Examining the Efficacy of a Supplemental Program for Improving Student Performance toward Meeting District Graduation Requirements

Matthew Davidoff
Grace Kim
Hiyanthi Peiris
Giselle Saleet
Mela Still

May 5, 2017

Table of Contents

Executive Summary

The creation of No Child Left Behind (NCLB) in 2001 led to a movement towards holding schools accountable for student success. NCLB represents the first time that funding was tied directly to school districts demonstrating (via test scores) that students across different groups were progressing towards Annual Yearly Progress (AYP) (Cross, 2010). Each state, however, was able to set their own standards and timeline for all students to reach proficiency (Ryan, 2004). Because of the need for schools to demonstrate that students were improving on their standardized test scores, the Act created an incentive for states to create their own tests and develop systems to support their students to improve test scores.

Though NCLB's testing requirements broadened the market for Study Island, the program itself was created before those requirements came into effect. Study Island is part of an increasingly popular genre of educational technology, called Computer Aided Instruction (CAI), which is used throughout the country to support students in preparing for state education assessments. There is, however, a limited amount of research demonstrating the efficacy of such programs. This research seeks to improve our understanding of CAI's on student performance on three Keystone Assessments: Algebra I, Biology, and Literature. While Study Island offers dozens of services, this report only addresses the specific Keystone assessment supports that are used by the Neshaminy School District (NSD). NSD uses Study Island to prepare and benchmark students on their standardized state exams. In Pennsylvania, students across the state are required to take the PSSAs and Keystone Exams to fulfill federal testing requirements.

The Keystone Assessments are one component of Pennsylvania's current system of high school graduation requirements. It is required that all students score proficient or higher (defined as a score of 1500 or above) on Algebra I, Biology, and Literature to graduate high school; each exam can be taken up to three times until students reach proficiency (Keystone Exams Technical Report, 2015).

NCLB's supplemental educational services (SES) provision began a federally mandated after-school tutoring intervention for schools that do not meet adequate yearly progress and resulted in a proliferation of supplemental educational services (Burch, Steinberg, & Donovan, 2007). In response, there are currently thousands of supplemental education providers, like Study Island, throughout the country (Ascher, 2006). Those offerings, however, do not always address students' needs nor do all students have the same array of SES options to choose from; while the majority of students receive tutoring in small groups, a quarter of students receive one-on-one sessions, and others participate in online classes (Ascher, 2006). This paper focuses on one computer assisted for-profit supplemental educational services program: Study Island, which under-achieving students in Neshaminy School District are required to use.

There is, however, a lack of information about the effectiveness of these programs. Current information on effectiveness of SES is limited to a few internal evaluations done by larger providers of supplemental programs, like Study Island (Ascher, 2006; Burch, Steinberg, & Donovan, 2007; Munoz, Potter, & Ross, 2008). One meta-analysis of such programs found that the overall effects of these programs on student achievement is small; rather, the researchers find that the quality of the provider has the greatest impact on student outcomes and growth (Chappell, Nunnery, Pribesh, & Hager, 2011). Though the federal government mandates that states choose providers whose methods are research-based, more research needs to be done to see if the programs are effective

(Ascher, 2006). These results point to the need for more independent research on the individual programs.

This report examines the relationship between Study Island usage and Keystone Exam performance on the Algebra I, Biology, and Literature tests for students in grades seven through twelve in the Neshaminy School District. Neshaminy School District is a suburban community in Bucks County, Pennsylvania, located northeast of Philadelphia and currently includes about 8,400 students in total (Neshaminy School District [NSD], 2017). Students who score below proficient on their first test attempt are required to participate in the Study Island program.[1] We focused on the following questions in order to better understand the influence of Study Island on students' test performance:

1. Do academic outcomes (PA Keystone Assessments in Math/Literature/Science) improve more in one content area vs. another when using Study Island as a remediation tool in secondary education?
2. Do academic outcomes (PA Keystone Assessments in Math/Literature/Science) improve more for some underperforming students over others when Study Island is used as a remediation tool in secondary education?

The data used in this analysis was provided by the Neshaminy School District for school years 2013-2014; 2014-2015; and 2015-2016. After excluding students who were eligible for Study Island, but did not have a re-take test event, our final sample population consisted of 4,694 students. Of the sample of 4,694, 1,107 (or 23.6 percent) students participated in Study Island. Our analysis excluded students who were eligible for Study Island remediation (i.e., scored below 1500) but did not have a recorded test re-take, as we could not measure their growth across attempts.

Our empirical findings indicate that there was a significant increase in exam scores across retake attempts for Study Island Student scores on Algebra I. This increase in mean score, however, did not hold true for change in Literature and Biology Keystone Exam scores. Among students who took Study Island for Algebra I, White, female, male, not historically underperforming, and not economically disadvantaged students had statistically significant differences in mean scores between various test attempts, as did students without Individualized Education Plans (henceforth referred to as IEPs), and those who are not classified as English Language Learning (henceforth referred to as ELL). The differences were most pronounced for White, not economically disadvantaged, and non-ELL students, whereas some other groups also showed significant differences that their counterparts reflected (for example, females showed significant differences in test scores across attempts, but so did males). Because there were only significant differences between scores for the Algebra I test scores, the efficacy of Study Island is limited to one subject of the three examined. In the future we recommend that dosage (or length of time students spend on Study Island), supervision during programming, and extra support for historically underperforming students be investigated.

[1] Students who have an IEP (Individualized Education Plan) and score below proficient are given the option to opt out of participating in Study Island.

Introduction

It was not until the *No Child Left Behind Act* (NCLB) was created in 2001 that school districts were held accountable for individual students' demonstrated proficiency on standardized state examinations. NCLB aimed to maintain rigorous academic standards and eliminate the achievement gap while simultaneously holding stakeholders more accountable for these goals. NCLB revolutionized education reform by taking a technocratic approach and used data to drive its decision making to raise American students' international competitiveness.

NCLB was intended to mandate that all students in the United States, especially historically neglected groups (low income minority students, students with disabilities, and English Language Learners) achieve educational proficiency within twelve years. To achieve these goals, NCLB allowed each state the flexibility and authority to establish annual benchmarks for student proficiency. Furthermore, states were required to disaggregate student results by subgroup (racial minorities, socio-economic status, special education, and English Language Learners). For the first time local, state, and federal government would be able to identify where the gaps exist and how to fix them (Ryan, 2004).

The challenging standards were created to force schools, districts, and states to close the achievement gap between disadvantaged and minority students and their more affluent peers. The opposite was achieved, however, as schools quickly realized that these standards could be arbitrary without any reference to past achievement levels or rates of growth (Ryan, 2004). Because the states were allowed to set their own standard of "challenging" academic measures, achievement levels were not uniform across the country. Some states redefined "proficient" to mean that students met the minimal level of educational standards; other states decreased the difficulty of the tests to increase the number of students deemed "proficient" (Ryan, 2004). NCLB was designed to inspire states to raise academic standards to push students, unfortunately states did not have the proper guidance to create those rigorous standards.

Through Race to the Top (RTTT), the federal government wanted to encourage states and local governments to lead education reform. RTTT offered states the opportunity to compete for additional funds, instead of withholding funding, as previous legislation had done. The grants were awarded to states that changed their education system in four specific areas: revising standards and assessments that prepared students to be successful in their post-secondary endeavors and to compete in the global economy; using data to measure student growth and success to inform school improvement; effective teachers and principals; and turning around lowest- achieving schools (Duncan, 2009). In order for states to be eligible for this grant they had to adopt a nationwide standards initiative, the Common Core State Standards Initiative (CCSS).

The Common Core State Standards were created by the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO) to establish consistent educational standards across the states as well as ensure that students graduating from high school are prepared to enter credit-bearing courses at two- or four-year college programs or to enter the workforce (Common Core State Standards Initiative, 2017). Common Core detailed what kindergarten through twelfth grade students should know in English/ Language Arts and Mathematics by the end of each grade. The initiative grew out of a need for students in the United States to compete academically with their international counterparts and a need to lessen the existing gaps between different student groups within the country.

Attempts at Raising Graduation Rates

Though high school graduation rates have been rising throughout the past century, "NCLB renewed interest among researchers in estimating high school graduation rates because it made increasing high school graduation one of its goals and required states and schools to monitor them as measures of adequate yearly progress" (Heckman & LaFontaine, 2010, p. 244). And though National Center for Educational Statistics (NCES) data confirms that today, graduation rates are higher than they've ever been before, this growth is disproportional. Currently, the national average graduation rate is around 90 percent, though, in low-income communities, this rate is closer to 65 percent (Steinberg & Almeida, 2008). Schools on the lower end have clear incentives to address this issue. As Steinberg and Almeida stated, though NCLB and its proceeding accountability acts required increased graduation rates, "Neither the federal government nor the state governments [held] a blueprint outlining the steps to getting there" (p. 1).

One attempt to increase graduation rates has been to provide "...early and continuous support for struggling students" (Steinberg & and Almeida, 2008, p. 4). Standardized exams have been the main mechanism to locate students who fall behind, though states, districts, and schools all differ in how they assist these students. Common practices, as defined by Madden and Slavin (1987) are pull-outs (where a specialist pulls students out of their scheduled classes for specialized help) or small group instruction (where small groups work with a teacher after school or during lunch). However, these interventions require extra personnel, which is increasingly difficult to provide, given that the majority of states are providing less per-student funding today than they were at the inception of the Common Core (Leachman, et al. 2016). This tension between incentivizing extra services without proper funding has encouraged growth in the educational technology market, most notably, CAI's.

The standards of NCLB, in addition to being highly varied, were unable to prepare students for post-secondary endeavors. In 2013, ACT published an annual report on the number of students taking the ACT who met its college readiness benchmarks. This report stated that out of the 54 percent of all high school graduates who took the ACT, only 26 percent of them reached the college readiness level in all four areas tested (English, reading, mathematics, and science) (Conley, 2014). The introduction of CCSS aligned standards across the country and detailed what students should know in order to be successful after high school. CCSS attempted to remediate the inconsistencies of NLCB, through aligning standards across the country and detailing proficiency students need in order to be successful. CCSS identified the deeper cognitive processes and learning strategies needed to develop skills and knowledge necessary for college and careers (Conley, 2014). Many states used CCSS to develop and implement new state standardized test, with many states introducing comprehensive high school exams.

State Response

Following CCSS protocol, states have developed new standardized tests that have been integrated into graduation requirements for high school students. Specifically, Pennsylvania state administrators developed a new system designed to ensure that students are performing at levels equivalent to their peers nationally and are graduating with the knowledge and skills needed to be successful in their post-secondary endeavors. With Pennsylvania's Common Core State Standards at the helm, the Keystone State Assessment was developed as a graduation competency assessment. The Keystone Exams are one component of Pennsylvania's current system of high school graduation requirements and have replaced locally developed final exams in Algebra I, Biology, and Literature. Currently exams in Algebra I, Geometry, Chemistry, English Composition, Civics and Government, U.S.

History, and World History are in development. Students take the Keystone Exam when they complete the related coursework. For example, the earliest grade that a Pennsylvania student can take a Keystone exam is in 7th grade (Pennsylvania Department of Education, n.d.).
The assessment is meant to help school districts evaluate student proficiency. Pennsylvania state standards have been established to prepare students for postsecondary success in college or other post-secondary endeavors. School districts in Pennsylvania require all students to score proficient or higher (a score of 1500 or above) on Algebra I, Biology, and Literature on the Keystone Exam to graduate high school. The final score is then categorized as Advanced, Proficient, Basic, or Below Basic, and students need to achieve a score categorized as Advanced or Proficient, 1500 and above, to graduate. The Keystone Exam can be taken up to three times until proficiency or higher is obtained (Keystone Exams Technical Report, 2015).

This report examines Neshaminy School District's use of Study Island as the district's supplemental program to support the academic performance of high school students struggling to achieve proficiency on the Keystone Assessments and help students meet the district's graduation requirement. Neshaminy School District covers 27.6 square miles northeast of Philadelphia, PA in Bucks County, PA and educates approximately 8,400 students in one of ten schools within the district (NSD, 2017). The District's high school, Neshaminy High School, which most students in the sample attend, currently serves 2,533 students in grades nine through twelve (NCES, 2016b). In the district, 23 percent of students have Individualized Education Program, and 1 percent are of Limited English Proficiency. Like many suburban school districts, Neshaminy High School is predominantly White (86 percent) (NCES, 2016b). Furthermore, of the current student body 21 percent are eligible for free or reduced price lunch (NCES, 2016b). The intended use of Study Island for Neshaminy School District is to provide remediation for academically struggling students and to help students meet the graduation requirement.

Summary of Research Questions

The purpose of this study is to examine the efficacy of a supplemental remediation program entitled Study Island, on increasing the passing rate of high school students in the Neshaminy School District on the Keystone Assessment. Therefore, the researchers were presented with the following research questions:

1. Do academic outcomes (PA Keystone Assessments in Math/Literature/Science) improve more in one content area vs. another when using Study Island as a remediation tool in secondary education?
2. Do academic outcomes (PA Keystone Assessments in Math/Literature/Science) improve more for some underperforming students over others when Study Island is used as a remediation tool in secondary education?

The overarching goal behind both these questions was to assess the effectiveness of Study Island as a remediation tool for students who had not gained proficiency in Algebra, Biology, and/or Literature Keystone examinations in the Neshaminy School District. The dataset listed student information by subject test taken by year, which allowed for analysis of change across test attempts for each subject area. In order to answer the research question, the researchers took the following approach:

1. *Do academic outcomes (PA Keystone Assessments in Math/Literature/Science) improve more in one content area vs. another when using Study Island as a remediation tool in secondary education?*

The researchers divided the dataset into two groups of students--one group of students who reached proficiency in their first test attempt (i.e., scored 1500 or above) and the other group of students who did not. For the purposes of answering this research question, the researchers focused on the group that did not reach proficiency in their first test attempt. The researchers were able to calculate mean (i.e., average) test scores for each subject across the test attempts to analyze how students showed change in test scores across the number of attempts made. The researchers used paired samples T-tests to compare mean test scores across different attempts of the same group of students to determine statistical significance. The researchers were informed by the Neshaminy School District that Study Island was used as a remediation tool for students between test attempts. However, changes in test scores cannot be attributed solely to Study Island, as there was no mandated or fixed dosage to allow for an accurate comparison and the study conducted was only an observation of how test scores changed.

2. *Do academic outcomes (PA Keystone Assessments in Math/Literature/Science) improve more for some underperforming students over others when Study Island is used as a remediation tool in secondary education?*

Researchers analyzed the change in subject test scores by different demographic groups to allow for comparisons between groups. In order to analyze mean test scores across test attempts, the researchers used paired samples T-tests to compare mean test scores and determine statistical significance. Similar to the first research question, changes in test scores cannot be attributed solely to Study Island, as there was no mandated or fixed dosage to allow for an accurate comparison and the study conducted was only an observation of how test scores changed.

Demographic group comparisons

In addition to the research questions addressed above, the researchers also compared demographic characteristics of students in Study Island (i.e., students who did not score a proficient test score on their first attempt) and students who are not in Study Island (i.e., students who scored a proficient

test score on their first attempt). Researchers conducted proportion significance testing to analyze which demographic groups were over or under represented in either group (Study Island versus Non-Study Island).

Competitor analysis

The researchers also conducted literature research of similar CAIs and provided comparative information on Edgenuity, IXL, Achieve 3000, iReady, Wowzers, and Scootpad. The researchers compared these CAI offerings with Study Island on subjects offered, price range, and grades for which remediation was offered.

Description of Study Island

What it is

Study Island is an interactive, subscription-based educational software service, part of a genre known as Computer-assisted instruction (CAI). CAI's take a different approach to online education than Massive Online Open Courses (MOOCS). MOOCS maintain instruction at their core, and are usually structured like traditional classes, featuring lessons in a particular order. CAI's, however, are centered around assessment and evaluation, individually tailored to a student's needs.

While Study Island offers a vast array of subjects, its main objective is to support students from kindergarten to twelfth grade master specific, grade level Common Core State Standards, in a "fun and engaging manner" (Study Island, 2017b). Study Island uses computer adaptive lessons, practice tests, and games to do this. Study Island also offers benchmarking capabilities, giving teachers real-time assessments of student levels on state tests before the exams are administered. In addition to state exam alignment, Study Island offers an ever-expanding assortment of products, including courses in Math, English Language Arts, Science, History, ACT, SAT, Graphic Novels, College and career readiness, the works of Shakespeare, AP Examination Preparation, College transition lessons, and GED preparation (Bernard, 2013).

Study Island provides services to districts, schools, and families (usually those of homeschooled children) and differentiates cost depending on the number of students, number of subjects, length of contract, and access to technical features, like benchmarking and exam creation. Study Island typically costs anywhere between $2 and $15 per student per month (Study Island, 2017a).

In Study Island, students answer practice questions with real-time feedback and explanations are provided for questions answered incorrectly. The software is adaptive, so students who score well receive increasingly difficult questions, and those who answer questions incorrectly are moved to the "base" questions of the topic. For example, if an Algebra Study Island student misses several bivariate Algebra questions, the software will sense this and bring the student back to univariate questions. Study Island refers to this as a "short cycle assessment feedback loop," claiming it is at the core of Study Island's differentiation for individual students (Laing, 2011). "The Study Island program records statistics for each user session in a real-time report card. These statistics measure progress, streamline the learning process and can be customized by student, subject, class, grade and school. In this system, both the teacher and the student can vary the learning style in which the questions are delivered" (Bernard, 2013, p. 10). "The software is also designed so that teachers can differentiate instruction. Options include setting a student on a lower or higher grade level to study the topic being covered, offering text-to-speech for students with reading difficulties, and setting the number of answer choices for the students" (Shoemaker, 2013, p. 33). Also, based on research conducted by Magnolia Consulting (Styers, 2012a), contracted by Edmentum (the parent company of Study Island), though keeping students engaged is often a challenge to CAI, Study Island's integration of games (which students can use to compete against their friends) and "blue ribbon" rewards keep students engaged in Study Island lessons (Laing, 2011).

In addition to online lessons, Study Island enables teachers to create printable worksheets, quizzes, and exams based on Common Core State Standards, other Study Island content areas, and student or class strengths and weaknesses. Parents are able to log on to Study Island to check their child's progress. Teachers and school leaders are able to compare between students and across classes. This creates opportunities to gauge how an individual student is performing, how a class is performing, how a class compares to other classes in the school, and potentially how a school compares to others in the district (Benthall, 2015).

A final aspect of Study Island noted by Bruce-Simmons (2013), is that it meets the four basic principles set forth by the *No Child Left Behind Act* (NCLB): stronger accountability for results (made easier by Study Island's data gathering capabilities), increased flexibility and local control (through Study Island's differentiated lessons), expanded options for parents (giving parents the ability to monitor students' homework completion), and an emphasis on teaching methods that are proven to work (a bit less clear, as Study Island is still relatively young and research is inconclusive). For these reasons, as well as the details listed above, Study Island has become popular in schools and districts across the country.

SI History

Study Island was founded by Cameron Chalmers, a computer scientist, and David Muzzo, an economist and marketer, in 2000. It was developed in conjunction with the Ohio Department of Education to prepare students for the Ohio Proficiency Test Program (Archipelago Learning, 2010). It was initially intended to help Ohio schools reach their state standards (Bracht, 2011). After expanding to nearly half of the school districts in Ohio, Study Island began to expand across the United States. Today, Study Island is used in all 50 states, the District of Columbia, and in several regions of Canada (Archipelago Learning, 2010). Study Island's rapid expansion was driven largely by NCLB's reporting requirements.

