Spring 4-2011

The Relationship Between Study Island and Student Achievement

Nathan Travis Bracht
Lindenwood University

The Relationship Between Study Island and Student Achievement

by

Nathan Travis Bracht

A Dissertation submitted to the Education Faculty of Lindenwood University

in partial fulfillment of the requirements for the

degree of

Doctor of Education

School of Education

The Relationship Between Study Island and Student Achievement

by

Nathan Travis Bracht

This dissertation has been approved as partial fulfillment of the requirements for the

degree of

Doctor of Education

at Lindenwood University by the School of Education

Dr. Bryan Williams, Dissertation Chair	Date
Dr. Pam Sloan, Committee Member	4-29-11 Date
Dr. Jennifer Patterson, Committee Member	4-29-11 Date

Declaration of Originality

I do hereby declare and attest to the fact that this is an original study based solely upon my own scholarly work here at Lindenwood University and that I have not submitted it for any other college or university course or degree here or elsewhere.

Full Legal Name: Nathan Travis Bracht

Signature: N. Travis Bracht Date: 4-29-11

Acknowledgements

Thanks to the following people for encouragement, support, feedback, and direction throughout the entire dissertation process: Dr. Vicki Hedges-Oldani, Dr. Beth Kania-Gosche, Dr. Sherrie Wisdom, Dr. Pam Sloan, Dr. Jennifer Patterson, and Dr. Bryan Williams. Likewise, my wife Kelly provided valuable guidance during my dissertation endeavor while managing to finish her own dissertation at the same time.

I would also like to thank the school district in this study for providing access to the necessary data and being supportive of the study. It is truly a blessing to work with such dedicated professionals on a daily basis.

Finally, I must express a heartfelt thanks to Dr. Kelly Bracht, my wife, for her love and support. I could not have accomplished this without her guidance and encouragement.

Abstract

Many schools in America have issues with raising overall achievement as well as the achievement of subgroup populations on state tests required by No Child Left Behind. This quantitative study determined whether an online program called Study Island significantly effected overall and subgroup achievement on the Missouri Assessment Program (MAP) tests in communication arts and mathematics at the elementary and middle school levels. The results will inform school officials in this district and similar districts on whether Study Island can meet the needs of their teachers and students.

The students in the study began using Study Island in preparation for the 2009 MAP state tests. Therefore, the average scale scores before using Study Island (2008 MAP) and after using Study Island (2009 MAP) formed the basis for the data analysis. The *z* tests and *t* tests (95% confidence interval) performed on random samples from the total population and seven subgroup populations provided the results for this study. The subgroup populations for the district in this study included Asian/Pacific Islander, Black, Hispanic, White, Free and Reduced Lunch, Individualized Education Program, and Limited English Proficiency. A significant difference existed in the 2008 and 2009 MAP overall population and each subgroup at the elementary level in communication arts and mathematics. Conversely, at the middle school level, no significant difference existed in the 2008 and 2009 MAP overall population and each subgroup , with the exception of the mathematics subgroups Asian/Pacific Islander, Hispanic, Free and Reduced Lunch, and Individualized Education Program,

This study will not provide evidence that Study Island was the sole factor that effected student achievement. However, when reviewing the amount of time spent on

Study Island and the number of questions answered by the schools in this study, evidence exists that the use of Study Island represented a significant change in the practice of teachers as well as opportunities for students when comparing 2007-2008 data to 2008-2009 data.

Table of Contents

Acknowledgements i

List of Tables xii

Chapter One: Introduction 1

Background of the Study 5

Significance of the Study 7

Problem Statement 9

Independent variable 9

Dependent variable. 10

Hypotheses 10

Null hypothesis #1 10

Null hypothesis #2 10

Null hypothesis #3 10

Null hypothesis #4 10

Alternative hypothesis #1. 10

Alternative hypothesis #2 11

Alternative hypothesis #3 11

Alternative hypothesis #4. 11

Definition of Terms 11

Adequate yearly progress 11

Formative assessment 11

Missouri Assessment Program 11
No Child Left Behind Act 11
Study Island. 12
Limitations 12
Sample demographics 12
Timeframe of study. 12
Study Island usage 12
Missouri Assessment Program 12
Data analysis 13
Subgroups. 13
Other variables 13
Conclusion 13
Chapter Two: Review of the Literature 15
Background 15
Study Island 17
School Improvement 20
Effective schools movement. 20
What works in schools. 22
Frequent Monitoring of Student Progress 23
Goal setting. 24

