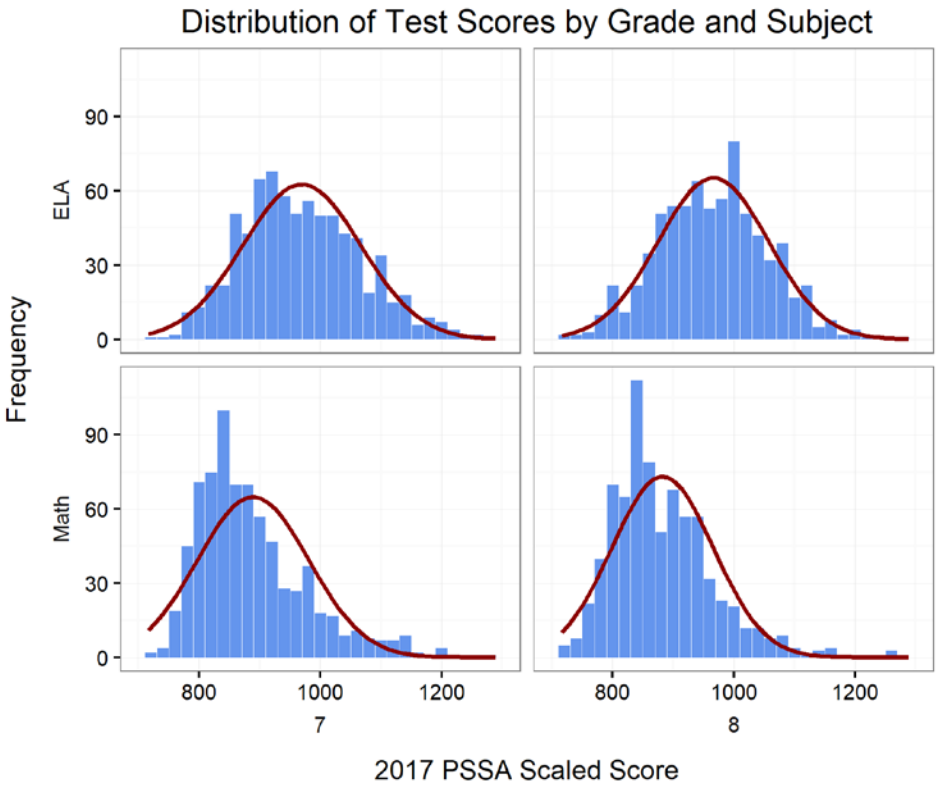
5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