NCLB created a beneficial environment for Study Island to thrive. For example, provisions in NCLB required schools to demonstrate annual growth on state exams. Because Study Island had experience with Ohio's exams, they were able to more easily align their content with other state exams and relieve teachers of the additional responsibility of creating test prep materials. Furthermore, NCLB's accountability requirements mandated that schools report student progress disaggregated by specific student demographic groups to districts. Study Island's software enabled teachers and schools to easily share such data with districts (Bracht, 2011). The requirements of NCLB led to many schools and districts receiving grants geared towards academic improvement, often directed towards "proven techniques" or specifically directed at technological supports. This resulted in a proliferation of "commercial companies [like Study Island] market[ing] their products to school leaders in an attempt to satisfy the growing demand for empirical results in learning, promising higher test scores if students use them" after the implementation of No Child Left Behind in 2001 (Liang, 2012). In comparison to drastic changes to the curriculum or personnel improvements, Study Island was relatively inexpensive (Dube, 2011). These factors contributed to Study Island's expansion and were complemented by the realization that continuous formative assessments (like those offered in Study Island) would help teachers and schools better track student progress throughout the year in preparation for their summative exams, thus informing educational policies on how to improve scores over time (Bracht, 2011).

As Study Island's reach increased, the founders began to develop and acquire other subscription-based products, such as Reading Eggs and ESL ReadingSmart. Archipelago Learning was founded as the umbrella organization, which was acquired by Edmentum in 2012. Today Study Island offers an array of courses and subject, far beyond standardized test preparation, and is used by approximately 11 million students across North America (Study Island, 2017a).

How it Works

Study Island's use varies widely across schools, districts, and states. In Pennsylvania, it is most often used for state exam preparation and benchmarking (giving teachers a snapshot of their students'

progress before state tests are administered). This use of Study Island is common though some schools only purchase subscriptions for students who have scored below grade-level, while others purchase subscriptions for all students. Also, some schools keep Study Island subscriptions year-round while others get them once a year, specifically to prepare for standardized exams.
Of those using Study Island, some teachers assign specific each night, while others merely require a weekly time commitment for Study Island homework activities. Other schools, like those in Neshaminy, provide a computer lab for Study Island use. While some schools have scheduled Study Island sessions, Neshaminy encourages students to work on Study Island during their free periods. A common limitation addressed in most research on the efficacy of Study Island is the variability of home access to a computer (Bract, 2011; Dube, 2011; Laing, 2011; Viviano, 2011). Neshaminy overcomes this limitation by offering Study Island in a computer lab, though this still gives students with home access an advantage.

In Neshaminy, Study Island is used as an educational supplement, only being offered to students who need subject-specific test preparation. Additionally, Study Island participation does not at the expense of other instructional time; students are not "pulled out" for Study Island. Indeed, this is not always the case. Viviano (2011) researched the use of Study Island in vocational schools where Study Island lessons were provided to integrate academics into a largely Career and Technical based curriculum. Dube (2011), on the other hand, wrote of his own experience as a teacher, where he used Study Island in class because the textbooks in his Michigan school were out of date. Finally, Parlapanides (2008) researched Study Island in a school where students could opt out of physical education courses in order to use Study Island for one class period (44 minutes) per week. Though this is not an exhaustive list of Study Island's use across the country, in demonstrates the variability in how it is assigned.

Study Island is frequently used for accountability measures as well. For example, teachers use Study Island assignment participation or assessment performance to grade their students. Also, many school leaders and districts require Study Island assessment results from teachers to ensure students are progressing and meeting set achievement goals.

Competition with Study Island
Who's Involved. Study Island is one of many (20 to 30) CAI's, all of which provide children with supplemental and alternate forms of instruction, engagement, and assessment. Each of these CAI's share a few similarities: all contain content separated by subject and grade-level; all utilize a system of lessons, exercises, and games; and all are meant to supplement classroom content, not to provide content as MOOCS do. Viviano (2011) refers to CAI's as "drill and practice" platforms. She asserts that, while there is no evidence of the efficacy of computers alone, the results of computers as a supplement have been "overwhelmingly positive" (Viviano, 2011).

While some CAI's focus on one specific subject, like *MathSeeds* or *SpellingCity*, most are similar to Study Island in that they offer several subjects and align their lessons to state exams, putting them in the same field as Study Island.

How the Field Has Progressed? Dube (2011) notes that, as NCLB provisions set a national tone for accountability and Common Core mandated a standardized version of that accountability, Study Island gained appeal with schools and districts. This is true of most CAI's largely due to their ability to provide data (which is used as an accountability measure) and to provide standardized (yet personalized) content (i.e., students in Neshaminy receive the same Algebra lessons as those in

Pittsburg). That said, it is worth noting that many CAI's were founded as startups in Silicon Valley in the 1990's. It wasn't until many years later that large tech and education corporations began to get involved in the CAI field. Notably, *Aleks* (a math CAI) was founded in Irvine in 1994, and acquired by McGraw Hill in 2013 (aleks.com). Similarly, *TenMarks* was founded in Foster City, California, and acquired by Amazon.com in 2013 (tenmarks.com).

Who's the Best? It is difficult to compare across CAI's because there is so much variability in their services, their potential effects, and their accessibility. Please see *Appendix A* for a comparison of several prominent CAI's, including Study Island, organized by price and functionality. Due to limitations expressed later in this paper, however, assessing which CAI is most effective is beyond the scope of our research. As demonstrated in this table, Study Island is extremely competitive, based on the low cost and high number of services offered.

Do CAI's Work? As CAI's are still relatively young, few studies have been conducted to determine which are more effective, or, more importantly, if CAI's are effective at all. A 2009 Mathematica-conducted meta-analysis (Campuzano, et al.) found no significant value-added from CAI mathematics and reading instruction. This study included matched-comparison data from 9,458 students in 428 classes from 33 districts in three states (California, New Mexico, and Iowa). In this study, participating teachers opted to include CAI instruction in their classroom practices. These classes were then matched with similar classes who had not used any CAI instruction. These comparisons, which included 16 CAI's --such as *Plato Focus*, *LeapFrog*, and *Destination Reading*—found no significant impact on state test scores. It is worth noting that these studies all required students to lose traditional instructional time in exchange for CAI participation, which (as noted above) is not the only manner in which CAI's are used. Also, it is not indicated in this analysis whether these CAI's were aligned to state standards, as Study Island is.

In contrast to this study, however, another 2009 study (Barrow, Markman, & Rouse), which examined matched comparison results from 1,873 pre-Algebra students in three states, found CAI's to be beneficial overall. In this study, state test scores from classes who had used only CAI instruction were compared to traditionally taught classes. Overall scores were stronger in CAI participants, which the authors attribute to the individualized instructional aspects of CAI's. That said, the authors note that the scope is limited (only pre-Algebra) and should, thus, not be used to generalize across CAI's or across subjects.

The discrepancies between these two studies, the variation of services among CAI's, and the variables in CAI use all demonstrate a need for much more research before any reliable conclusions can be made about CAI efficacy.

Does Study Island Work?

The research regarding the efficacy of Study Island is inconsistent; some research indicates that it is not effective, while others find that it has an influence on students' results. There is very little information on the three subjects that the Neshaminy School District uses, and of Biology, Literature, and Algebra I, much of the research focuses on the last. Bruce-Simmons (2013) investigated the impact of computer-assisted instruction mathematics achievement (using Study Island) of underachieving fifth-grade students. Students who did not "achieve standard" the Palmetto Assessment of State Standards (PASS) Test at the end of fourth-grade were asked to participate in a year-long academic assistance program (Bruce-Simmons, 2013, p. 72). Bruce-Simmons (2013) compared two years of students' math PASS scores – one year before the

implementation of Study Island and the year after the program was implemented. Students who scored above basic, but not proficient were scheduled into the computer-assisted instruction labs once a week, while those who scored below basic were scheduled to use Study Island in the labs two or more times a week (Bruce-Simmons, 2013). The study found no significant different in fifth-grade math achievement (understood as students' PASS scores) for those students who used Study Island just once a week (Bruce-Simmons, 2013). There was, however, a significant difference in the math scores of students who received the computer-assisted instruction two or more times a week (Bruce-Simmons, 2013).

This highlights one of the limitations of our understanding of Study Island: the influence of dosage and how to track student engagement with the program. One study of Study Island's effect on ninth grade students' Algebra scores found that regular use of Study Island (for 90 minutes each day for ten weeks) had a statistically significant influence on increased mathematic assessment scores (Ramsay, 2014). Ramsey (2014) used a quasi-experimental nonequivalent (pretest and posttest) control-group design - students who participated in Study Island were compared with a control group who received traditional Algebra instruction (Ramsay, 2014). The study is somewhat limited in that just 28 students were assigned to work with Study Island during this time period and therefore the researcher was not able to include high achieving or special needs students, or disaggregate the data by student demographics (Ramsay, 2014).

Meehan (2016) found that Study Island's remediation program did not increase student scores for the students in one Pennsylvania high school compared with students who were not in the remediation program. Similar to the program's use in Neshaminy, in this study students who scored below proficient on the May Algebra exam were placed in the Math Lab course in order to improve their scores by the time they retook the exam in January (Meehan, 2016). Meehan's (2016) results are in agreement to those found in this study and point to the need for more independent research into the efficacy of Study Island on other subjects (especially Literature and Biology).

Limitations of Evaluating Study Island and Other CAI's

Dosage. Throughout the research on Study Island, participants are often referred to as those "who did Study Island" or "those who had Study Island." This implies that exposure to Study Island is binary, rather than differentiated by dosage (Bract, 2011; Dube, 2011; Laing, 2011; Viviano, 2011). Therefore, it is assumed that participants can be considered a treatment group and non-participants can be considered as a control group, allowing for them to be compared to test the efficacy of Study Island. However, as mentioned above, Study Island is less of a "course" and more of a collection of exercises. In the studies cited in this paper, very few researchers were able to control for dosage. As an educational supplement, Study Island is usually assigned as homework, and, it is therefore difficult to ascertain that all students with access used it in the same way or for the same amount of time. And, though Study Island specifically mentions the teacher's ability to monitor each student's usage, this usage data appears is not included in most (including our own) analyses.

Bernard (2013) discusses this limitation by saying, "The participants in the study might not all have had equal access to the program Study Island, which can be accessed, at any time, through nearly any device that has internet. This might put students who do not have access to the internet outside of the school day at a disadvantage" (p. 14). Additionally, even those who do have access to appropriate technology may not be using Study Island in the same manner as their classmates.

Bracht (2011), who conducted studies of Study Island efficacy across several schools in a district, said, "Each school in the study had their own Study Island usage plan. The amount of usage and the manner in which students used the program varied" (p. 12).

No Matched Trials. To truly understand the efficacy (or lack thereof) of having Study Island access, matched comparison trials would need to be conducted, measuring test score growth on a treatment group with another non-Study Island group. However, most studies of Study Island did not have experimental design in that they did not have a randomly assigned treatment and control group. Instead, they examined growth in control groups in comparison to growth in similar groups of students in years prior, when Study Island was not available (Benthall, 2015; Bernard, 2013; Bracht, 2011; Gernert, 2014; Rich, 2016).

Bernard (2013) cites this limitation, saying, "This study is not a true experimental design with a control and experimental group, but the before and after effects could be examined because of natural breaks in treatments. A pseudo-before and after test was created because Study Island was used in 7th and 8th grades and not in 5th and 6th grades. The researcher used this break to try to examine a before and after treatment effect" (p. 15). Indeed, this was a common factor in most Study Island studies. Similarly, for Neshaminy School District, there is no control group, limiting the ability to assess Study Island's overall efficacy in this manner.

Limited Data. It should also be noted that for most previous studies of Study Island, data has been limited due to a small sample size, and/or an examination of only one Study Island subject. Dube (2011) concluded that "the main limitation [of his research on Study Island's efficacy] relates to the number of students included in the study," in the study (p. 61). The researcher felt that "the number of kids in the experimental group was small" and therefore "having more kids in the experimental group would lead to better and possibly different results" (Dube, 2011, p. 61). Though often not explicitly stated, this was true of many of our studies.

Furthermore, most of the studies that assessed Study Island focused on one subject. Of 14 studies examined, eight focused on math, three focused on reading/ communication arts, two focused on reading and math, and only one examined math, science, and reading. For this reason, the current paper also adds value to the literature on Study Island as (to the best of our knowledge) no other study has evaluated Study Island usage over more than one school year.

For these reasons, the data provided by Neshaminy, with a sample size of nearly 700, a timespan of three years, and information on three subjects, is quite robust, giving us a unique ability to evaluate Study Island. This team also believes our analysis can be used to inform many studies in the future.

Teacher Tech Issues & Hacks. In our review of existing literature, several studies included minor technological issues in their footnotes. Many found that teachers had varying levels of ability and familiarity with the Study Island website and interface, affecting how they utilized its features. Dube (2011) specifically mentions this in their "limitations" section, mentioning that (as a practitioner) he had forgotten his password at times and had limited access to the school's technology support personnel. Additionally, though not included in any formal studies, it is worth noting that for a significant fee (over $1,000 but varying based on location and time duration) Study Island offers in-person lessons for teachers, taught by Study Island specialists (Study Island, 2017a). Further, a Google search for "Study Island" brings up countless informal (usually teacher-produced) "how-to" videos, blogs, and discussion boards. These two factors demonstrate that Study Island's software

may be difficult to navigate for some teachers, increasing the possibility that it is used differently across schools, districts, and states.

Also, though not included in any studies, a YouTube search of "Study Island" produces hundreds of videos entitled "Study Island Hack," produced by and for students. While we have no information on the validity or effect of such hacks, it there are, indeed, ways to "outsmart" Study Island's usage tracking or assessment, this could affect findings (Youtube: Study Island, 2017).

These issues limit the reliability of Study Island data significantly. If usage is not uniform, or if results are not necessarily reflective of actual student growth, all data must be reviewed critically.

Data and Analysis Methods

The data set our team received from Neshaminy School District contains Keystone Assessments student data from three school years; 2013-2014 (SY14); 2014-2015 (SY15); and 2015-2016 (SY16). Our data set contains demographic and academic variables about each student in our sample population. Each student was identified using a "Master Student ID." The Master Student ID was a unique, anonymous individual identifier that allowed us to differentiate between students and keep track of each student when compiling multiple data sets.

Academic Data

There were several academic variables relevant to our analysis:
Student Grade Level: Students can complete the Algebra I Keystone exam between 7th and 10th grade. The Literature exam is taken by all students in 10th grade, and Biology is completed in either 9th or 10th grade.
Keystone Subject: Keystone Assessment data are from the Algebra I, Literature, and Biology exams. Students are administered the assessment upon completing their coursework in each of these subjects.
Number of Keystone Attempts: Students are able to retake each Keystone Assessment a maximum of three times before graduating high school.[2] The number of test attempts is dependent on whether students scored proficient--if students do not meet proficiency, they retake Keystone Exams the next year. All students in the data set took at least one of the Keystone exams a minimum of one time. For those who did not achieve proficiency on their first attempt, there was a second attempt, and, potentially a third, if necessary.
Keystone Exam Scaled Score: The data that we used for our analysis was the administration scaled score, the number used by the district to determine Keystone proficiency. Scores from previous Keystone Assessments were also included in the data set. Proficiency is defined by the state of Pennsylvania as a score of 1500 or higher on the exam.

Our study included students who did not achieve proficiency on their first attempt. These students were required to participate in a supplemental education program called Study Island (SI) and are referred to in this report as "Study Island Students" (SIS). Students who achieved proficiency on their first attempt were not required to participate in Study Island, and are referred to as Non-Study Island Student (NSIS); these students were not included in our analysis. Finally, there were some SIS who had no recorded re-takes. These students were also excluded, as it was impossible to measure any change in scores. Students who have an IEP (Individualized Education Plans) were not required to participate in Study Island, but could opt in.

[2] Students who do not pass the exam even after three attempts can complete a project in order to receive credit.

Table 1. Outline of Academic variables used.

Academic Variable	**Levels**
Year	2013-14 (SY14), 2014-15 (SY15), 2015-16 (SY-16).
Grade Level	7th, 8th, 9th, 10, 11th, or 12th.
Keystone Exam Subject	Algebra I, Literature, Biology subject tests
Scaled Score	1500 is considered proficient
Number of Test Attempts	First, Second, and/or Third

Demographic Data

The demographic variables included in our analysis were: gender, IEP (Individual Education Plan) (not gifted) status, ELL (English Language Learners) status, race/ethnicity, socioeconomic status (economically disadvantaged), and historically underperforming status. Economically disadvantaged students are students eligible for free or reduced price lunch, as per Neshaminy School district definitions. ELL students are those whose dominant language is not English who have not passed the exit criteria for English language proficiency and/or are within their first two years of monitoring. (Pennsylvania Department of Education, 2016a). Students with IEPs are defined by the Pennsylvania State Department of Education as students with provisional services for those who are disabled or but not gifted. Students also meet one of the 13 disability categories defined by IDEA and requires specifically designed instruction to receive educational benefits (Pennsylvania Department of Education, 2016b). Furthermore, Neshaminy School District defines historically underperforming students as a non-duplicated count of students with disabilities, economically disadvantaged students, and English Language Learners enrolled for a full academic year taking the PSSA/PASA/Keystone Exams (Pennsylvania Department of Education, 2015). If a student is in more than one of the individual groups of special education students, English Language Learners, and economically disadvantaged students, s/he is only included in the Historically Underperforming Student group and not counted as part of the other individual groups, which ensures non-duplicity (Pennsylvania Department of Education, 2015).

Table 2. Outline of Demographic variables used.

Demographic Variable	**Levels**
Gender	Male or Female
IEP (not gifted)	Students with provision of services for those who have disabilities and students with disabilities. Does not include gifted students. Students also meet one of the 13 disability categories defined by IDEA and requires specifically designed instruction to receive educational benefits
ELL	students whose dominant language is not English who have not passed the exit criteria for English language proficiency and/or are within their first two years of monitoring
Race/Ethnicity	American Indian, Asian, Black, Hispanic, Mixed Race, Multiracial, Pacific Islander, White
Economically Disadvantaged	Students are students on free or reduced price lunch, as per Neshaminy School district definitions.
Historically Underperforming	Non-duplicated count of students with disabilities, economically disadvantaged students, and English Language Learners enrolled for a full academic year taking the PSSA/PASA/Keystone Exams

Table 3 illustrates demographic characteristics of students who participated or did not participate in Study Island (determined by the number of test attempts and cut score) scored proficient. Around 25 percent of Neshaminy School District (NSD) students participate in Study Island across the three school years that we examined. In all three school years, there is a significant overrepresentation of IEP students and historically underperforming students participating in Study Island (at the 99% confidence level). In 2013-14 and 2014-15, there was a significant overrepresentation of Black/African American students (at the 99% confidence level for both years), Hispanic students (at the 99% confidence level for 2013-14 and 95% confidence level for 2014-15), and economically disadvantaged students (at the 95% confidence level for both years) in Study Island. These students were not overrepresented in 2015-16.