Feedback. 25

Opportunity to Learn and Time on Task.. 27

Positive Home-School Relationships.. 28

Assessment Research 29

Assessment and student achievement. 31

Formative assessment generalizations and practices. 33

Formative assessment and classroom instruction. 35

Formative assessment and student motivation.. 36

Formative assessment and feedback. 37

Formative assessment and alignment.. 38

Formative assessment and technology.. 39

No Child Left Behind... 40

Accountability for results... 40

Research-based actions. ....**Error! Bookmark not defined.**

Expanded parental options. 42

Expanded local control and flexibility.. 43

Missouri Annual Performance Report 43

Student Achievement 44

Quality and limitations of state tests. 46

Achievement gap trends... 48

Achievement levels. 50

Scale scores, standard deviations, and effect sizes. 53

Conclusion 54

Chapter Three: Methodology 57

Research Overview 57

Research Design and Perspective 57

Instrumentation 58

Missouri Assessment Program reliability and validity. 59

Population and Sampling Procedures 64

Determining the eligible sample population. 64

Determining the actual sample population. 65

Research Design and Procedure 65

Preparing the data for analysis. 67

Data Analysis 67

Conclusion 69

Chapter Four: Results 70

Overview 70

Data Analysis 71

Elementary communication arts. 71

School total 72

Asian/Pacific Islander 74
Black 76
Hispanic 78
White 80
Free and reduced lunch 82
Individualized education program 84
Limited English proficiency 86
Elementary mathematics. 88
School total 89
Asian/Pacific Islander 91
Black 93
Hispanic 95
White 97
Free and reduced lunch 99
Individualized education program 101
Limited English proficiency 103
Middle school communication arts. 104
School total 106
Asian/Pacific Islander 108
Black 110

Hispanic 112

White 114

Free and reduced lunch 116

Individualized education program 118

Limited English proficiency 120

Middle school mathematics. 121

School total 123

Asian/Pacific Islander 125

Black 127

Hispanic 129

White 131

Free and reduced lunch 133

Individualized education program 135

Limited English proficiency. 137

Study Island Data 139

Number of sessions and time spent using Study Island. 139

Study Island questions attempted and percent correct. 143

District Achievement Data 146

Conclusion 148

Chapter Five: Implications and Recommendations 151

Overview ... 151

Data Analysis ... 152

Elementary communication arts and mathematics. ... 152

Middle school communication arts and mathematics. ... 153

Implications ... 154

Curriculum. ... 155

Instruction. ... 156

Assessment. ... 158

Professional development. ... 160

Study Island usage. ... 161

Subgroups ... 163

Recommendations ... 163

Extending the timeframe of the study. ... 164

Case study and qualitative components. ... 164

Control group. ... 165

Different statistical measures. ... 166

Disaggregate Study Island usage data. ... 166

Conclusion ... 167

References ... 171

Appendix A ... 182

Appendix B 184

Vitae 185

List of Tables

Table 1: School District Enrollment and Demographics 2009 6

Table 2: Findings on the Effects of Different Types of Feedback................................... 27

Table 3: Gain Associated with Number of Assessments Over 15 Weeks 36

Table 4: Missouri's Adequate Yearly Progress Annual Performance Targets 41

Table 5: Missouri Assessment Program Grade-Span Assessment Timeline 60

Table 6: Missouri Assessment Program Grade-Level Assessment Timeline 60

Table 7: 2009 Comm. Arts Missouri Assessment Program Grade-Level Assessments ... 63

Table 8: 2009 Math Missouri Assessment Program Grade-Level Assessments 63

Table 9: Descriptive Statistics for Elementary Communication Arts School Total 73

Table 10: Quantitative Analysis for Elementary Communication Arts School Total 73

Table 11: Descriptive Statistics for Elementary Comm. Arts Asian/Pacific Islander 75

Table 12: Quantitative Analysis for Elementary Comm. Arts Asian/Pacific Islander 75

Table 13: Descriptive Statistics for Elementary Communication Arts Black 77

Table 14: Quantitative Analysis for Elementary Communication Arts Black 77

Table 15: Descriptive Statistics for Elementary Communication Arts Hispanic 79

Table 16: Quantitative Analysis for Elementary Communication Arts Hispanic............. 79