ELA and Math Grade 7 & 8, PSSA 2016

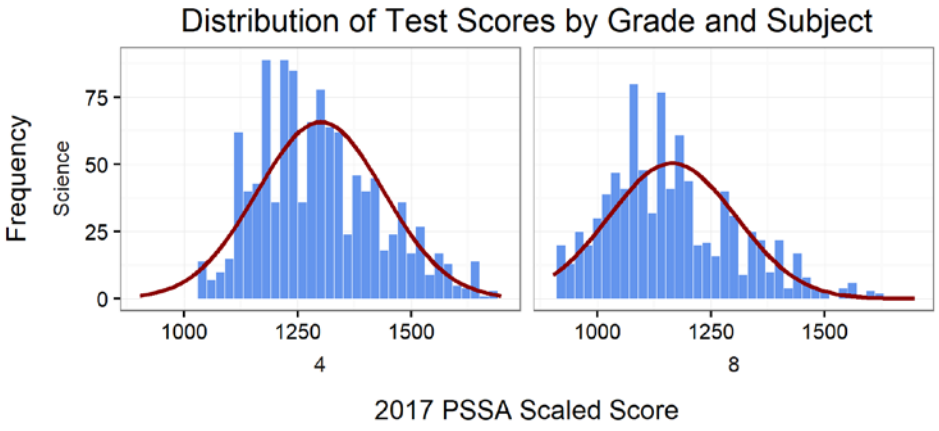


5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

ELA and Math Grade 7 & 8, PSSA 2017

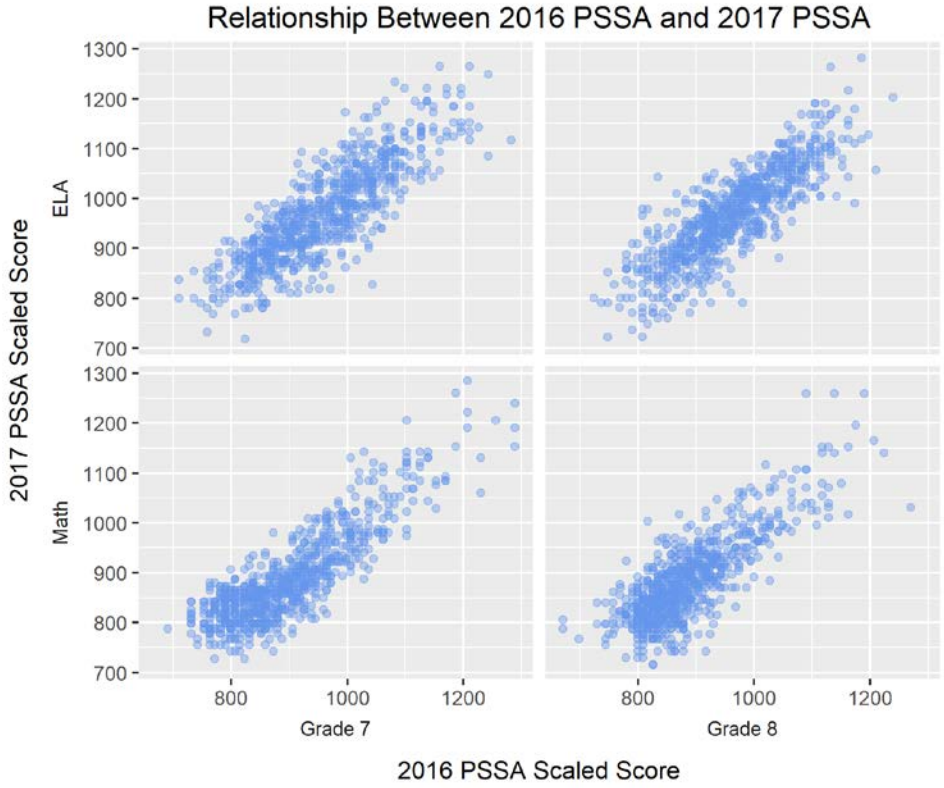


Science Grade 4 & 8, PSSA 2017