Eleventh grade students are overrepresented in the Study Island participant group across all three school years (at the 99% confidence level). This might be due to the fact that by 11th grade, students have taken all three Keystone Exam subjects and therefore, those who were not proficient in previous grade must participate in Study Island. Students can complete the Algebra I Keystone in 7-10th grade. The Literature Keystone is taken by all students in 10th grade, and Biology is completed in either 9th or 10th grade.

Table 3. All Student (NSD), Study Island (SIS) and Non-Study Island Student (NSIS) Demographics for all school years.

	School year 2013-14			School year 2014-15			School year 2015-16			Average for all years		
Student Demographics	**All NSD**	**SIS**	**NSIS**	**All NSD**	**SIS**	**NSIS**	**All NSD**	**SIS**	**NSIS**	**All NSD**	**SIS**	**NSIS**
Female	.51	.49	.51	.49	.47	.50	.52	.47	.52	.51	.48	.51
American Indian/ Alaskan Native	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Black/ African American	.05	.10***	.03**	.04	.09***	.02**	.04	.05	.03	.04	.05	.03
Hispanic (any race)	.03	.06***	.02*	.03	.05**	.02*	.03	.05	.02*	.03	.05	.02
White	.86	.80***	.88	.86	.82**	.89***	.87	.85	.87	.86	.85	.88
Multi-racial	.01	.02*	.01	.02	.02	.01	.02	.02	.02	.02	.02	.01
Asian	.04	.03	.05	.05	.03	.06	.05	.03	.05	.05	.03	.05
Native Hawaiian or Pacific Islander	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
7th Grade	.00	.00	.00	.06	.00***	.08**	.08	.00**	.09	.05	.00	.05
8th Grade	.22	.05***	.30***	.18	.02***	.25***	.13	.00**	.15	.18	.02	.15
9th Grade	.30	.29	.30	.30	.27	.31	.30	.06***	.35**	.30	.28	.31
10th Grade	.37	.34	.39	.34	.32	.34	.38	.29**	.40	.36	.33	.38
11th Grade	.10	.32***	.01***	.13	.39***	.02***	.11	.65***	.01***	.11	.45	.01
12th Grade	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Economically Disadvantaged	.18	.24**	.15**	.20	.26**	.17**	.19	.21	.18	.19	.21	.18
English Language Learners (Current and former)	.02	.03	.02	.03	.04	.02	.03	.05	.03	.03	.04	.02
IEP (Not Gifted)	.14	.31***	.07***	.14	.33***	.07***	.10	.29***	.07**	.13	.31	.07
Historically Underperforming	.28	.46***	.21***	.31	.50***	.2***	.26	.43***	.23	.28	.46	.23
Total Count of Students	1590	455	1135	1699	513	1186	1405	215	1190	1565	394	1170
Percentage of Students	-	.29	.71	-	.30	.70	-	.30	.70	-	.30	.70

Note: *denotes significance in a t-test, at a 90% confidence level; **denotes significance in a t-test, at a 95% confidence level; ***denotes significance in a t-test, at a 99% confidence level. We compared the NSD counts to SIS and NSIS proportions.

"Historically Underperforming Students are defined as a non-duplicated count of students with disabilities, economically disadvantaged students, and English Language Learners enrolled for a full academic year taking the PSSA/PASA/Keystone Exams. If a student is in more than one of the individual groups (e.g., special education and English Language Learner), s/he is only included in the Historically Underperforming Student group one time – a non- duplicated count. This group is not a cohort but rather students currently in the building meeting the definition during the reported year" as defined by the Neshaminy School District (http://paschoolperformance.org/FAQ)

A "0" does not necessarily mean that there were no students of that racial/ ethnic groups, just that there were so few that the proportion rounded to 0.

There is 1 student who did not report their gender and ethnicity for school year 2013-14, 1590

There is 1 student who did not report their ethnicity, but the total number of students in 2014-15 is 1699

Table 4 illustrates what proportion of SIS students took which Keystone Exam based on demographics across the three school years. The subject combination categories (Algebra I; Biology; Literature; Algebra & Biology; Algebra & Literature; Biology & Literature Algebra, Biology, & Literature) are mutually exclusive, meaning that students who took more than one subject were categorized in the respective subject combination.

Table 4: All Study Island (SIS) by Subject Test and demographics for all school years.

	School Year 2013-14							School Year 2014-15							School year 2015-16						
Student Demographics	Algebra One	Biology	Literature	Algebra & Biology	Algebra & Literature	Biology & Literature	Algebra, Biology, & Literature	Algebra One	Biology	Literature	Algebra & Biology	Algebra & Literature	Biology & Literature	Algebra, Biology, & Literature	Algebra One	Biology	Literature	Algebra & Biology	Algebra & Literature	Biology & Literature	Algebra, Biology, & Literature
Female	0.48	0.5	0.43	0.77	0.3	0.46	0.63	0.55	0.5	0.41	0.37	0.4	0.52	0.46	0.64	0.62	0.35	0.56	0.5	0.48	0.56
American Indian/Alaskan Native	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Black/African American	0.05	0.17	0.04	-	0.2	0.15	0.22	0.07	0.1	0.05	-	0.13	0.04	0.09	0.07	0.1	0.03	0.06	0.21	-	0.07
Hispanic (any race)	0.05	-	0.09	0.05	0.05	0.09	0.05	0.14	-	0.08	-	0.03	0.07	0.06	0.14	0.05	0.03	0.06	0.17	0.06	0.02
White	0.87	0.67	0.79	0.82	0.75	0.72	0.73	0.79	0.85	0.82	0.89	0.78	0.84	0.81	0.79	0.82	0.87	0.81	0.83	0.88	0.85
Multi-racial	0	-	0.09	-	-	0.2	-	-	-	0.01	0.11	0.01	-	0.02	-	-	-	0.06	-	-	0.05
Asian	0.03	0.17	0	0.17	-	0.3	-	0	0.05	0.04	-	0.04	0.04	0.02	-	0.03	0.06	-	-	0.06	0.02
Native Hawaiian or Pacific Islander	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Economically Disadvantaged	0.2	0.5	0.26	0.14	0.35	0.27	0.34	0.31	0.2	0.2	0.37	0.33	0.2	0.25	0.21	0.15	0.13	0.25	0.17	0.19	0.6
ELL	0.02	0.17	-	0.09	-	0.05	0.02	0.04	-	0.05	0.05	0.04	0.09	0.03	-	0.03	0.03	-	-	0.08	0.08
7th Grade	-	-	-	-	-	-	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-
8th Grade	0.11	-	-	-	-	-	-	0.08	-	-	-	-	-	-	-	-	-	-	-	-	-
9th Grade	0.46	0.67	-	0.95	0.15	0.03	0.07	0.65	0.65	-	0.79	0.04	0.07	0.03	0.36	-	-	0.38	0.33	-	0.02
10th Grade	-	0.17	0.34	0.05	-	0.96	0.85	0.03	-	0.08	0.16	0.33	0.7	0.9	0.07	0.05	0	-	0.67	0.19	0.79
11th Grade	0.43	0.17	0.66	-	0.85	0.01	0.07	0.24	0.35	0.92	0.05	0.63	0.24	0.08	0.57	0.95	0.97	0.63	-	0.81	0.2
12th Grade	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IEP	0.25	-	0.3	0.09	0.6	0.32	0.66	0.24	0.2	0.26	0.05	0.58	0.33	0.49	0.07	0.23	0.23	0.31	0.17	0.4	0.34
Historically underperforming	0.39	0.5	0.45	0.23	0.7	0.48	0.8	0.47	0.3	0.41	0.37	0.7	0.5	0.59	0.29	0.33	0.35	0.38	0.33	0.52	0.52
Total Count of Students	215	6	47	22	20	28	41	157	20	98	19	67	46	106	14	39	31	16	6	48	61
Percentage of Students	0.57	0.02	0.12	0.06	0.05	0.07	0.11	0.31	0.04	0.19	0.04	0.13	0.09	0.21	0.07	0.18	0.14	0.07	0.03	0.22	0.28

Note. A "0" does not necessarily mean that there were no students of that racial/ ethnic groups, just that there were so few that the proportion rounded to 0.

The percentage of students who were not proficient and therefore participated in Study Island was largest for Algebra in SY14 and SY15, with 57% of SIS in SY14 and 31% of SIS in SY15. SIS who participated in only Algebra was 7% in SY16. In SY16, the proportion of SIS was largest for those who were not proficient in all three subjects at 28%. SIS proportions were lowest for Biology in SY14 and 15, with 2% of SIS in SY14 and 4% in SY15. 18% of SIS were not proficient in Biology in SY16. The smallest proportion of SIS in SY16 was for Algebra and Literature, at just 3%. Across the three years, the majority of SIS not proficient in Literature were male with 57% in SY14, 59% in SY15, and 75% in SY16.

A large proportion of IEP and historically underperforming students did not meet proficiency in Algebra and Literature. The proportion of IEP students who did not meet proficiency in Algebra and Literature was 60% in SY14, 58% in SY15, and 17% in SY16, showing a sharp decrease. The proportion of historically underperforming students who did not meet proficiency in Algebra and Literature was 70% in SY14, 70% in SY15, and 33% in SY16, showing a similar decline to their IEP peers. Additionally, a substantial proportion of IEP students and historically underperforming students did not meet proficiency in all three subject areas of Algebra I, Biology, and Literature. For IEP students, the proportion of students who did not meet proficiency in all three subjects was 60% in SY14, 58% in SY15, and 17% in SY16, again showing a large decrease. For historically underperforming students, the proportion of students who did not meet proficiency in all three subjects was 70% in SY14, 70% again in SY15, and 30% in SY16, showing a decrease as with IEP students.

Analysis

The research questions presented by the Neshaminy School District were twofold, and our analysis therefore addressed them separately. Our analysis for both questions focused solely on students who did not reach proficiency (defined as a score of 1500 or higher) on their first attempt on the Keystone Exam, as both research questions (stated below) are focused specifically on students who participated in Study Island remediation.

Underperforming students (defined here as students who scored below 1500 on one or more of their subject tests at the Keystone Exam on their first test attempt) were required to participate in Study Island remediation prior to attempting the test a second time (and sometimes a third time). Although the research question refers to Math, Literature, and Science test scores, these tests are recorded as Algebra I, Literature, and Biology in the dataset, and therefore we will refer to them as such henceforth.

The researchers used a quasi-experimental design because the student participants were not randomly assigned, also called ex post facto design. Students scoring below proficient on their first attempt at the Keystone Exam are allowed to retake the exam until they gain proficiency, or take the exams three times, as per Neshaminy School District requirements. Therefore, for both research questions, researchers used students first and last exam results to compare the students who received Study Island remediation. Students who did not receive Study Island remediation (i.e. students who reached proficiency on their first attempt) were used as a comparison group in both research questions.

Research Question 1: Do academic outcomes (PA Keystone Assessments in Math/Literature/Science) improve more in one content area vs. another when using Study Island as a remediation tool in secondary education?
In order to answer this question, we analyzed trends in growth across test attempts for all students in Study Island differentiated by subject name. We first categorized students by number of test attempts so that there were two groups of students - one group that attempted the test twice and a second group that attempted the test three times. We conducted t-tests to test significance of change in average test scores across test attempts by comparing mean scores across attempts for each subject. We tested the difference in mean scores at the 90 percent, 95 percent, and 99 percent confidence level.

Research Question 2: Do academic outcomes (PA Keystone Assessments in Math/Literature/Science) improve more for some underperforming students over others when Study Island is used as a remediation tool in secondary education?
We answered this question by analyzing the differences in test scores across test attempts for each demographic group. We first categorized students by number of test attempts so that there were two groups of students - one group that attempted the test twice and a second group that attempted the test three times. We conducted t-tests to test significance of change in average test scores across test attempts by comparing mean scores across attempts for each subject and demographic group. For example, we analyzed the differences in average test scores across attempts for males versus females, to test significance in differences of mean test scores. We tested the difference in mean scores at the 90 percent, 95 percent, and 99 percent confidence level.

Additionally, we conducted descriptive analysis on the proportion of students from various demographic groups in the group of Study Island participants and compared these demographic statistics to the group of students who did not participate in Study Island.

Empirical Results and Findings

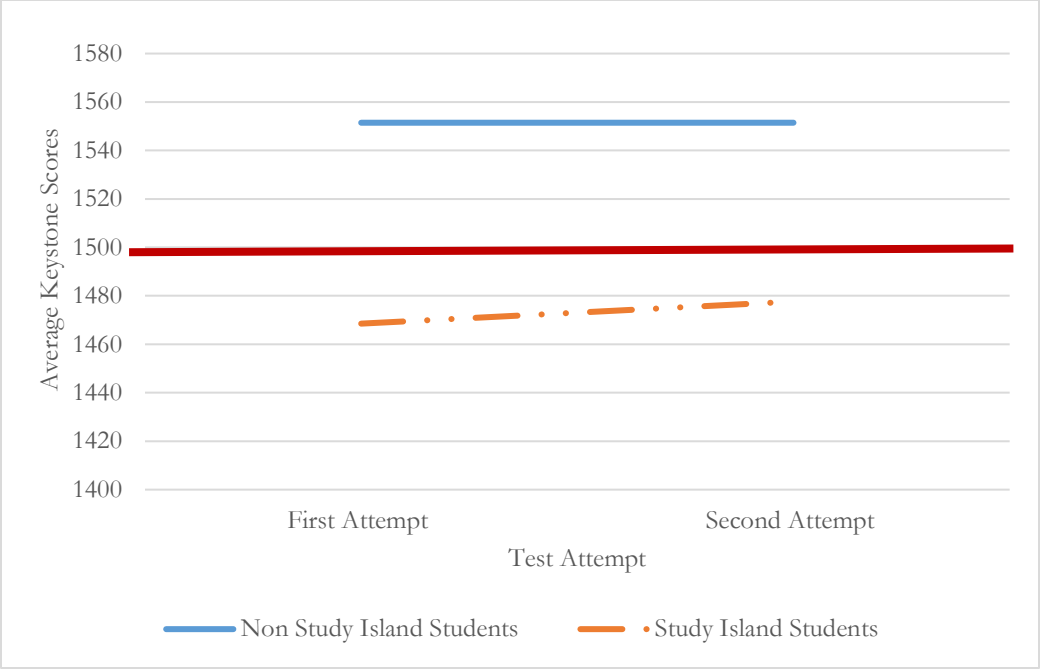
Research Question 1: Is Study Island more effective in one content area vs. another (PA Keystone Assessments in Math/Literature/Science) for underperforming students when used as a remediation tool in secondary education?

Algebra

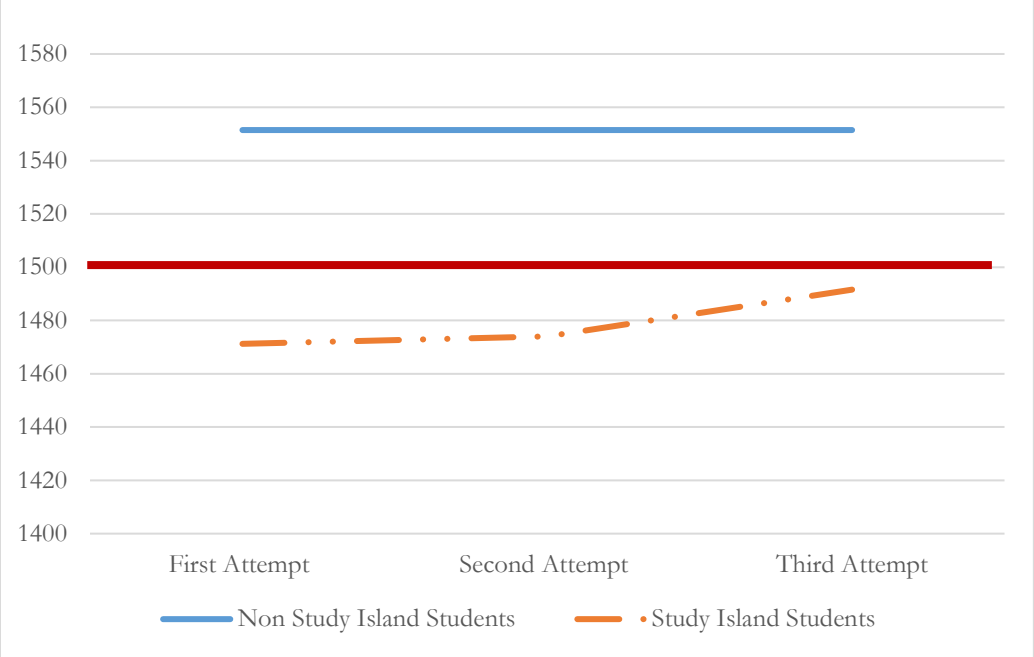
There was incremental change in Algebra exam scores across attempts among students who did not reach proficiency on their first attempt. All changes in test scores are standardized and reported as t-values in Appendix A.

Among all students who scored below 1500 in Algebra on their first test attempt (1500 is the proficiency cut score), there was a statistically significant difference of 9.099 between the initial Algebra I Keystone mean exam scores and the second attempt mean exam scores among students who attempted the exam twice. Among students who attempted the test three times there was an increase of test scores of 2.813 between the first and second test attempts that was not statistically significant. There was a statistically significant increase of 17.521 between mean scores of test attempts two and three. Out of 319 students who attempted Algebra twice, 50 students reached proficiency after two attempts. Out of 24 students who had three test attempts in Algebra, 8 students reached proficiency on their third attempt.

Figure 1. All students with Algebra I test attempts.
Panel A: Students who took the Algebra I Keystone twice.



Panel B: Students who took the Algebra I Keystone three times.



Biology

However, for Biology, there was no statistically significant difference Out of mean scores for each attempt. Among students who scored below proficient in the first test attempt for Biology, there was no significant difference in the Biology Keystone exam scores between the first test attempt and the second test attempt (after Study Island participation). The difference in mean scores across the two attempts was -3.302, which means that the mean score showed a decline from test attempt one to two, although this decline was not statistically significant. There were no students who took the Biology test a third time (because the first attempt for Biology happened comparatively later than other subjects). Out of 83 students who took the Biology test twice, 12 students met proficiency on their second attempt. All changes in mean test scores are standardized and displayed as t-values in Table 1 in Appendix A.

Figure 2. All students with Biology test attempts.