Table 17: Descriptive Statistics for Elementary Communication Arts White.................. 81

Table 18: Quantitative Analysis for Elementary Communication Arts White 81

Table 19: Descriptive Statistics for Elementary Comm. Arts Free and Reduced Lunch . 83

Table 20: Quantitative Analysis for Elementary Comm. Arts Free and Reduced Lunch. 83

Table 21:Descriptive Statistics for Elementary Comm. Arts IEP.................................... 85

Table 22: Quantitative Analysis for Elementary Comm. Arts IEP.................................. 85

Table 23: Descriptive Statistics for Elementary Comm. Arts LEP 87

Table 24: Quantitative Analysis for Elementary Comm. Arts LEP................................. 87

Table 25: Descriptive Statistics for Elementary Mathematics School Total 90

Table 26: Quantitative Analysis for Elementary Mathematics School Total 90

Table 27: Descriptive Statistics for Elementary Mathematics Asian/Pacific Islander 92

Table 28: Quantitative Analysis for Elementary Mathematics Asian/Pacific Islander 92

Table 29: Descriptive Statistics for Elementary Mathematics Black 94

Table 30: Quantitative Analysis for Elementary Mathematics Black 94

Table 31: Descriptive Statistics for Elementary Mathematics Hispanic 96

Table 32: Quantitative Analysis for Elementary Mathematics Hispanic.......................... 96

Table 33: Descriptive Statistics for Elementary Mathematics White.............................. 98

Table 34: Quantitative Analysis for Elementary Mathematics White 98

Table 35: Descriptive Statistics for Elementary Mathematics Free & Reduced Lunch . 100

Table 36: Quantitative Analysis for Elementary Mathematics Free & Reduced Lunch 100

Table 37: Descriptive Statistics for Elementary Mathematics IEP................................ 102

Table 38: Quantitative Analysis for Elementary Mathematics IEP............................... 102

Table 39: Descriptive Statistics for Elementary Mathematics LEP............................... 104

Table 40: Quantitative Analysis for Elementary Mathematics LEP.............................. 104

Table 41: Descriptive Statistics for Middle School Comm. Arts School Total............. 107

Table 42: Quantitative Analysis for Middle School Comm. Arts School Total............ 107

Table 43: Descriptive Statistics for Middle School Comm. Arts Asian/Pacific Islander109

Table 44: Quan. Analysis for Middle School Comm. Arts Asian/Pacific Islander 109

Table 45: Descriptive Statistics for Middle School Communication Arts Black 111

Table 46: Quantitative Analysis for Middle School Communication Arts Black 111

Table 47: Descriptive Statistics for Middle School Communication Arts Hispanic 113

Table 48: Quantitative Analysis for Middle School Communication Arts Hispanic 113

Table 49: Descriptive Statistics for Middle School Communication Arts White........... 115

Table 50: Quantitative Analysis for Middle School Communication Arts White.......... 115

Table 51: Descriptive Stats for Middle School Comm. Arts Free and Reduced Lunch. 117

Table 52: Quan. Analysis for Middle School Comm. Arts Free and Reduced Lunch ... 117

Table 53: Descriptive Statistics for Middle School Comm. Arts IEP 119

Table 54: Quantitative Analysis for Middle School Comm. Arts IEP 119

Table 55: Descriptive Statistics for Middle School Comm. Arts LEP 121

Table 56: Quantitative Analysis for Middle School Comm. Arts LEP 121

Table 57: Descriptive Statistics for Middle School Mathematics School Total............. 124

Table 58: Quantitative Analysis for Middle School Mathematics School Total............ 124

Table 59: Descriptive Stats for Middle School Mathematics Asian/Pacific Islander..... 126

Table 60: Quan. Analysis for Middle School Mathematics Asian/Pacific Islander 126

Table 61: Descriptive Statistics for Middle School Mathematics Black 128

Table 62: Quantitative Analysis for Middle School Mathematics Black 128

Table 63: Descriptive Statistics for Middle School Mathematics Hispanic 130

Table 64: Quantitative Analysis for Middle School Mathematics Hispanic 130

Table 65: Descriptive Statistics for Middle School Mathematics White........................ 132

Table 66: Quantitative Analysis for Middle School Mathematics White....................... 132