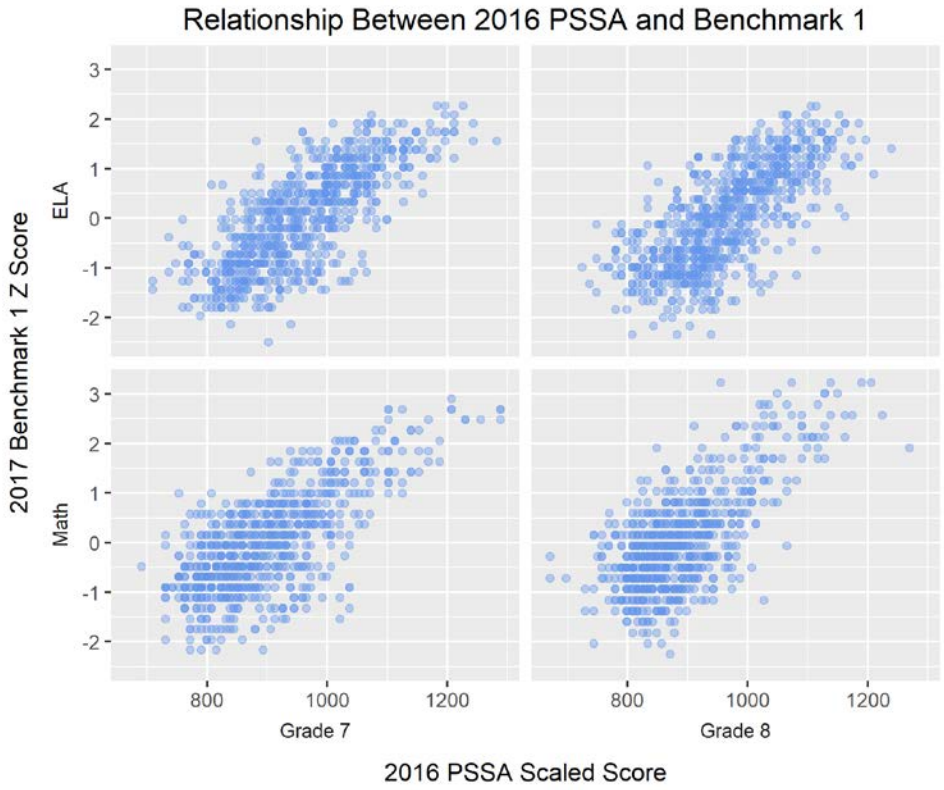


5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

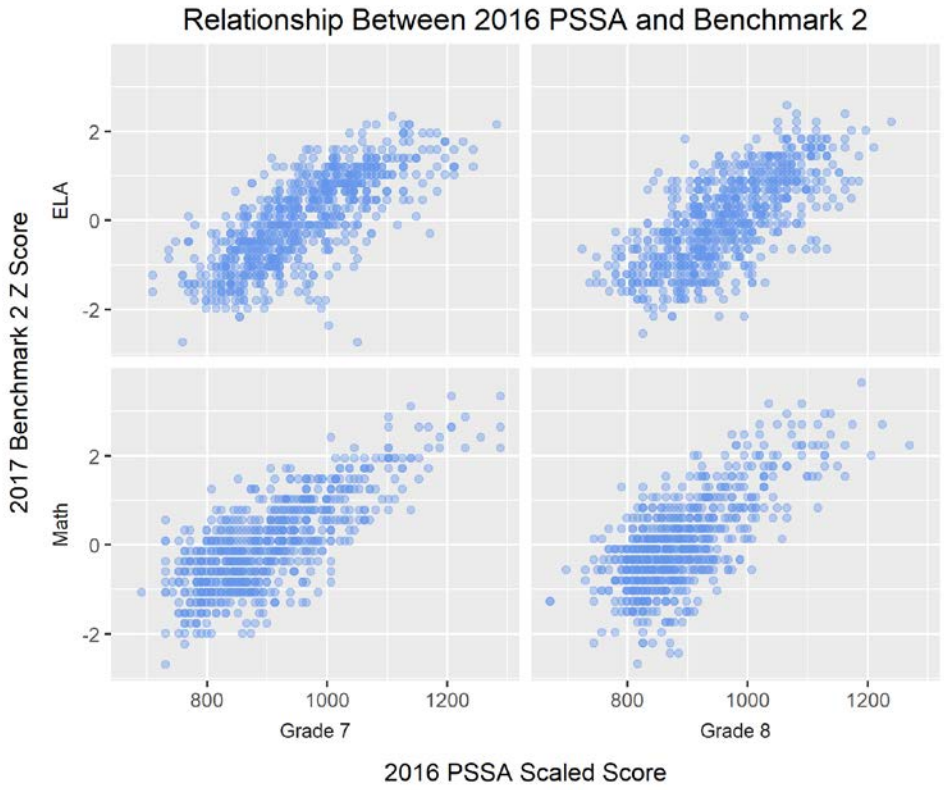
Appendix E: Scatterplots Showing Correlations between Study Island Benchmarks and PSSA Scores