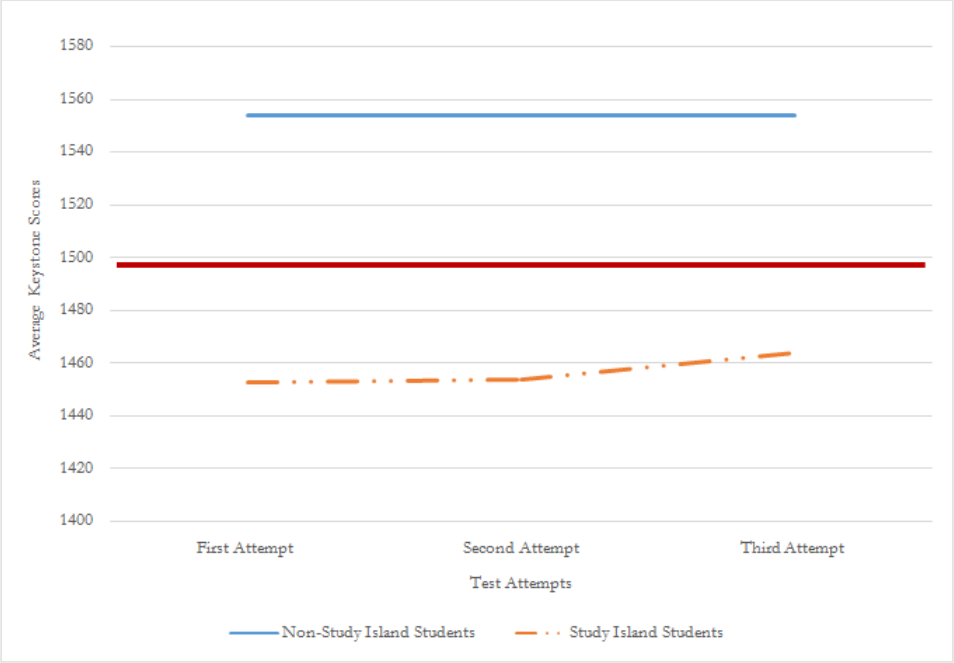
Literature

Overall, there was also no statistically significant difference in growth between attempts. All changes in mean test scores are standardized and displayed as t-values in Table 1 in Appendix A. Out of mean scores for test attempts in Literature. There was an increase of 4.186 in Literature exam scores between the first attempt and second attempt among two-time test takers, although this was not statistically significant. There was an increase of 1.5 between Literature mean exam scores on the first attempt and second attempt among three-time test takers and an increase of 9.625 between attempts two and three although neither change was statistically significant. Out of 205 students who attempted the test twice in Literature, 68 students reached proficiency on their second attempt. Out of 4 students who attempted the test three times, one student met proficiency on their third attempt.

Figure 3. All students with Literature test attempts.

Panel A. Students who took the Literature Keystone twice.

Panel B. Students who took the Literature Keystone three times.



Therefore, analysis of mean exam scores indicates that there is a significant difference in mean exam scores between the re-takes and initial test attempt for Algebra I, but not for Literature and Biology. However, we cannot make a definitive statement about the impact of Study Island specifically on these improvements (or lack thereof), due to the absence of a randomized control trial design experiment. Additionally, Study Island dosage was variable and there was no mandatory dosage that was administered to students. Therefore, the researchers conclude that test scores across attempts improved in Algebra, but not in other subjects.

Research Question 2: Does the Study Island program contribute to improved academic outcomes (PA Keystone Assessments in Math/Literature/Science) for underperforming students when used as a remediation tool in secondary education?

Algebra

Among students who participated in Study Island for Algebra, those who were in the following demographic groups had statistically significant differences in mean scores between test attempts (e.g., between initial and second test attempt, between second and third test attempt): White, female, male, historically underperforming, non-historically underperforming, non-economically disadvantaged, IEP, non-IEP, and non-ELL students. All changes in mean test scores are standardized and displayed as t-values in Tables 1-7 in Appendix A.

Demographic: Race- White, Black, and Hispanic students[3]

White students in Study Island made significant growth across Algebra test attempts. There was a statistically significant growth of 8.57 points between the initial Algebra I Keystone exam scores and the second attempt exam scores among students who attempted the test twice. Although the mean score between first and second test attempts increased by 2.81 points among two time test takers, it

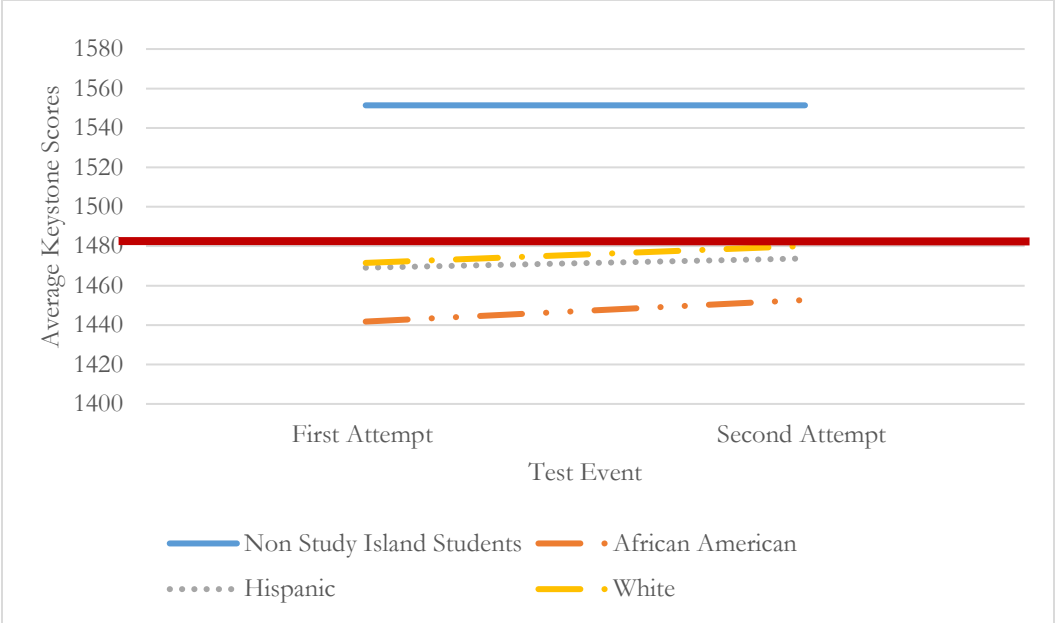
3 For this analysis, researchers only included White, Black, and Hispanic students as other racial group sample sizes were quite small. The researchers included the other racial groups' analysis in the appendix.

was not a statistically significant increase. The test score increase of 19.18 between second attempt and third attempt for three-time test takers was statistically significant. Out of 259 White students who attempted the Algebra test twice, 82 White students met proficiency on their second test attempt. Out of 16 White students who attempted the Algebra test three times, 6 White students achieved proficiency on their third test attempt.

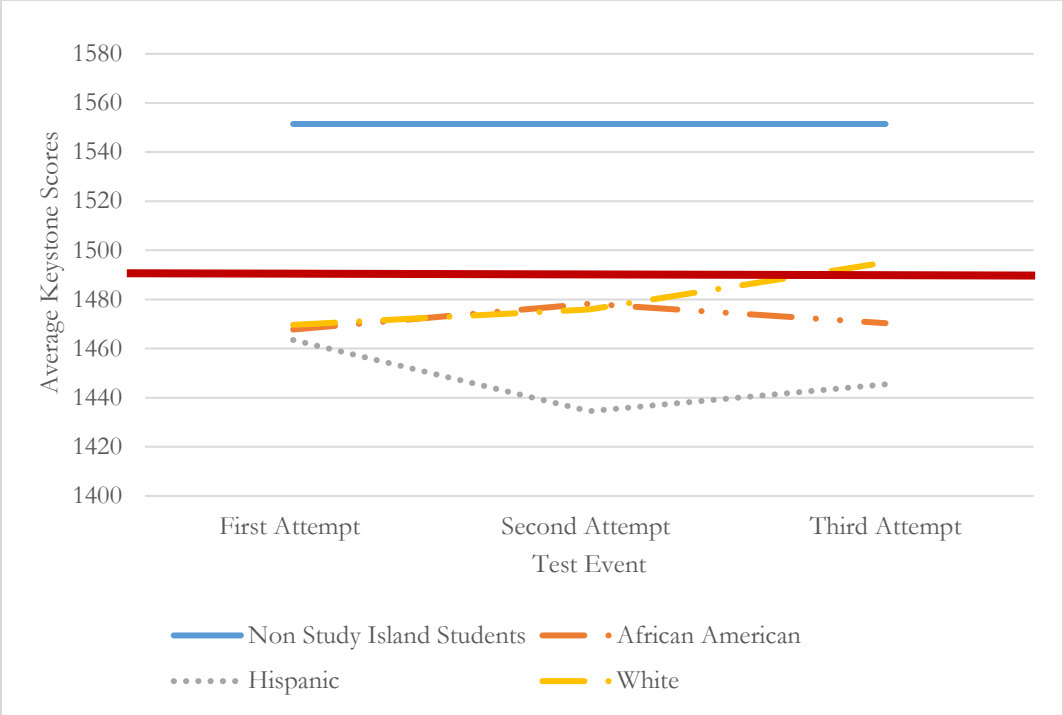
Black students in Study Island did not make significant growth across Algebra test attempts. The mean test score between attempts one and two increased by 10.92 among two time test takers. However, the increase was not statistically significant. The mean test score between attempts one and two increased by 10.5 among Black students who took the test three times and decreased by 7.86 between attempts two and three. However, neither of these changes for three time Black test takers was statistically significant. It should be noted that the sample size of the number of Black students who had a third test attempt for Algebra was very small (N=4). Out of 25 Black students who attempted the Algebra test twice, 4 Black students reached proficiency on their second test attempt. Out of 4 students who attempted the Algebra test three times, no students reached proficiency.

Hispanic students in Study Island did not make significant growth in Algebra exam scores across the test attempts. The mean score across the first and second test attempts increased by 4.67 points, but this was not a statistically significant among two-time test takers. The mean test score decreased between attempt one and two by 29 points among three-time test takers and increased by 11 points between second and third attempts. However, these differences could not be tested for statistical significance as only one Hispanic student took the Algebra test a third time. Out of 18 Hispanic students who attempted the Algebra test twice, 2 students reached proficiency on their second test attempt. The one student who took Algebra the third time did not meet proficiency.

Figure 4. All students with Algebra I test attempts, by race/ethnicity
Panel A. Students who took the Algebra I Keystone twice.



Panel B. Students who took the Algebra I Keystone three times.



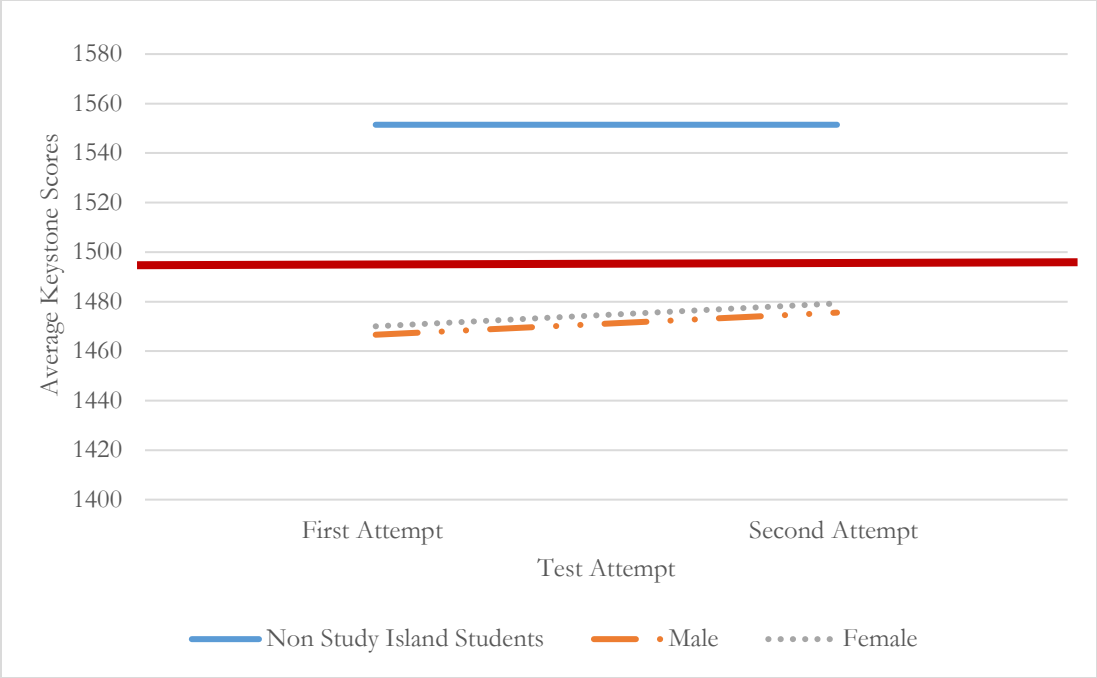
Demographic: Gender- Male students and Female students

Female students who participated in Study Island made statistically significant growth across test attempts. There was a statistically significant increase of 9.23 between the initial Algebra I Keystone exam scores and the second attempt exam scores among two-time test takers. There was an increase of 4.96 between the first Algebra I Keystone exam scores and the second attempt exam scores among three-time test takers, although this was not statistically significant. There was also an increase of 22.63 between the second and third attempt mean Algebra scores among female two-time test takers. Out of 178 female two-time Algebra test takers 52 students reached proficiency on their second test attempt. Out of 12 female three-time Algebra test takers, 6 reached proficiency on their third test attempt.

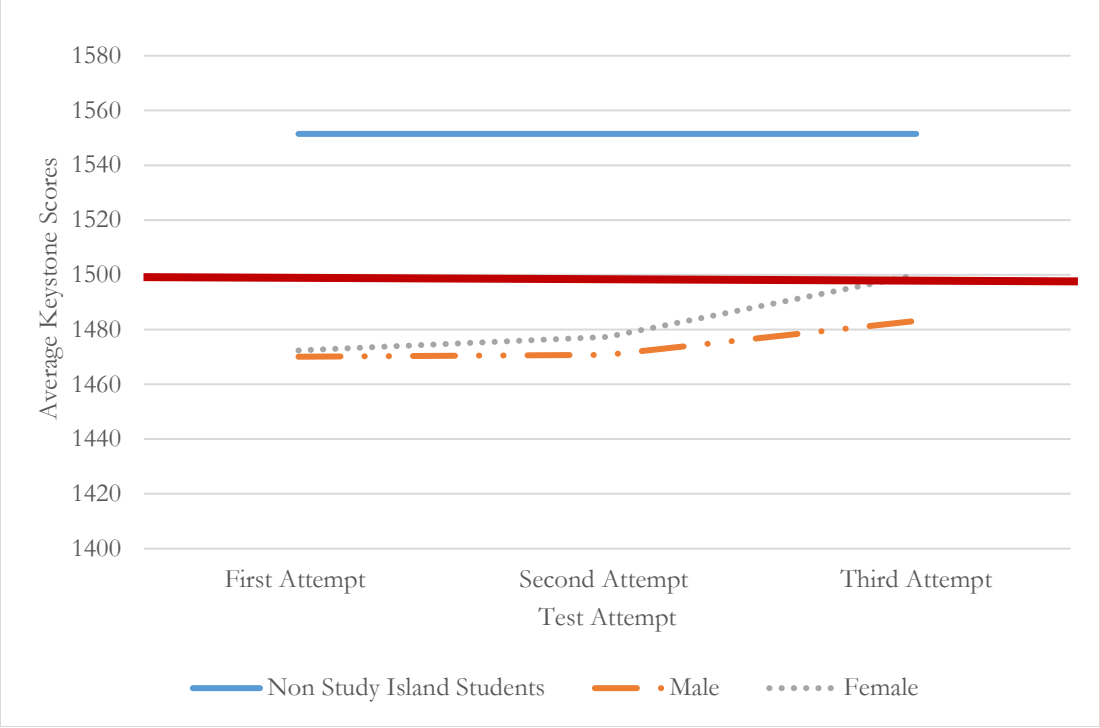
Male students who participated in Study Island made significant growth across test attempts. There was a statistically significant increase of 8.57 between the initial Algebra I Keystone exam scores and the second attempt exam scores among two-time test takers. There was an increase of 2.81 between the first Algebra I Keystone exam scores and the second attempt exam scores among three-time test takers, although this increase was not statistically significant. There was also an increase of 19.18 between the second and third attempt mean Algebra scores among male two-time test takers. Out of 141 male two-time Algebra test takers, 46 students reached proficiency on their second test attempt. Out of 12 male three-time test takers, 2 reached proficiency on their third test attempt.

Across all Study Island participants, female and male students who participated in Study Island made similar growth from the first test attempt to the second test attempt, as growth in exam scores for both groups was approximately 9 points. However, females made more growth than males from the second test attempt to the third test attempt with females increasing by an average of approximately 20 points and males by an average of 8 points.

Figure 5. All Algebra I test-takers, by gender.
Panel A. Students who took Algebra I Keystone twice, by gender.



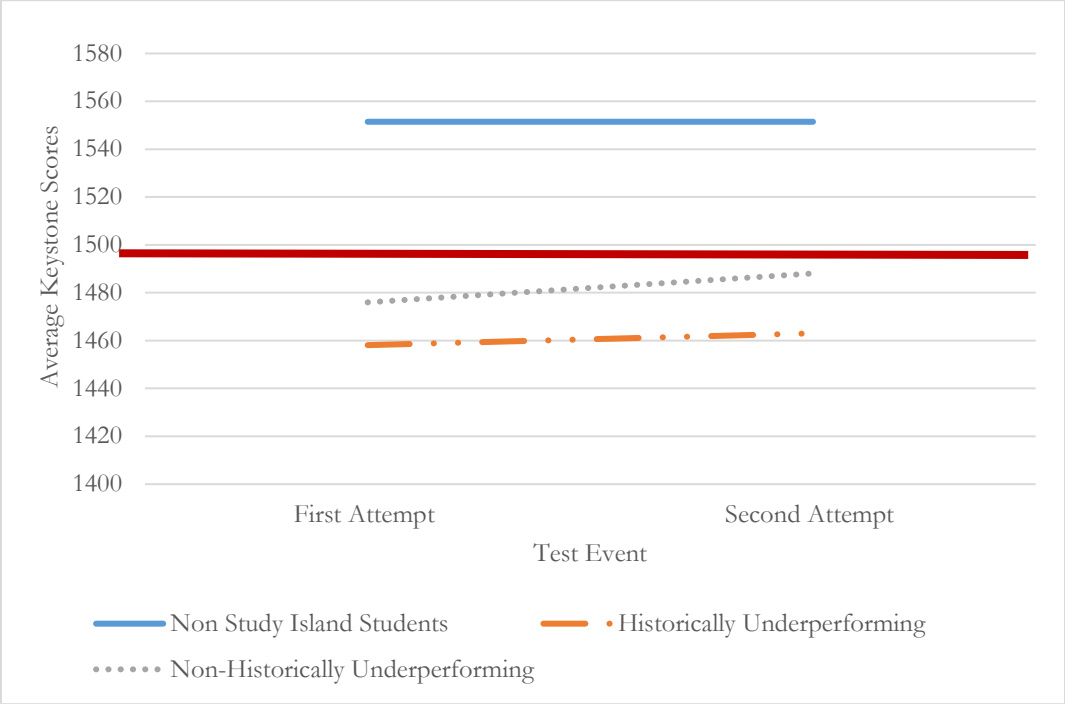
Panel B. Students who took Algebra I Keystone three times, by gender.



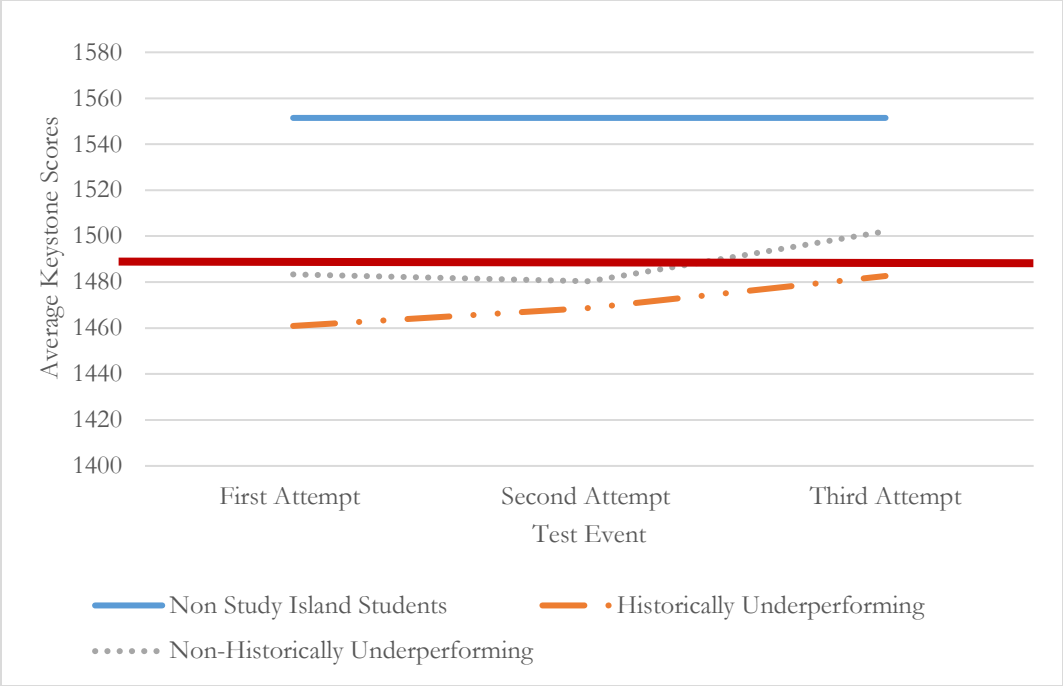
Demographic: Historically underperforming students in comparison to others
Historically underperforming students in Study Island showed partial growth. There was no statistically significant difference (a mean score increase of 4.93) in the initial Algebra I Keystone exam scores and the second attempt exam scores among two time test takers. Students who attempted the test three times also showed growth across attempts that was not statistically significant. There was an average increase of 7.77 between first and second test attempts among three-time test takers and 13.96 among second and third attempt. Out of 133 historically underperforming students who had two test attempts in Algebra I, 27 reached proficiency on their second test attempt. Out of 13 historically underperforming students who attempted the exam three times, 2 students reached proficiency on their third attempt.

For non- historically underperforming students in Study Island, there was a statistically significant increase of 12.08 between the initial Algebra I Keystone exam scores and the second attempt exam scores. There was no statistically significant difference in the Algebra I Keystone exam scores between the first and second attempts among three-time test takers, and the mean score showed a decline of 3.05. Test scores increased by 21.73 among second and third attempts for three-time test takers. Out of 186 non-historically underperforming students with two test attempts in Algebra 1, 70 reached proficiency on their second test attempt. Out of 11 non-historically underperforming students with three test attempts in Algebra, 6 reached proficiency on their third test attempt.

Figure 6. All students who took the Algebra I Keystone exam, by historically underperforming status.
Panel A. Students who took the Algebra I test twice, by historically underperforming status.



Panel B. Students who took the Algebra I test three times, by historically underperforming status.



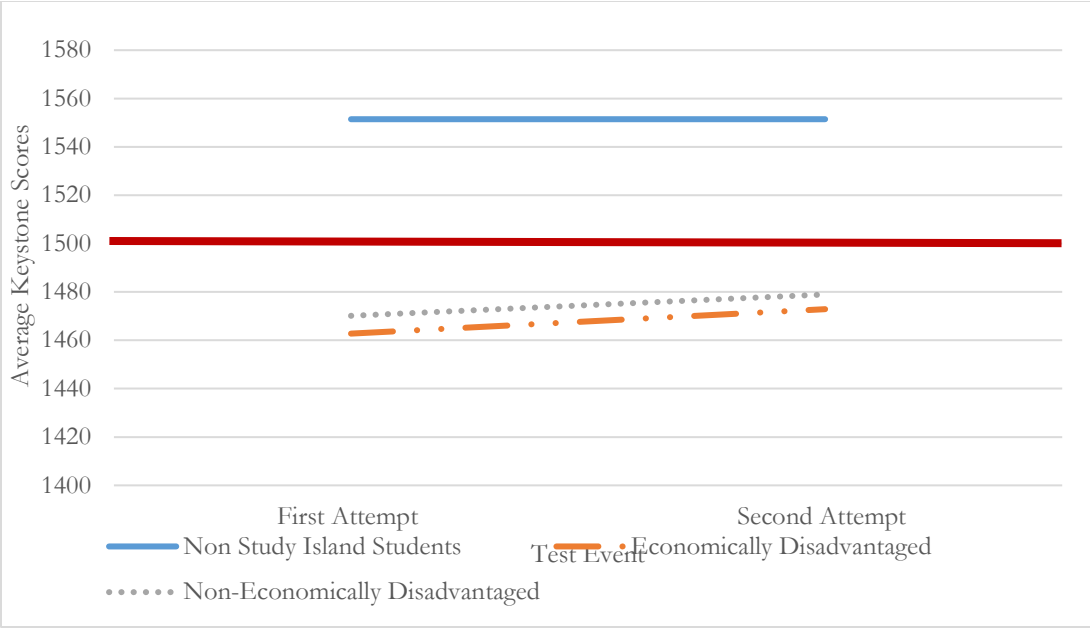
Demographic: Economically disadvantaged students in comparison to others
Students in Study Island that were not economically disadvantaged showed growth with Study Island participation while students in Study Island that were economically disadvantaged did not show significant growth with Study Island participation.

Students who were economically disadvantaged and took the Keystone exam twice did not show significant growth with Study Island participation across first and second test attempts. The mean scores increased by 10.108, but this was not a statistically significant increase. The mean scores between attempts one and two for three-time test takers increased by 5.19, which was not a statistically significant increase. However, the mean score increase of 19.18 between attempts two and three for three-time test takers was statistically significant. However, the sample size of the number of students who took the test a third time was very small (n=8). Out of 70 economically disadvantaged students with two test attempts in Algebra, 20 students reached proficiency on their second test attempts. Out of 8 economically disadvantaged students with three test attempts in Algebra, 1 reached proficiency on their third test attempt.

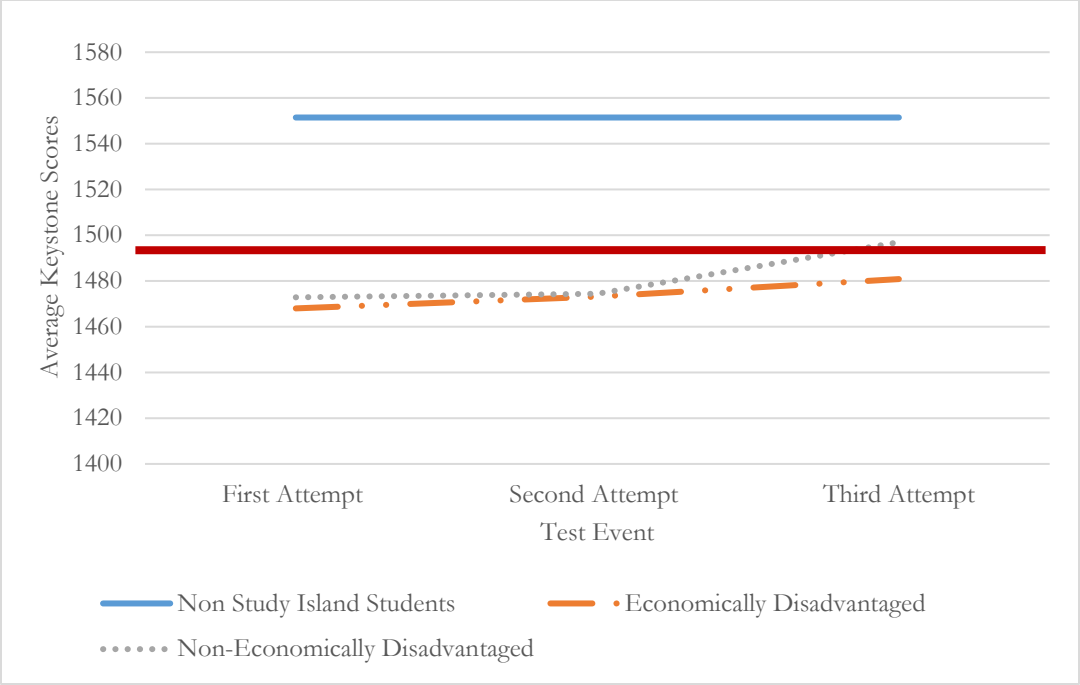
Students who were not economically disadvantaged and took the Keystone exam twice showed significant growth of 8.82 with Study Island participation across first and second test attempts. The mean scores between attempts one and two for three-time test takers increased by 1.64, which was not a statistically significant increase. However, the mean score increase of 22.47 between attempts two and three for three-time test takers was statistically significant. Out of 249 non-economically disadvantaged students who attempted the Algebra 1 test twice, 28 reached proficiency on their second test attempt. Out of 16 non-economically disadvantaged students who attempted the Algebra 1 test three times, 7 reached proficiency on their third test attempt.

Figure 7. All students who took the Algebra I Keystone exam, by economically disadvantaged status.

Panel A. Students who took the Algebra I Keystone exam twice, by economically disadvantaged status.



Panel B. Students who took the Algebra I Keystone exam three times, by economically disadvantaged status.



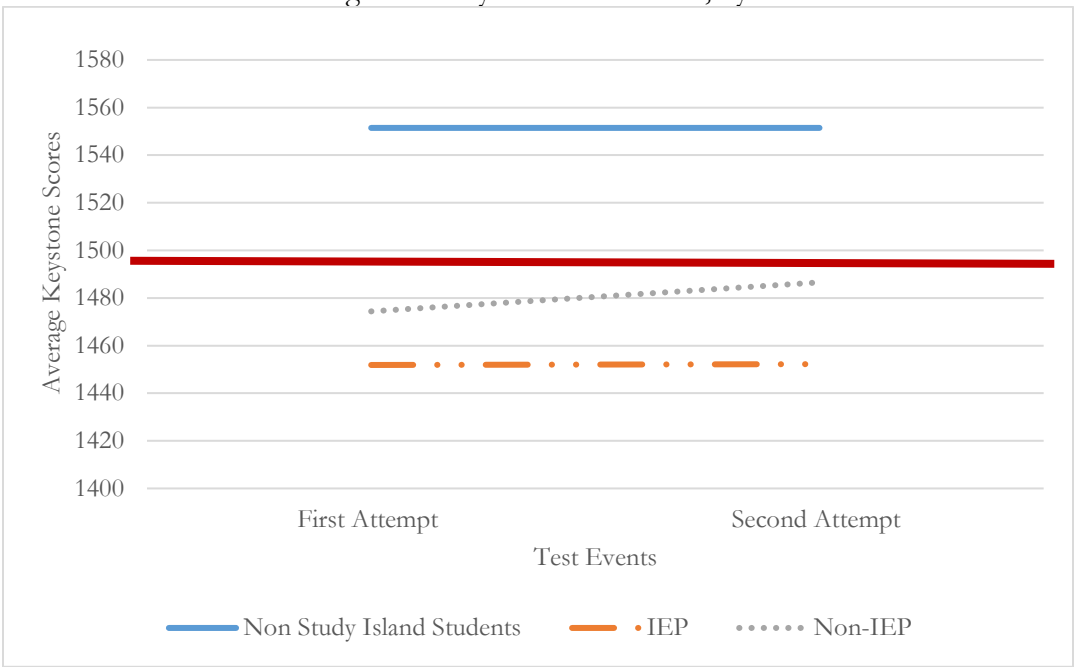
Demographic: Students on an Individualized Education Plan in comparison to those who are not

Students who were on an IEP and took the Keystone exam twice showed an increase in mean scores by 0.35 points across first and second test attempts, which was not statistically significant. The mean scores between attempts one and two for three-time test takers increased by 11.85, which was not a statistically significant increase. The mean score increase of 20.38 between attempts two and three for three-time test takers was also not statistically significant. However, the sample size of the number of students who took the test a third time was very small (n=8). Out of 83 IEP students who had two test attempts in Algebra I, 10 reached proficiency on their second test attempt. Out of 8 IEP students with three test attempts, 2 reached proficiency on their third test attempt.

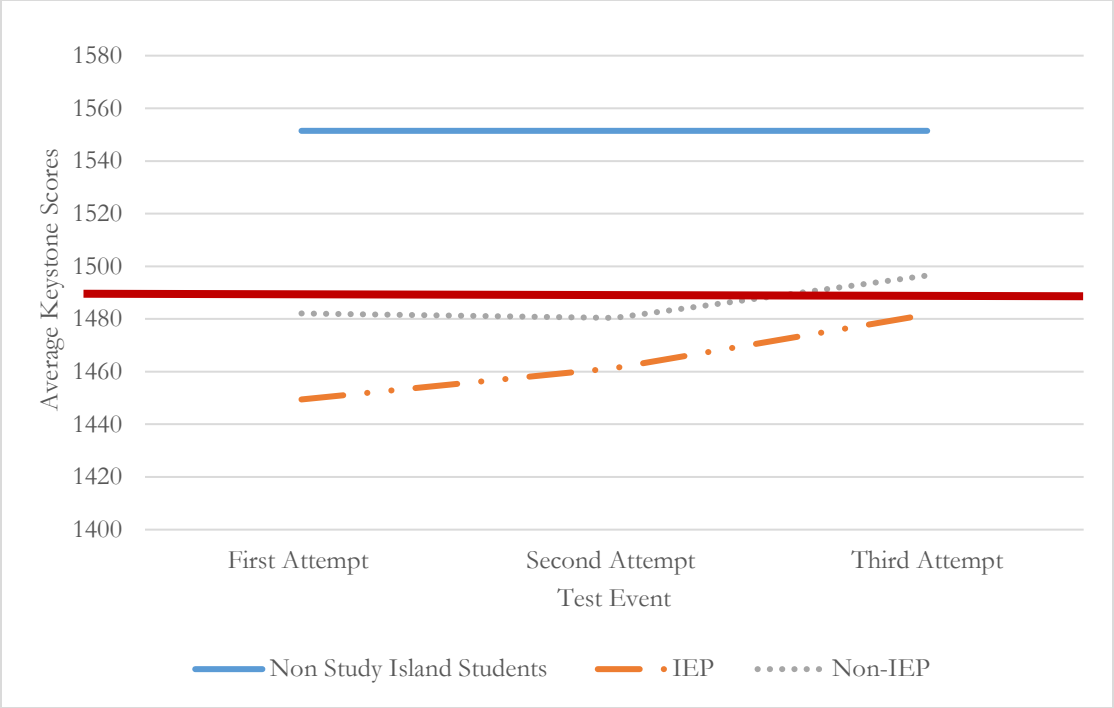
Students who were not on an IEP and took the Keystone exam twice showed an increase in mean scores by 12.18 points across first and second test attempts, which was statistically significant. The mean scores between attempts one and two for three-time test takers decreased by 1.72, which was not a statistically significant increase. The mean score increase of 16.09 between attempts two and three for three-time test takers was statistically significant. However, the sample size of the number of students who took the test a third time was very small (n=16). Out of 236 non-IEP students with two test attempts for Algebra I, 88 reached proficiency. Out of 16 non-IEP students with three test attempts for Algebra I, 6 reached proficiency.

Figure 8. Students who took the Algebra I Keystone exam, by IEP status.

Panel A. Students who took the Algebra I Keystone exam twice, by IEP status.



Panel B. Students who took the Algebra I Keystone exam three times, by IEP status.

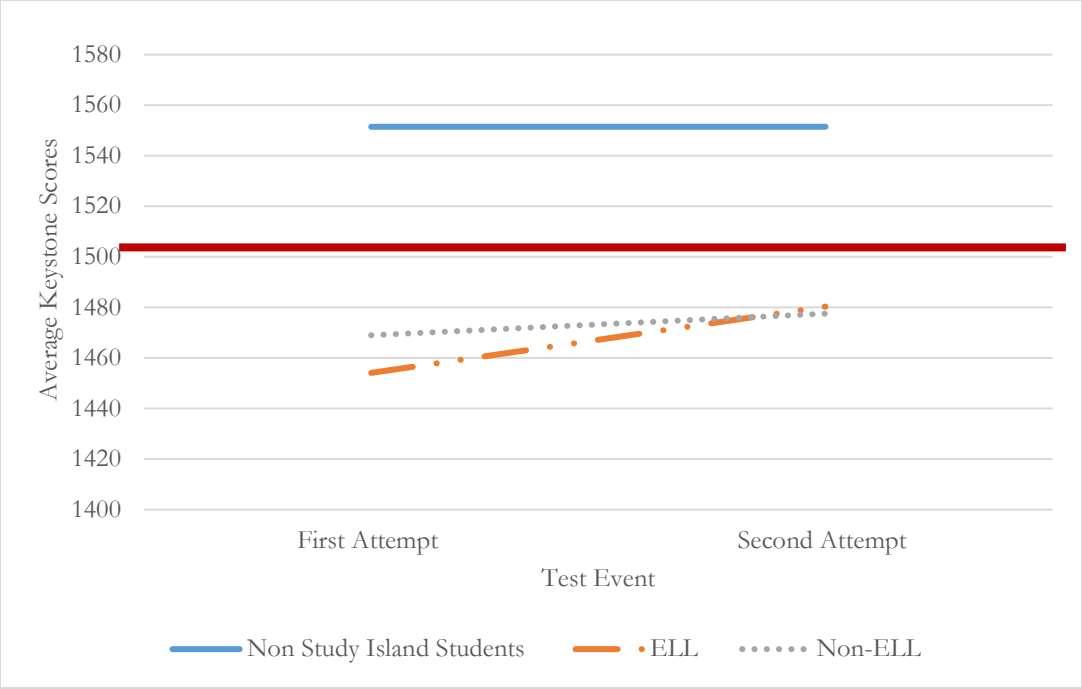


Demographic: English Language Learners and those who are not

ELL students who participated in Study Island did not make significant growth across test attempts. Although the mean score between the first and second test attempts increased by 4.93 points for two-time test takers, it was not a statistically significant increase between the first attempt and the second attempt. It should be noted that there were only ten ELL students who participated in Study Island for Algebra I. There were no ELL students that had a third test attempt for Algebra I. Out of 10 ELL students who attempted the Algebra I test twice, three reached proficiency on their second attempt. No students who were ELL attempted the test a third time.

For students in Study Island who were not ELL, there was a statistically significant increase of 8.55 between initial Algebra I Keystone exam scores and the second attempt exam scores for two-time test takers. There was also a statistically significant difference of 17.52 in the Algebra I Keystone exam scores between the second test attempt scores and the third test attempt scores for three-time test takers. The increase in mean scores of 2.81 between first and second attempts for three-time test takers was not statistically significant. Out of 309 non-ELL students who had two test attempts in Algebra I, 95 reached proficiency on their second test attempt. Out of 24 non-ELL students with three test attempts, 8 reached proficiency on their third test attempt.

Figure 9. Students who took the Algebra I Keystone exam, by ELL status.



Biology
No demographic groups demonstrated significant growth in Biology. As previously stated, there were no students who took the Biology test a third time (because the first attempt for Biology happened comparatively later than other subjects). All changes in mean test scores are standardized and displayed as t-values in Table 1 in Appendix A.