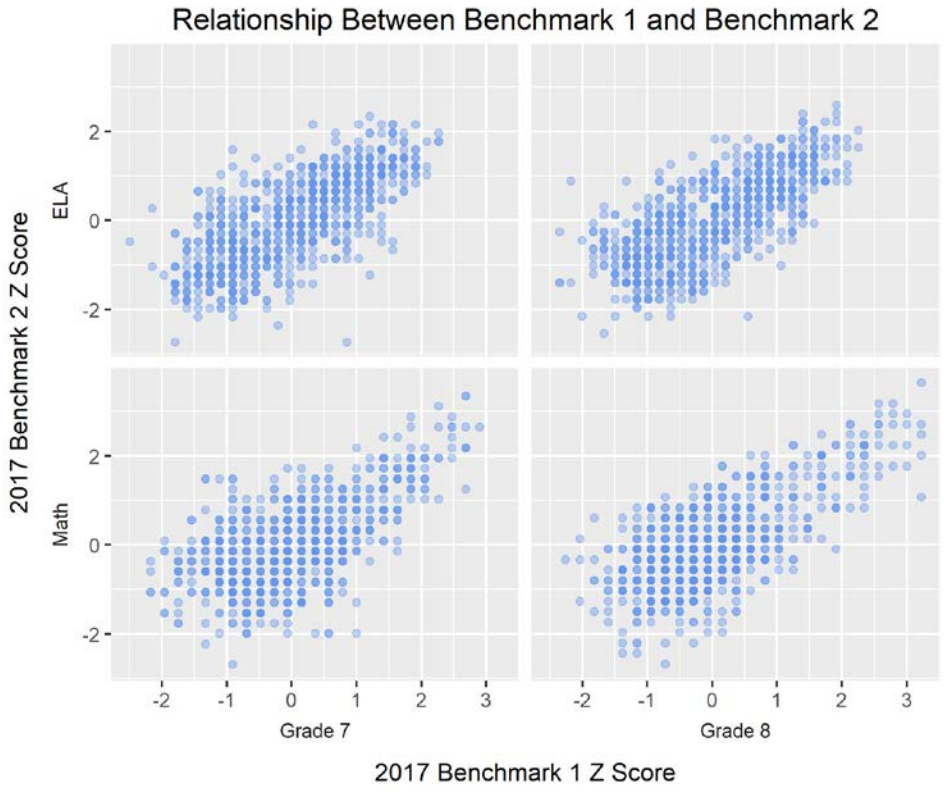
5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