Demographic: Race- White, Black, and Hispanic students[4]

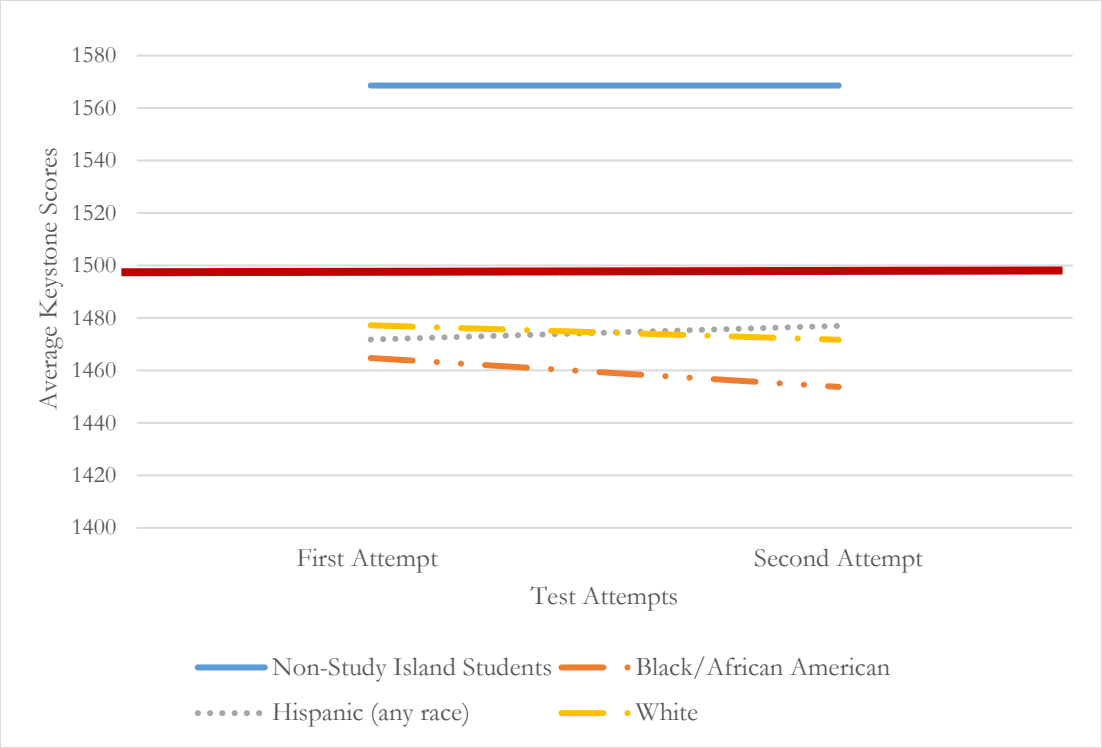
White students in Study Island did not make significant growth across Biology test attempts. Although the mean score between the first and second test attempts decreased by 5.5 points, it was not a statistically significant decrease between the first attempt and the second attempt. Out of 71 White students who attempted the test twice in Biology, 10 reached proficiency on their second attempt.

Black students in Study Island did not make significant growth across Biology test attempts. Although the mean score between the first and second test attempts decreased by 11 points, it was not a statistically significant decrease The sample size of Black students who participated in Study Island for Biology was very small (n=4). No Black students reached proficiency in Biology on their second attempt.

Hispanic students in Study Island did not make significant growth across Biology test attempts. The mean score across the first and second test attempts increased by 5.2 points, but this was not a statistically significant increase between the first attempt and the second attempt. The sample size of the number of Hispanic students who participated in Study Island for Biology was very small (n=5). Out of 5 Hispanic students who attempted the Biology test twice, one Hispanic student reached proficiency on their second attempt.

4 For this analysis, researchers only included White, Black, and Hispanic students as other racial group sample sizes were small. The researchers included the other racial groups analysis in the appendix.

Figure 10. Students who took the Biology Keystone exam, by race.



Demographic: Gender- Male students and Female students

Female students in Study Island did not make significant growth across Biology test attempts. Although the mean score across the first and second test attempts decreased by 5.5 points, it was not a statistically significant decreased between the first attempt and the second attempt. Out of 44 female students who attempted the Biology test twice, 8 reached proficiency on their second attempt.

Male students in Study Island did not make significant growth across Biology test attempts. Although the mean score across the first and second test attempts decreased by 3.22 points, it was not a statistically significant decrease between the first attempt and the second attempt. Out of 39 male students who attempted the Biology test twice, 4 students reached proficiency on their second attempt.

Figure 11. Students who took the Biology Keystone exams, by gender.

Demographic: Economically disadvantaged students in comparison to others

Economically disadvantaged students in Study Island did not make significant growth across Biology test attempts. Although the mean score across the first and second test attempts decreased by 2.75 points, it was not a statistically significant decrease between the first attempt and the second attempt. Out of 16 economically disadvantaged students who attempted the test twice in Biology, one student reached proficiency on their second attempt.

Non-economically disadvantaged students in Study Island did not make significant growth across Biology test attempts. Although the mean score across the first and second test attempts decreased by 3.45 points, it was not a statistically significant decrease between the first attempt and the second attempt. Out of 259 non-economically disadvantaged students who attempted the Biology test twice, 11 students achieved proficiency on their second test attempt.

Figure 12. Students who took the Biology Keystone exams, by economically disadvantaged status.

Literature

Hispanic students and non-IEP students showed significant growth in Literature after participating in Study Island. All changes in mean test scores are standardized and displayed as t-values in Tables 2 and 4 in Appendix A.

Demographics: Race- White, Black, and Hispanic students[5]

White students in Study Island did not make significant growth across Literature test attempts. Although the mean score across the first and second test attempts increased by 3.31 points for two-time test takers, it was not a statistically significant increase. Test scores increased by 10.25 between the first and second attempts for three-time test takers, which was also not statistically significant. The mean score from the second test attempt to third test attempt increased by 13 points, which was not a statistically significant increase in scores. However, the sample size of the number of students who took the test a third time was very small (n=2). Out of 165 White students who had two test attempts in Literature, 57 reached proficiency on their second test attempt. Out of 2students with three test attempts, one reached proficiency on their third test attempt.

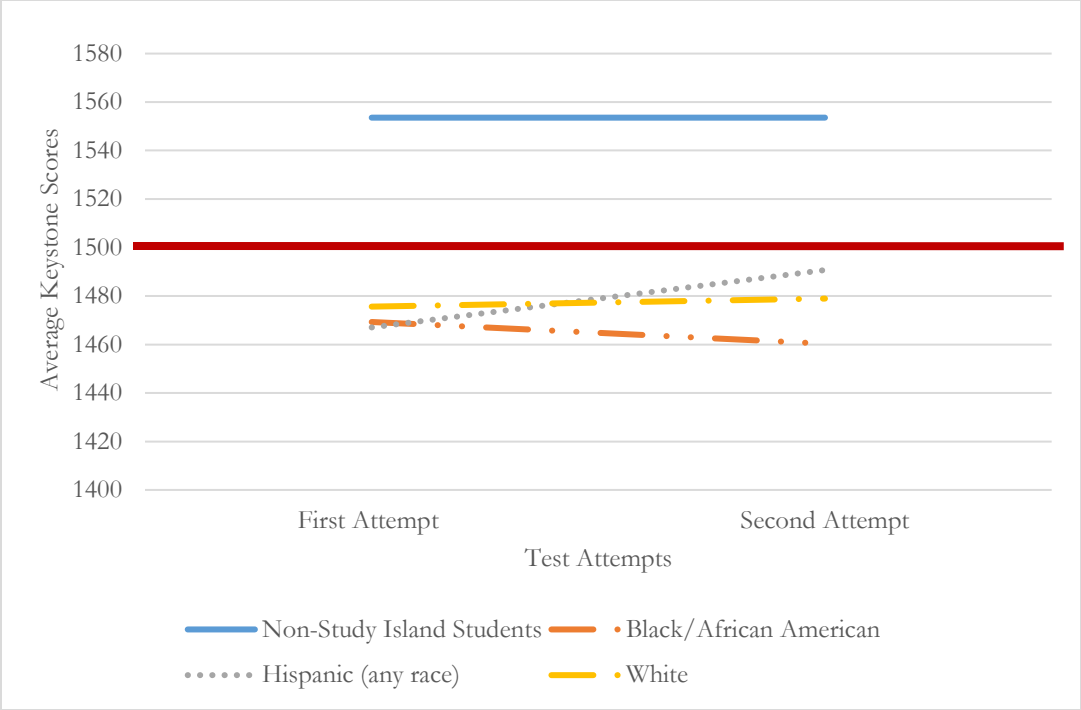
Black students in Study Island did not make significant growth across test attempts. Although the mean score across the first and second test attempts increased by 13 points for two-time test takers, it was not a statistically significant decrease between the first and second attempt. There were no students who were recorded as having taken the test a third time, which did not allow for comparison across second and third attempts. Out of 18 Black students who had two test attempts in Literature, 4 reached proficiency on their second test attempt.

Hispanic students in Study Island made significant growth in Literature exam scores across the first and second attempt. The mean score across the first and second test attempts increased by 23.67 points, and this was a statistically significant increase between the first attempt and the second

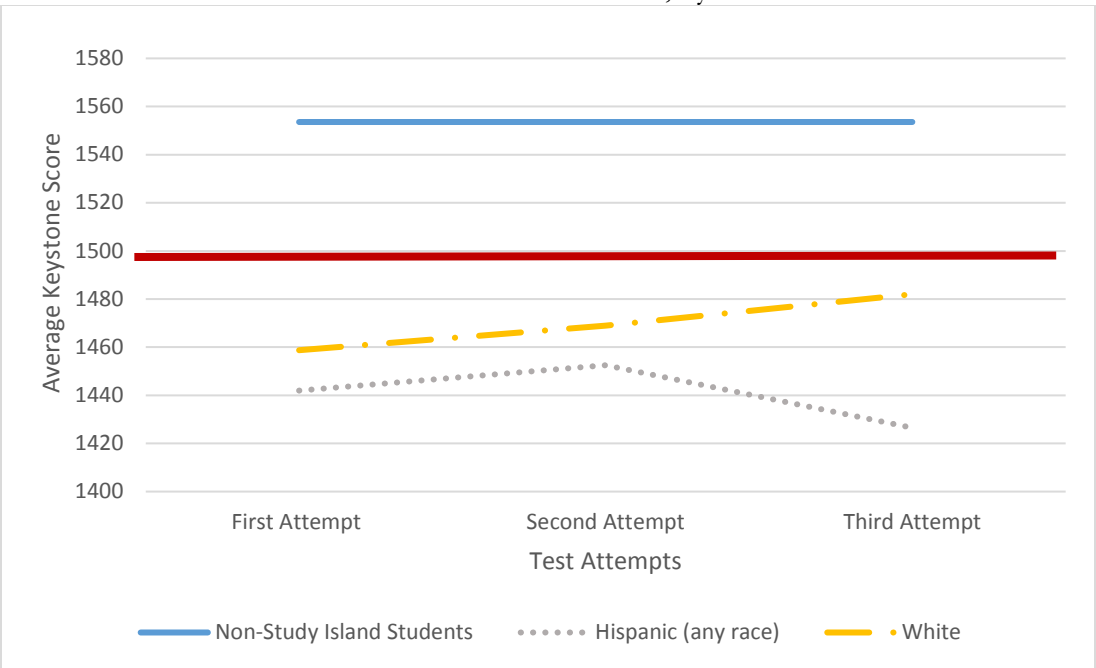
[5] For this analysis, researchers only included White, Black, and Hispanic students as other racial group sample sizes were small. The researchers included the other racial groups analysis in the appendix.

attempt. The mean score from the first to second test attempt among three-time test takers increased by 10.5 points and lowered by 26 points between second and third test attempts, which could not be tested for statistical significance as only one student took the Literature test a third time. Out of 12 Hispanic students who had two test attempts in Literature, 3 reached proficiency on their second test attempt. The one Hispanic student with three test attempts did not reached proficiency.

Figure 13. Students who took the Literature exam, by race.
Panel A. Students who took the Literature exam twice, by race.



Panel B. Students who took the Literature exam three times, by race.



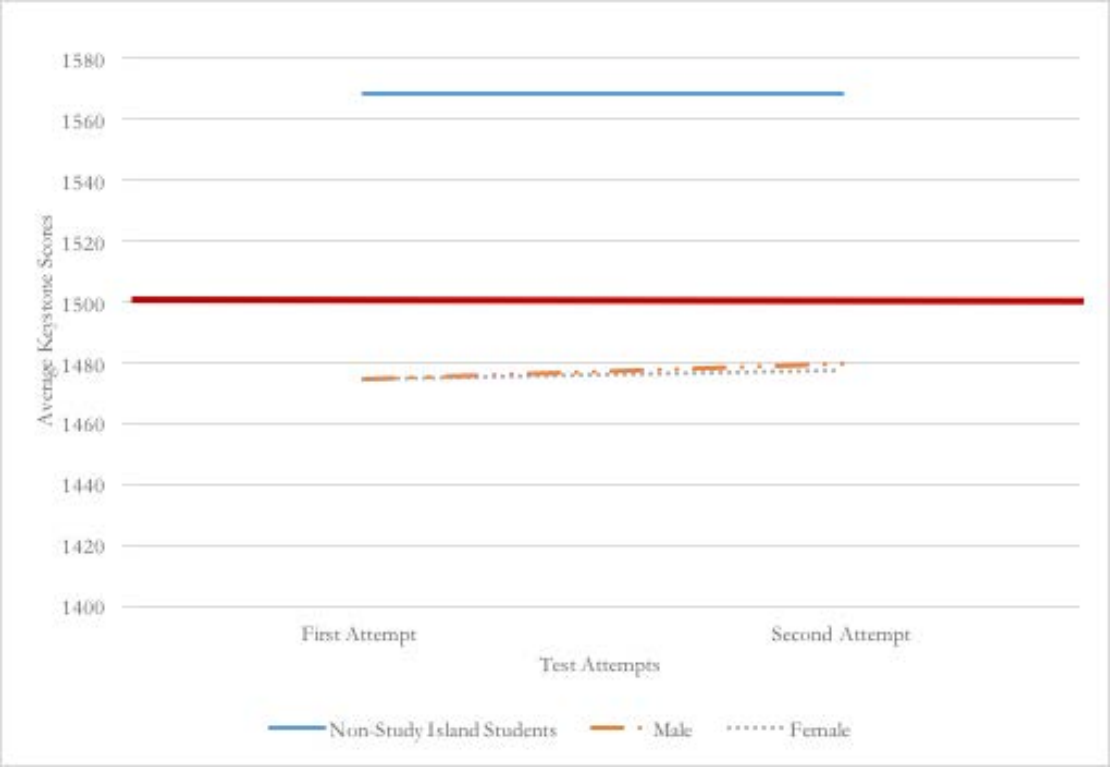
Demographic: Gender- Male students and Female students

Female students in Study Island did not make significant growth across test attempts. Although the mean score across the first and second test attempts increased by 3.16 points, it was not a statistically significant increase between the first attempt and the second attempt. The mean score from the second test attempt to third test attempt lowered by 8 points, which was also not a statistically significant decrease in scores. However, the sample size of the number of students who took the test a third time was very small (N=2). Out of 82 female students who had two test attempts in Literature, 28 reached proficiency on their second test attempt. Out of 3 female students with three test attempts, one reached proficiency on their third test attempt.

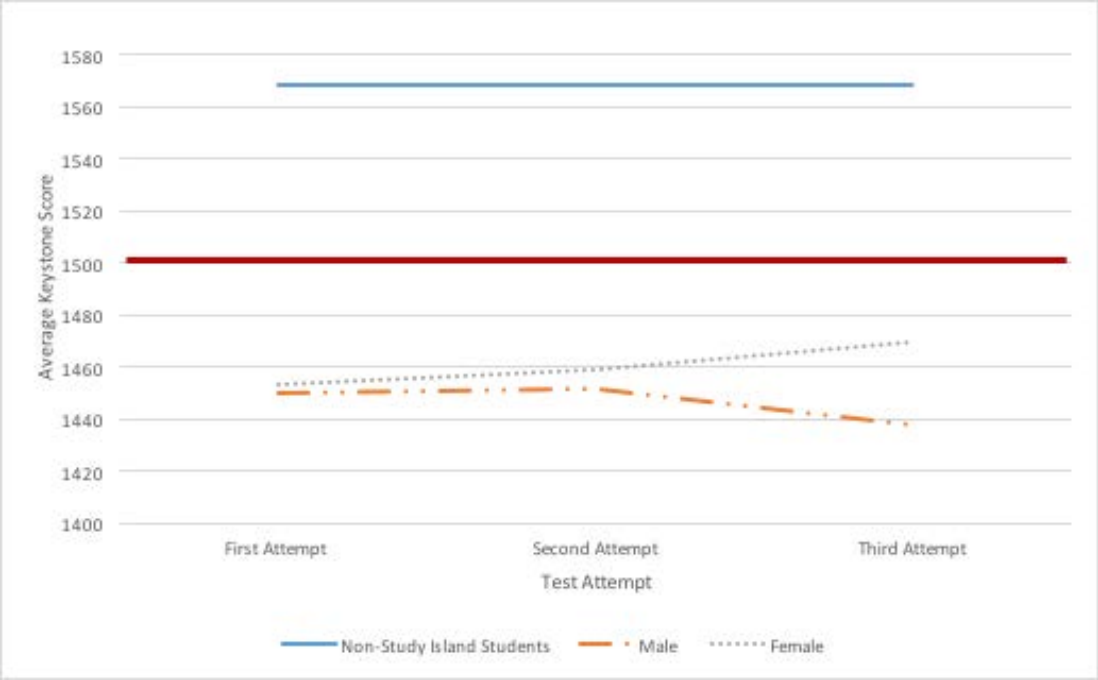
Male students in Study Island did not make significant growth across test attempts. Although the mean score across the first and second test attempts increased by 4.94 points, it was not a statistically significant increase between the first attempt and the second attempt for two-time test takers. The mean test scores between attempt one and two for three-time test takers grew by 2 points which could not be tested for statistical significance. The mean score from the second test attempt to third test attempt lowered by 14 points, but could not be tested for statistical significance as only one individual took the test three times in Literature. Out of 123 male students who had two test attempts in Literature, 40 reached proficiency on their second test attempt. The one student who had three test attempts did not meet proficiency.

Figure 14. Students who took the Literature exam, by gender.

Panel A. Students who took the Literature exam twice, by gender.



Panel B. Students who took the Literature exam three times, by gender.