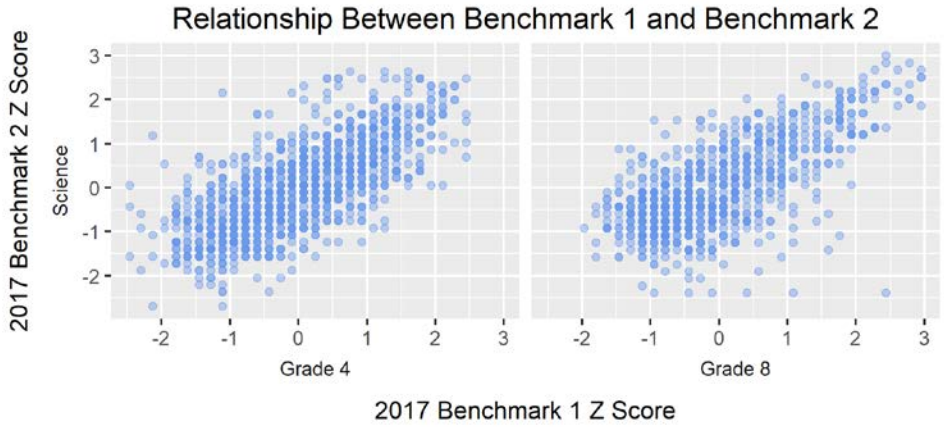


5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437

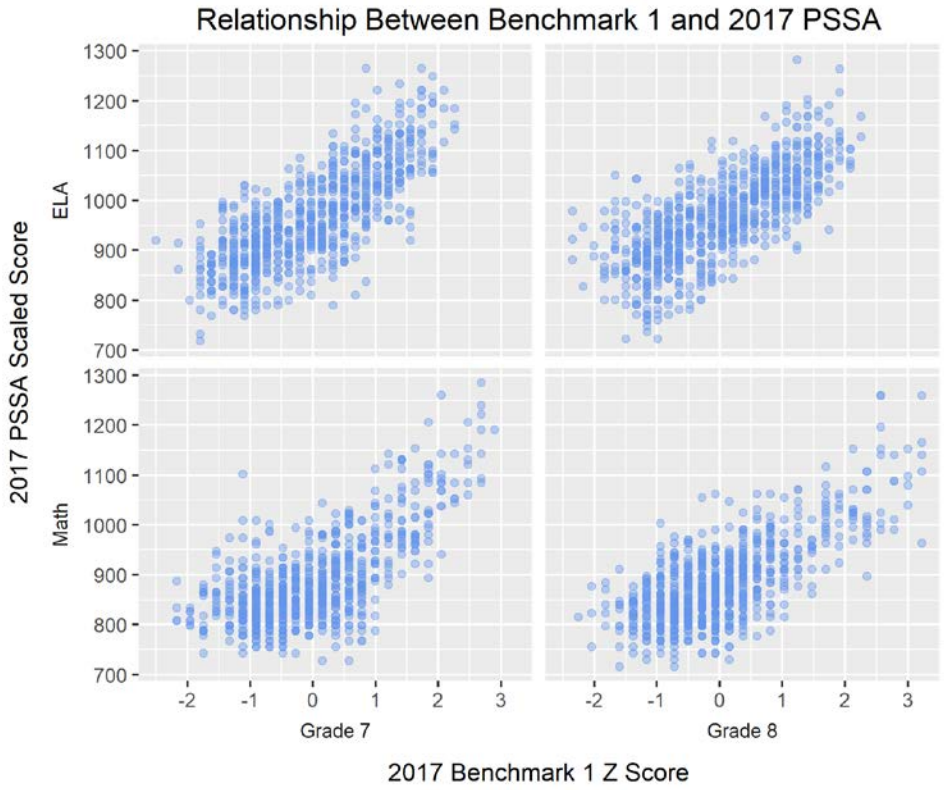


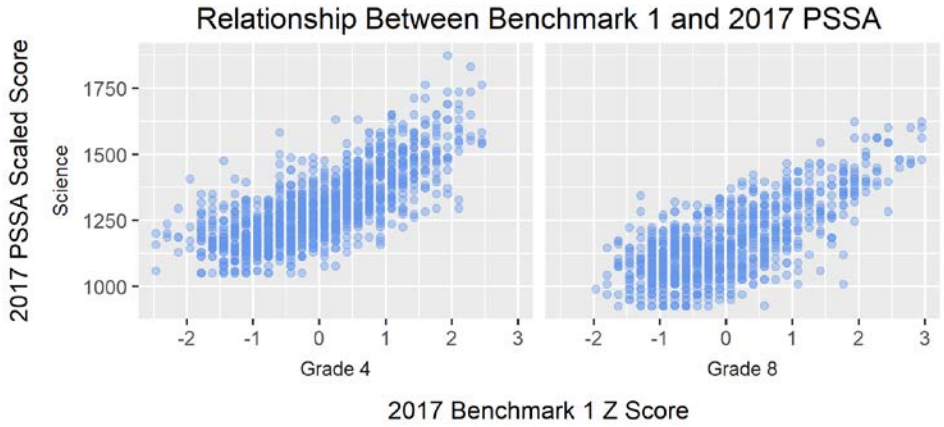
5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437