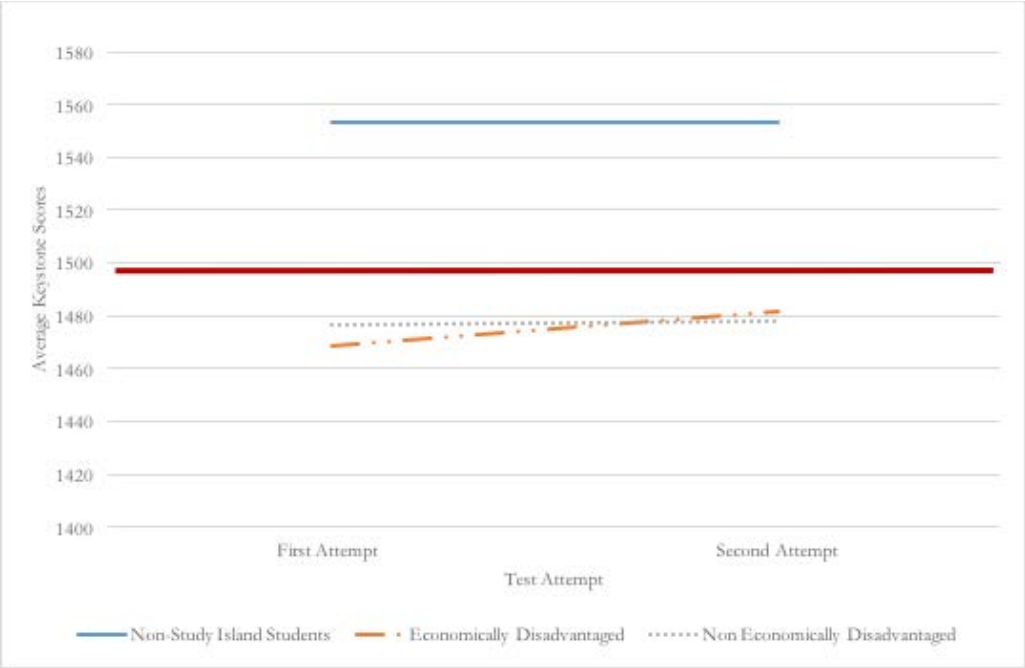


Demographic: Economically disadvantaged students in comparison to others

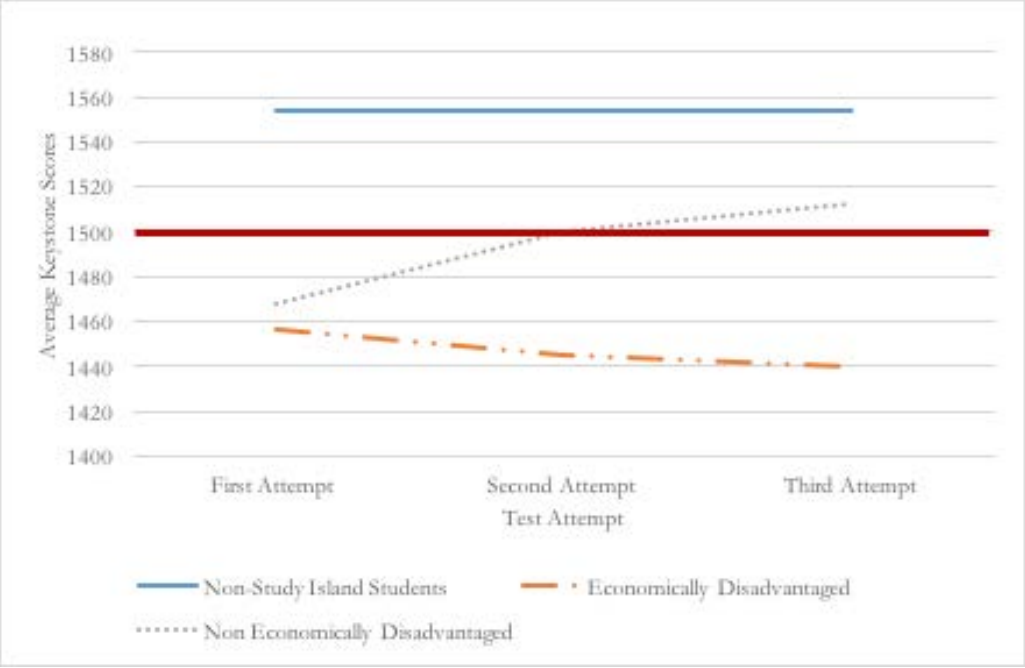
Economically disadvantaged students in Study Island did not make significant growth across Literature test attempts. Although the mean score for two-time test takers between the first and second test attempts increased by 13.1 points, it was not a statistically significant increase between the first attempt and the second attempt. The mean score decreased by 0.75 between attempts one and two for three-time test takers. The mean score between the second test attempt to third test attempt lowered by 6 points, which was also not a statistically significant decrease in scores. However, the sample size of the number of students who took the test a third time was very small (N=2). Out of 42 economically disadvantaged students with two test attempts for Literature, 17 reached proficiency on their second test attempt. Out of 2 economically disadvantaged students with three test attempts, none reached proficiency.

Non-economically disadvantaged students in Study Island did not make significant growth across test attempts. Although the mean score across the first and second test attempts increased by 1.72 points for two-time test takers, it was not a statistically significant increase between the first attempt and the second attempt. There was one non-economically disadvantaged student who had three test attempts for literature. This student made a gain of 32.3 points between the first and second attempts. We could not assess statistical significance as only one student took the Literature test three times only one student took the exam three times. This student had a gain of 12 points from the second test attempt and third test attempt and reached proficiency (scored over 1500) on the third test attempt. Out of 164 non-economically disadvantaged students who had two test attempts in Literature, 51 reached proficiency on their second test attempt.

Figure 15. Students who took the Literature Keystone exam, by economically disadvantaged status. Panel A. Students who took the Literature Keystone exam twice, by economically disadvantaged status.



Panel B. Students who took the Literature Keystone exam three times, by economically disadvantaged status.



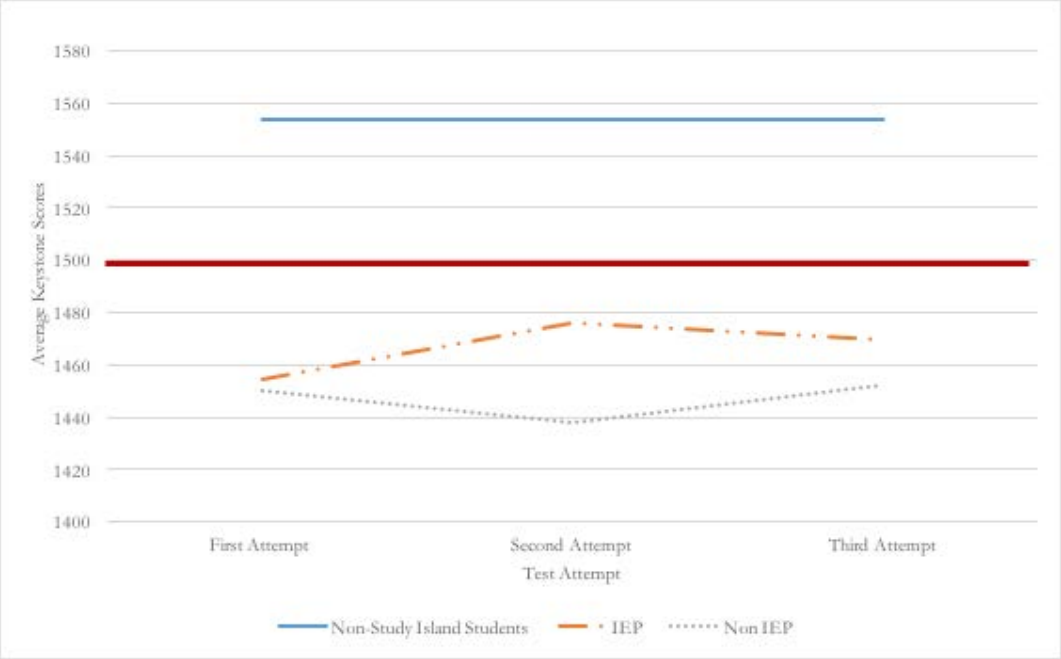
Demographic: Students on an Individualized Education Plan in comparison to those who are not

IEP students in Study Island did not make significant growth across Literature test attempts. Although the mean score across the first and second test attempts decreased by 3.49 points, it was not a statistically significant decrease between the first attempt and the second attempt for two-time test takers. The mean score between first and second attempts for three-time test takers grew by 21.5 which was not statistically significant. The mean score from the second test attempt to third test attempt decreased by 7 points, which was also not a statistically significant decrease in scores. However, the sample size of the number of students who took the test a third time was very small (N=2). Out of 68 IEP students who had two test attempts in Literature, 10 students reached proficiency on their second test attempt. Out of 2 IEP students with three test attempts for Literature, 1 reached proficiency on their third test attempt.

Non IEP students in Study Island did not make significant growth in Literature exam scores between the first and second attempt. The mean score between the first and second test attempts increased by 7.76 points among two time test takers, and this was not a statistically significant increase between the first attempt and the second attempt for two time test takers. The mean score between the first and second test attempts decreased by 12 points between attempts one and two among three time test takers, and this was not a statistically significant increase. The mean score from the second test attempt to third test attempt increased by 14 points, which could not be tested for statistical significance as only one student took the Literature test a third time. Out of 138 non-IEP students who had two test attempts in Literature, 58 reached proficiency on their second test attempt. The one IEP student who had three test attempts did not reach proficiency.

Figure 16. Students who took the Literature Keystone exams, by IEP status.
Panel A. Students who took the Literature Keystone exams twice, by IEP status.

Panel B. Students who took the Literature Keystone exams three times, by IEP status.



Conclusions and Opportunities for Further Research

Conclusions

This report investigates the potential efficacy of Study Island use on Neshaminy School District's Keystone exam scores. We examined students' Literature, Biology, and Algebra I exam scores across test events for those who did not meet proficiency (i.e. score above 1500 on the first event), which is required to graduate in Pennsylvania. We could not directly analyze the efficacy of Study Island--whether Study Island has a causal relationship with growth in test scores--because we did not perform a randomized control trial (which is the only way to determine causation). This report, however, will contribute to the literature as a preliminary examination of Study Island's potential effects in Neshaminy School District. It examines the growth of students who participated in Study Island for Algebra I, Biology, and Literature. There is only one past study that examined Study Island's usage for math, science, and reading (Bernard, 2013). It is notable that our study examines the demonstrated growth of Study Island participants by demographic groups, as there is a marked lack of literature regarding the topic.

To address the first research question ("Is Study Island more effective in one content area than another"), we analyzed trends in growth across test attempts for all students in Study Island differentiated by subject name to determine if students made more growth in certain subjects after participating in Study Island. We differentiated between students who took the test two times and those who sat the exam three times. Overall, students demonstrated significant growth for Algebra I but not for Biology and Literature. Students who had two test attempts as well as those with three test attempts showed significant growth in Algebra I.

To answer the second research question we investigated the demonstrated growth of different demographic groups of students on the three Keystone content areas to determine if Study Island contributes to improved academic outcomes for underperforming students. To address this, we analyzed the differences in test scores across test attempts for each demographic group by subject and examined the results for students who took the test two times compared with those who took the exam three times. First, the following demographic groups showed significant growth in Algebra I (which was the only subject in which students demonstrated significant growth): White, female, male, non-historically underperforming, non-economically disadvantaged, IEP, non-IEP, and non-ELL students.

Only White students demonstrated significant growth after Study Island participation for the Algebra I Keystone. White students who had two test attempts as well as three test attempts showed significant growth. This was not true for Black and Hispanic students, although this might be partially explained by the fact that, while there were 275 White students, just 29 Black students and 19 Hispanic students participated in Study Island for Algebra I. It is still worth noting that there is more than a 20-point gap between White students and racial minority students in Algebra I Keystone test scores. In order to gain a clearer understanding regarding different racial/ethnicity group growth rates, the number of students who participated in Study Island for each racial/ethnicity group should be similar.

Female and male students with two test attempts who participated in Study Island for Algebra I had significant growth. For those who had three test attempts, however, only females made significant growth. Although we cannot offer definite statements regarding this phenomenon, a possible explanation might be that girls in general tend to have higher levels of academic motivation (Bugler,

McGeown, & St. Clair-Thompson, 2013). The potential impact of gender motivation on Study Island investment needs further investigation since we do not know the dosage of students and whether it was possible for students to voluntarily devote more time to Study Island participation.

There were significant gains in Algebra I exam scores for students who were not part of the historically underperforming group. Non-historically underperforming students made significant gains in both the cases of students who had two test attempts and students who had three test attempts. None of the historically underperforming students (those with two test attempts or those with three test attempts) showed significant growth after Study Island participation. Non-historically underperforming students who had three test attempts had a mean score of 1483.41 for their first attempt and after making significant growth, scored a mean of 1502.09 for their third attempt, meeting the proficiency score of 1500, while historically underperforming students who had three test attempts had a mean score of 1460.92 for their first test attempt, lagging behind their non-historically underperforming counterparts by more than 20 points, and scored a mean of 1482.65 for their third test attempt, still smaller than non-historically underperforming students' first attempt mean score with the 20 point gap persistent. This result further reinforces the notion that historically underperforming students require additional academic support to reduce the achievement gap that exists between these students and non-historically underperforming students and ultimately, reach proficiency.

Students who were not economically disadvantaged showed significant growth across test attempts in Algebra I. Those who had two test attempts and three test attempts both demonstrated significant growth. Students who were economically disadvantaged, however, did not show significant growth across any test attempts. Non-economically disadvantaged students who had three test attempts had a mean score of 1472.84 for their first attempt and after making significant growth, scored a mean of 1496.94 for their third attempt, coming close to the proficiency score of 1500, while economically disadvantaged students who had three test attempts had a mean score of 1468 for their first test attempt and scored a mean of 1480.81 on their third test attempt. The initial gap between students who took the test three times was not large at only a difference of about 4 points. However, the achievement gap *grew* over increased test attempts for this demographic group, amounting to more than a 15-point difference. Similar to their historically underperforming peers, economically disadvantaged students seem to require additional academic support to catch up with their peers and reach proficiency.

Students without an IEP made significant growth in Algebra I after participating in Study Island. Students who had two test attempts and three test attempts both showed significant growth. However, students with an IEP did not show significant growth in Algebra I after participating in Study Island. Although students with an IEP who had three test attempts show a large amount of growth with each test event (mean score from 1449.44 for the first test attempt to 1461.31 for the second test attempt to 1481.69 for the third test attempt), this growth might not have been statistically significant due to the small sample size of 8 students. Also, the third test attempt mean score for IEP students was 1481.69 which was still lower than the first test attempt mean score of non-IEP students with three test attempts of 1482.13. These students also had a third test attempt mean score of 1496.5, which is close to the proficiency score of 1500. These results demonstrate that IEP students require additional support in order to meet proficiency.

Similar to the IEP category, non-ELL students who participated in Study Island made significant growth in Algebra I after participating in Study Island while ELL students did not. Non-ELL

students who had two test attempts and three test attempts both showed significant growth. It should be noted that there were only 10 ELL students who participated in Study Island for Algebra I and that no ELL students that had a third test attempt for Algebra I.

To summarize, the demographic groups of students that showed significant growth and came close to achieving proficiency in Algebra I were racial majorities (White students) and students who were of higher socioeconomic status (non-historically underperforming students, non-economically disadvantaged students, non-IEP students, and non-ELL students). Also, students from disadvantaged backgrounds require additional support in order to meet proficiency and fulfill their graduation requirements.

There were no demographic groups that demonstrated significant growth in Biology. This finding, however, requires more investigation because Biology only had two recorded test attempts and there were far less students who participated in Biology (83, compared with 343 for Algebra I and 209 for Literature. Examining more students who participated in Study Island for Biology would offer a clearer understanding of whether the growth that students in Biology is significant.

There were also several demographic groups that showed a decrease in mean test scores for Biology, although none of these decreases were statistically significant. White, Black, female, male, economically disadvantaged, and not economically disadvantaged students showed non-significant decreases in their mean scores after Study Island participation. This decrease in mean scores merits further investigation.

Among students who participated in Study Island for Literature, Hispanic students and non-IEP students showed significant growth. Both Hispanic students and non-IEP students who had two test attempts for Literature demonstrated significant growth. There was only one Hispanic student and one non-IEP student who had three test attempts for Literature. In contrast to these groups that made significant growth, Black students and IEP students showed non-significant decreases in their mean scores after Study Island participation. It would help to further explain this phenomenon with a larger sample size, since there were just 17 Black students and 70 IEP students who participated in Study Island for Literature. This decrease in mean scores for Black and Hispanic students would be a point for further investigation.

Opportunities for Further Research

There is a marked dearth of research on the efficacy of Study Island in schools, leaving many opportunities for further studies. We recommend that in the future the following be examined: dosage in Study Island participation, a more in depth analysis of historically underperforming students and those with IEPs, and supervision and progress made during programming. To truly determine the efficacy of Study Island, however, it would be necessary to perform an experiment to ascertain if there is any sort of causal effect of the program. Even the most careful analysis is limited to correlation unless done using an experimental design.

Bruce-Simmons (2013) found a difference in effect based on amount of time students spent using Study Island (once per week or twice per week), which indicates a potential influence of dosage on results. Future research should examine intensive involvement in Study Island; it is important to better understand the extant dosage effect, which is currently missing from the literature. In the future, we recommend that the Neshaminy School District track how often and for how long

students participate in Study Island in order to best understand the potential influence of the program.

Our research indicates that there is a significant difference in the representation of historically underperforming and IEP students, which leads us to recommend that the district investigate the possibility of an opportunity gap in student learning. It appears that Study Island is especially ineffective for some of the most vulnerable populations in the Neshaminy School District. According to the Pennsylvania Department of Education, historically underperforming students are "a non-duplicated count of students with disabilities, economically disadvantaged students, and English Language Learners enrolled for a full academic year taking the PSSA/PASA/Keystone Exams" (2015, p. 5). By their definition, students in more than one of the individual groups (for example, if a student is both in special education and an English Language Learner), are only included in the Historically Underperforming student group one time (Pennsylvania Department of Education, 2015). It would be worthwhile to examine the reasons why historically underperforming and IEP students are overrepresented among students who are asked to participate in Study Island (those that scored below proficient on their Algebra I, Literature, or Biology Keystone). This overrepresentation implies that there is perhaps a greater difference in preparation for those specific student groups in comparison with their peers.

We recommend the Neshaminy School District investigate the supervision offered to students as they persist through the Study Island program. It would be beneficial to track who is tracking students' progress as they use the program. It is important to better understand the ease of access, as difficulty accessing the online program or using the computer could impact students' experience with Study Island and their progress with the program. Ideally, the next step in assessing Study Island's impact is to perform a randomized control trial.

References

Ascher, C. (2006). NCLB's supplemental educational services: Is this what our students need? *The Phi Delta Kappan, 88(2),* 136-141.

Barrow, L., Markman, L., & Rouse, C. E. (2009). Technology's edge: The educational benefits of computer-aided instruction. *American Economic Journal: Economic Policy*, *1*(1), 52-74.

Benthall, S. A. (2015). Is study island just a craze? A comparison of student Achievement test scores in math before and after a Technology-integrated intervention (Doctoral Dissertation). Retrieved from ProQuest 3721702

Bernard, B. T. (2013). Student Achievement and the Use of the Program Study Island (Doctoral Dissertation). Retrieved from UMI Dissertation Publishing 3599008.

Bernard, R.M, Abrami, P.C., Lou, Y., Borohovshi, E, Wade, A., Wozney, L., Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74(3),* 379-439.

Bracht, N. T. (2011). The Relationship Between Study Island and Student Achievement (Doctoral Dissertation). Retrieved from UMI Dissertation Publishing 3453962.

Bruce-Simmons, C. (2013). The Impact of Computer-Assisted Instruction on Mathematics Achievement of Underachieving Fifth-Grade Students (Unpublished Doctoral Dissertation). South Carolina State University, Orangeburg, South Carolina.

Bugler, M., McGeown, S. P., & St. Clair-Thompson, C. (2013). Gender differences in adolescents' academic motivation and classroom behaviour. *Educational Psychology*, 35(5), 541-556.

Burch, P., Steinberg, M., & Donovan, J. (2007). Supplemental educational service and NCLB: Policy assumptions, market practices, emerging issues. *Educational Evaluation and Policy Analysis, 29(2),* 115-133.

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts* (NCEE 2009-4041). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Chappell, S., Nunnery, J., Pribesh, S., & Hager, J. (2011). A meta-analysis of supplemental educational services (SES) provider effects on student achievement. *Journal of Education for Students Placed at Risk, 16(1),* 1-23.

Common Core State Standards Initiative. (2017). Preparing America's students for success. Retrieved April 10, 2017 from http://www.corestandards.org.

Conley, D.T. (2014). The common core state standards: insight into their development and purpose. *Council of Chief State School Officers,* 1-9.

Cross, C. (2010). *Political Education: National Policy Comes of Age.* New York:Teachers College Press.

Dube, P. J. (2011). Attempting to Improve Standardized Test Results Using Study Island's Web-Based Mastery Program (Unpublished Master's Thesis). Michigan Technological University, Houghton, Michigan.

Duncan, A. (2009). *States will lead way to reform.* http://www2.ed.gov/news/speeches/2009/06/06142009.html

Friedman, T. (2003, April 03). It's a Flat World After all. *New York Times.*

Gaytan, J. & McEwen, B.C. (2007). Effective online instructional and assessment strategies. *American Journal of Distance Education, 21(3),* 117-132.

Heckman, J. J., & LaFontaine, P. A. (2010). The American high school graduation rate: Trends and levels. *The review of economics and statistics*, *92*(2), 244-262.

Heinrich, C.J., Meyer, R.H., & Whitten, G. (2010). Supplemental education services under No Child Left Behind: Who signs up, and what do they gain? *Educational Evaluation and Policy Analysis, 32(2),* 273-928.

Kentnor, H.E. (2015). Distance education and the evolution of online learning in the United States. *Curriculum and Teaching Dialogue, 17(1&2)*, 21-34.

Laing, L. J. (2011). The Effects of the Web-Based Mastery Program Study Island on Tennessee Middle School Math Standardized Test Scores (Doctoral Dissertation). Retrieved from UMI Dissertation Publishing 3498294.

Leachman, M., Albares, N., Masterson, K., & Wallace, M. (2016). Most states have cut school funding, and some continue cutting. *Center on Budget and Policy Priorities*, 4.

Madden, N. A., & Slavin, R. E. (1987). Effective Pull-out Programs for Students at Risk.

National Center for Education Statistics. (2016). School Directory Information: Neshaminy HS. Retrieved April 2017 from https://nces.ed.gov/ccd/schoolsearch/school_detail.asp?Search=1&DistrictID=4216410&ID=421641006455

National Center for Education Statistics. (2016b). District Directory Information: Neshaminy SD. Retrieved April 2017 from https://nces.ed.gov/ccd/districtsearch/district_detail.asp?Search=2&details=1&ID2=4216410&DistrictID=4216410

Parlapanides, T. (2008). Effects of A Technology Treatment On Student Scores On The Standardized Grade 8 Proficiency Assessment (Gepa) In New Jersey. Retrieved from UMI Dissertation Publishing 3436907.

Pennsylvania Department of Education. (2015). Pennsylvania school performance profile: Frequently asked questions. Retrieved April, 10 2017 from http://paschoolperformance.org/FAQ

Pennsylvania Department of Education. What students and parents need to know about Pennsylvania's new state high school graduation requirements. Retrieved April 10, 2017 from (http://static.pdesas.org/Content/Documents/PDE_PSFactSheet.pdf).

Pennsylvania Department of Education. (2016b). Identification, Placement, and Exit. Retrieved April 30, 2017 from http://www.education.pa.gov/Teachers%20-%20Administrators/Curriculum/English%20As%20A%20Second%20Language/Pages/Identification-Placement-Exit.aspx#.Vbt7YmXD-Uk

Pennsylvania Department of Education. What students and parents need to know about Pennsylvania's new state high school graduation requirements. Retrieved April 10, 2017 from (http://static.pdesas.org/Content/Documents/PDE_PSFactSheet.pdf).

Pennsylvania Department of Education. (2016a). IEPS and 504 Service Agreements. Retrieved April 30, 2017 from http://www.education.pa.gov/K-12/Homebound%20Instruction/Pages/IEPs-and-504-Service-Agreements.aspx#tab-1

Ryan, J.E. (2004). The Perverse Incentives of the No Child Left Behind Act. *NYU Law Review*, July, 2004.

Shomekaer, T. L. (2013). Effect of Computer-Aided Instruction on Attitude and Achievement of Fifth Grade Math Students (Doctoral Dissertation). Retrieved from UMI Dissertation Publishing 3595355.

Steinberg, A, and Cheryl A. A.. "Raising Graduation Rates in an Era of High Standards: Five Commitments for State Action." *Jobs for the Future* (2008).

Study Island. (2017a). Retrieved April, 2017 from: www.studyisland.com/pricing/home

Study Island. (2017b). Retrieved April, 2017 from: www.studyisland.com/about

Styers, M.K. (2012a). Developing student mathematics skills: How Study Island aligns with best practice. Magnolia Consulting, LLC.

Styers, M.K. (2012b). Advance student achievement while lowering costs through digital learning. Magnolia Consulting, LLC.

Viviano, T. A. (2011). What Impact Does the Software Study Island Have on 4sight, PSSA, and NOCTI Assessments of Part-Time CTE Students? (Doctoral Dissertation). Retrieved from UMI Dissertation Publishing 3500974.

Youtube. (2017). Search: Study Island. Retrieved from https://www.youtube.com/results?search_query=study+island

Appendices

Appendix A

Appendix B

Appendix A

Table 1. Overall means, t-test statistics, and p-values for test events, by number of test events and test type.

	Attempt 1	Attempt 2	Attempt 3	Difference	p-value
Students who took the test twice (N=319)	1468.52 (29.41)	1477.62 (39.91)		9.11***	.00
Students who took the test three times (N=24)	1471.23 (23.20)	1474.04 (19.40)		2.81	.65
Students who took the test three times (N=24)		1474.04 (19.40)	1491.563 (31.61)	17.52**	.03
Students who took the test twice (N=83)	1475.72 (25.86)	1472.42 (32.75)		-3.30	.61
Students who took the test twice (N=205)	1474.83 (25.92)	1479.02 (40.10)		4.19	.21
Students who took the test three times (N=4)	1452.38 (10.77)	1453.88 (32.74)		1.50	.93
Students who took the test three times (N=3)		1453.88 (32.74)	1463.5 (43.91)	9.63	.75

Table 2. Means, t-test statistics, and p-values for test events, by number of test events, subject, and students' race/ethnicity.

	Attempt 1	Attempt 2	Attempt 3	Difference	p-value
Panel A: Algebra					
White students who took the test twice (N=259)	1471.54 (25.7)	1480.11 (38.06)		8.57***	0.0028
White students who took the test three times (N=16)	1469.66 (25.9)	1475.91 (20.15)		2.813	0.4526
White students who took the test three times (N=16)		1475.91 (20.15)	1495.09 (30.86)	19.18**	0.0478
Black students who took the test twice (N=25)	1441.78 (44.11)	1452.7 (44.22)		10.92	0.3864
Black students who took the test three times (N=4)	1467.75 (24.25)	1478.25 (7.63)		10.5	0.4694
Black students who took the test three times (N=4)		1478.25 (7.63)	1470.39 (22.91)	-7.86	0.5614
Hispanic students who took the test twice (N=18)	1469.11 (28.62)	1473.78 (42.84)		4.67	0.7034
Hispanic students who took the test three times (N=1)	1463.5	1434.5		-29	
Hispanic students who took the test three times (N=1)		1434.5	1445.5	11	
Panel B: Biology					
White students who took the test twice (N=71)	1477.22 (26.21)	1471.72 (32.54)		-5.5	0.2694
Black students who took the test twice (N=4)	1464.75 (21.21)	1453.75 (25.7)		-11	0.5384
Hispanic students who took the test twice (N=5)	1471.8 (26.8)	1477 (27.08)		5.2	0.769
Panel C: Literature					
White students who took the test twice (N=165)	1475.62 (24.49)	1478.93(39.6)		3.31	0.7992
White students who took the test three times (N=2)	1458.75 (12.37)	1469 (43.84)		10.25	0.8038
White students who took the test three times (N=2)		1469(43.84)	1482 (42.43)	13	0.8136

Black students who took the test twice (N=259)	1469.28 (31.81)	1460.33 (48.97)		-8.95	0.5206
Hispanic students who took the test twice (N=12)	1467.04 (29.84)	1490.71 (34.05)		23.67*	0.0844
Hispanic students who took the test three times (N=1)	1442	1452.5		10.5	
Hispanic students who took the test three times (N=1)		1452.5	1426.5	-26	

Table 3. Means, t-test statistics, and p-values for test events, by number of test events, subject, and students' gender.

	Attempt 1	Attempt 2	Attempt 3	Difference	p-value
Panel A: Algebra					
Male students who took the test twice (N=141)	1466.62 (31.29)	1475.55 (41.66)		8.57***	.00
Male students who took the test three times (N=12)	1470.08 (21.43)	1470.75 (18.93)		2.81	.45
Male students who took the test three times (N=12)		1470.75 (18.93)	1483.17 (24.45)	19.18**	.05
Female students who took the test twice (N=178)	1470.03 (27.84)	1479.258 (38.51)		9.23***	.01
Female students who took the test three times (N=12)	1472.38 (25.76)	1477.33 (20.13)		4.96	.61
Female students who took the test three times (N=12)		1477.33 (20.13)	1499.96 (36.57)	22.63*	.08
Panel B: Biology					
Male students who took the test twice (N=39)	1468.55 (21.34)	1465.13 (33.68)		-5.50	.27
Female students who took the test twice (N=44)	1482.08 (28.01)	1478.86 (30.86)		-3.22	.31
Panel C: Literature					
Male students who took the test twice (N=123)	1474.82 (25.81)	1479.76 (40.27)		4.94	.80
Male students who took the test three times (N=1)	1450	1452			.80
Male students who took the test three times (N=1)		1452	1438		.81
Female students who took the test twice (N=82)	1474.85 (26.237)	1477.92 (40.08)			.56
Female students who took the test three times (N=3)	1453.17	1459.17 (37.94)			.81
Female students who took the test three times (N=3)		1459.17 (37.94)	1469.25 (60.46)		.83

Table 4. Means, t-test statistics, and p-values for Algebra and Literature test events, by number of test events, subject, and IEP or non-IEP students.

	Attempt 1	Attempt 2	Attempt 3	Difference	p-value
Panel A: Algebra					
IEP students who took the test twice (N=83)	1451.86 (34.19)	1452.21 (38.80)		.35	.95
IEP students who took the test three times (N=8)	1449.44 (27.90)	1461.31 (25.17)		11.88	.39
IEP students who took the test three times (N=8)		1461.31 (25.17)	1481.69 (27.57)	20.38	.15
Non-IEP students who took the test twice (N=236)	1474.381 (25.10)	1486.56 (36.36)		12.18***	.00
Non-IEP students who took the test three times (N=16)	1482.125 (9.08)	1480.41 (12.33)		-1.72	.66
Non-IEP students who took the test three times (N=16)		1480.41 (12.33)	1496.5 (33.16)	16.09*	.08
Panel B: Literature					
IEP students who took the test twice (N=68)	1463.48 (30.27)	1459.99 (40.71)		-3.49	.57
IEP students who took the test three times (N=2)	1454.75(18.03)	1476.25(33.59)		21.50	.51
IEP students who took the test three times (N=1)		1476.25(33.59)	1469.25(60.46)	-7.00	.90
Non-IEP students who took the test twice (N=138)	1480.25 (21.52)	1488.01 (36.67)		7.76	.03
Non-IEP students who took the test three times (N=1)	1450	1438		-12.00	.81
Non-IEP students who took the test three times (N=1)		1438	1452	14.00	.83

Table 5. Means, t-test statistics, and p-values for Algebra test events, by number of test events, and Historically underperforming and non-historically underperforming

	Attempt 1	Attempt 2	Attempt 3	Difference	p-value
Panel A: Algebra					
Historically Underperforming students who took the test twice (N=133)	1458.09 (35.56)	1463.02 (43.37)		4.93	.31
Historically Underperforming students who took the test three times (N=13)	1460.92 (26.36)	1468.69 (23.63)		7.77	.44
Historically Underperforming students who took the test three times (N=13)		1468.69 (23.63)	1482.65 (21.54)	13.96	.15
Non-Historically Underperforming students who took the test twice (N=186)	1475.978 (21.227)	1488.06 (33.66)		12.08***	.00
Non-Historically Underperforming students who took the test three times (N=11)	1483.409 (10.146)	1480.36 (10.71)		-3.05	.50
Non-Historically Underperforming students who took the test three times (N=11)		1480.36 (10.71)	1502.09 (38.94)	21.73*	.09

Table 6. Means, t-test statistics, and p-values for test events, by number of test events, subject, and economically and non-economically disadvantaged status.

	Attempt 1	Attempt 2	Attempt 3	Difference	p-value
Panel A: Algebra					
Economically disadvantaged students who took the test twice (N=70)	1462.77 (36.34)	1472.879 (45.257)		10.11	.15
Economically disadvantaged students who took the test three times (N=8)	1468 (21.75)	1473.19 (24.13)		5.19	.66
Economically disadvantaged students who took the test three times (N=8)		1473.19 (24.13)	1480.81 (17.65)	19.18**	.48
Non-economically disadvantaged students who took the test twice (N=249)	1470.139 (27.02)	1478.95 (38.27)		8.82***	.00
Non-economically disadvantaged students who took the test three times (N=16)	1472.84 (24.42)	1474.47 (17.47)		1.63	.83
Non-economically disadvantaged students who took the test three times (N=16)		1474.47 (17.47)	1496.94 (35.98)	22.47**	.03
Panel B: Biology					
Economically disadvantaged students who took the test twice (N=16)	1475.81 (18.58)	1472.36 (30.12)		-3.45	.43
Non-economically disadvantaged students who took the test twice (N=259)	1475.38 (46.21)	1472.63 (43.26)		-2,75	
Panel C: Literature					
Economically disadvantaged students who took the test twice (N=42)	1468.79 (29.02)	1481.89 (46.19)		13.10	.12
Economically disadvantaged students who took the test three times (N=2)	1446 (5.66)	1445.25(10.25)		-.75	.94
Economically disadvantaged students who took the test three times (N=2)		1445.25(10.25)	1439.25(18.03)	-6.00	.72
Non-economically disadvantaged students who took the test twice (N=164)	1476.23 (24.93)	1477.95 (38.61)		1.72	.63
Non-economically disadvantaged students who took the test three times (N=1)	1467.7	1500		32.30	
Non-economically disadvantaged students who took the test three times (N=		1500	1512	12.00	

Table 7. Means, t-test statistics, and p-values for Algebra I tests, by number of test events, subject, and English Language Learner status.

	Attempt 1	**Attempt 2**	**Attempt 3**	**Difference**	**p-value**
Panel A: Algebra					
ELL students who took the test twice (N=10)	1454.1 (43.6)	1480.3 (55.56)		4.93	.26
Non-ELL students who took the test twice (N=309)	1468.99 (28.82)	1477.53 (39.42)		8.55***	.00
Non-ELL students who took the test three times (N=24)	1471.23 (23.20)	1474.04 (19.40)		2.81	.65
Non-ELL students who took the test three times (N=24)		1474.04 (19.40)	1491.56 (31.61)	17.52**	.03

Appendix B. Study Island program offerings in comparison with competitors

		Program Name						
Program Details	**Explanation**	Study Island	Edgenuity	IXL	Achieve 3000	i-READY	Wowzers	Scoot Pad
Price	Amounts are based on price per student per school year**	NSD: $2.17* $20-$150	$350-$1000	$15	$40	$90	$30	$7
PreK-5	Curriculum is designed for Pre-Kindergarten through the fifth grade.	✓	✓	✓	✓	✓	✓	✓
6 to 8	Curriculum is designed for sixth through the eighth grade	✓	✓	✓	✓	✓	✓	✓
9 to 12	Curriculum is designed for ninth through the twelfth grade.	✓	✓	✓	✓	✓	-	-
History / Soc. Sci.	History, Social Studies, and Social Science Curriculum.	✓	✓	✓	-	-	-	✓
Foreign Language	Lessons in Spanish, French, German, Latin, or Mandarin.	-	✓	✓	✓	✓	-	-
Math	Mathematics Curriculum	✓	✓	✓	-	✓	✓	✓

Standard Aligned Math	Math lessons are aligned to state exam standards.	✓	✓	✓	-	✓	✓	✓
Science	Science Curriculum.	✓	✓	✓	✓	-	-	-
Standard Aligned Sci.	Science lessons are aligned to state exam standards.	✓	-	✓	-	-	-	-
ELA	English/ Language Arts Curriculum.	✓	✓	✓	✓	✓	-	✓
Standard Aligned ELA	ELA lessons are aligned to state exam standards.	✓	✓	✓	-	✓	-	✓
ESL/ ELL	Curriculum is designed for English Language Learners.	-	✓	-	✓	-	-	-
Novels	Curriculum is designed for use with provided e-books or graphic novels.	✓	✓	-	-	-	-	✓
SAT/ ACT	SAT/ ACT Exam Preparation	✓	✓	-	-	-	-	-
College and Career Readiness	Curriculum is designed to prepare students for college or for professional settings.	✓	✓	-	-	-	-	-
GED Prep	GED Exam Preparation	✓	✓	-	-	-	-	-

Benchmarking	An option for students to take online exams that will predict their scores on state aligned, standardize exams.	✓	✓	✓	-	-	✓	-
Adaptive Lessons	Questions are tailored to student's abilities based on previously answered questions.	✓	✓	✓	✓	✓	✓	✓
Quizzes	Assessments based on curriculum.	✓	✓	✓	✓	✓	✓	✓
Games	Educational games used as a reward for task completion, or as an engagement tool.	✓	✓	✓	✓	-	✓	-
Customizable Lessons	Teachers or parents have the ability to tailor lessons to specific students or classes.	✓	✓	✓	-	✓	✓	✓
Teacher Portal	Teachers can view and compare participation levels, achievement and progress information, and areas of weakness for their students. Often this	✓	✓	✓	-	✓	✓	✓

	comes with an ability to compare students to classmates and class to other classes.							
Parent Portal	Parents can log on, see a student's assignments, participation levels, achievement and progress information, and areas of weakness.	✓	✓	✓	-	✓	-	✓
Printables	Website allows teachers to create or download, and print paper worksheets, quizzes, and exams.	✓	✓	✓	-	-	-	-
Additional Elective Courses	A variety of supplemental courses, such as: AP exam prep, art, psychology, economics, computer science, and more.	✓	✓	-	-	-	-	-

* This price was arrived at by dividing Neshaminy's total Study Island expenditure ($4,483.41) by the number of participants (690) by number of years (3) between the academic years of 2013-2016. It is worth noting that estimates based on the Study Island website place the cost at $20-$150.

**Prices are all estimates, as real prices fluctuate based on number of students, length of contract, and breadth of services.