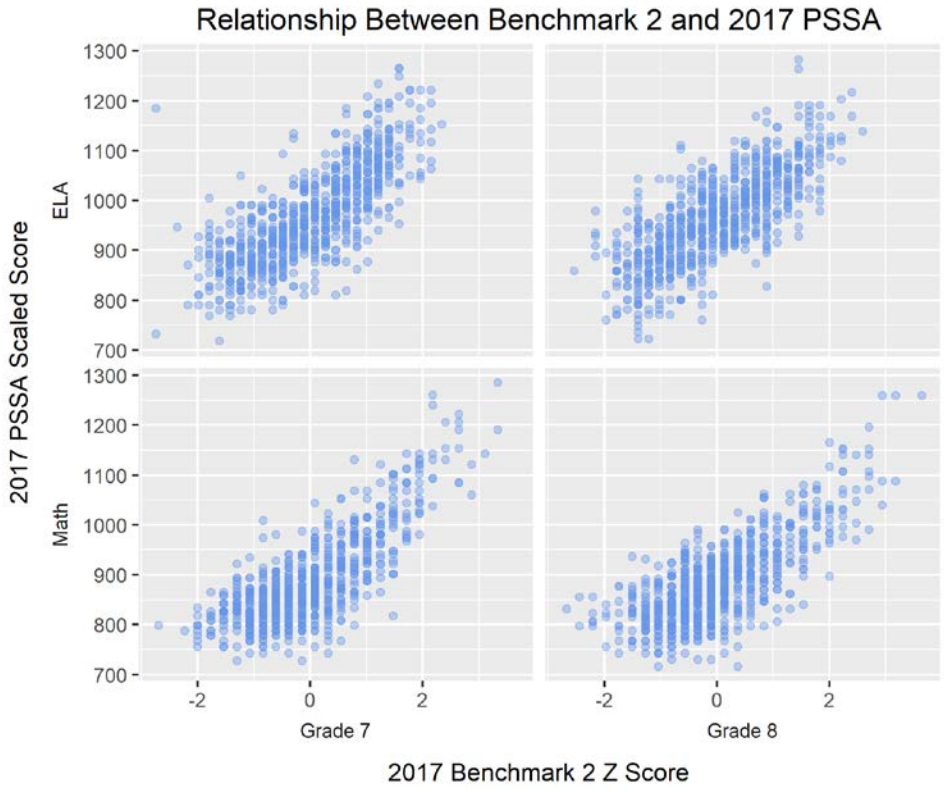


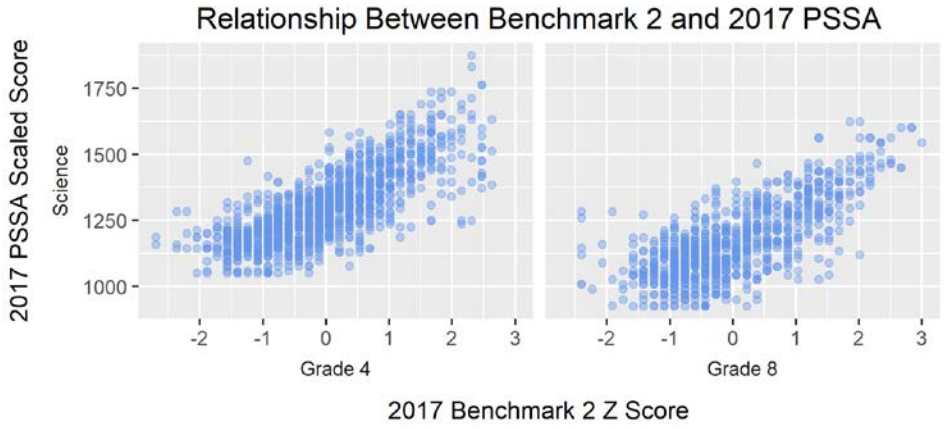
5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437





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





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