Table 67: Descriptive Stats for Middle School Mathematics Free & Reduced Lunch... 134

Table 68: Quan. Analysis for Middle School Mathematics Free & Reduced Lunch 134

Table 69: Descriptive Statistics for Middle School Mathematics IEP 136

Table 70: Quantitative Analysis for Middle School Mathematics IEP........................... 136

Table 71: Descriptive Statistics for Middle School Mathematics LEP 138

Table 72: Quantitative Analysis for Middle School Mathematics LEP 138

Table 73: District Comm. Arts Number of Sessions & Time Spent Using Study Island 140

Table 74: District Math Number of Sessions & Time Spent Using Study Island 141

Table 75: February Study Island Total Number of Sessions and Time Spent................ 142

Table 76: District Comm. Arts Study Island Questions Correct, Total, and Percent 144

Table 77: District Mathematics Study Island Questions Correct, Total, and Percent 144

Table 78: February Study Island Questions Correct, Total, and Percent........................ 145

Table 79: District Communication Arts Adequate Yearly Progress Percent Proficient. 147

Table 80: District Mathematics Adequate Yearly Progress Percent Proficient.............. 148

Chapter One: Introduction

Programs designed to raise student achievement are common in most school districts. However, the effectiveness of these programs can be difficult to measure. The difficulty arises because some programs help all students while others target specific students. Regardless of who uses a particular program, educators often rely on the results of state tests, as well as other measures, to determine if the program effected student achievement. This study was designed to answer whether Study Island had a significant effect on the achievement, as measured by state test scores, of all students.

The answer to that question is important for several reasons, one of which is cost. A program such as Study Island requires districts to purchase the right to access it on an annual basis. Therefore, district leaders need to be confident that the program is delivering the desired results. Across the nation, school district budgets are decreasing, while student achievement expectations are increasing. For example, each year since 2008, the district in this study has been forced to cut millions of dollars from its budget while state test score expectations have continued to rise. The district in this study began using Study Island to increase student achievement on state tests, which is the primary indicator used by the federal government to determine the effectiveness of public schools and districts. The schools used for this study were in the state of Missouri and in 2009, 278 out of 553 of all Missouri school districts and charter schools were considered in need of improvement as a result of not meeting Adequate Yearly Progress for two or more consecutive years in either communication arts, mathematics, or both (MO DESE, 2009b). Therefore, in just the state of Missouri the need for improved student achievement is paramount.

Increasing scores on state tests and meeting No Child Left Behind annual performance targets are challenges that all public schools have faced since 2002 when Congress passed the No Child Left Behind Act. The enactment of No Child Left Behind reauthorized the Elementary and Secondary Education Act of 1965, the federal law effecting education from kindergarten through high school. This reauthorization set the goal that every child would meet grade-level, state-defined educational standards in communication arts and mathematics by 2014. To measure whether students were meeting those grade-level standards, No Child Left Behind required authorities in each state to develop tests. States administer these tests on an annual basis to determine whether student achievement has increased (United States Department of Education [DOE], 2004).

Government officials created No Child Left Behind to raise overall student achievement and close achievement gaps. In education, an achievement gap refers to the disparity in academic performance between groups of students. Educators use it to describe the performance gaps between specific ethnic groups, such as African-Americans and Hispanics, and their white peers. It can also refer to the academic disparity between students from low-income and high-income households (Education Week, 2010).

The legislators that crafted No Child Left Behind mandated that the achievement of all groups of students be measured and reported, as opposed to just measuring and reporting the overall average achievement, which was how achievement had been reported in the past. Each group of students, separate from the overall population, is referred to as a subgroup which is a specific group of students defined by their ethnicity, race, socioeconomic status, limited English proficiency, or special education needs. Under No Child Left Behind, if a school or district has 30 or more students in a particular subgroup, it is accountable for the achievement of that subgroup (DOE, 2004). The district in this study is accountable for the overall average achievement, School Total, as well as the achievement of seven subgroups in both communication arts and mathematics, including Asian/Pacific Islander, Black, Hispanic, White, Free and Reduced Lunch, Individualized Education Program, and Limited English Proficiency.

Under No Child Left Behind, school leaders must ensure that the total school population and all subgroups meet grade-level standards in order to make Adequate Yearly Progress. Adequate Yearly Progress is the measure used by state officials under No Child Left Behind to determine whether a school has met annual performance targets (DOE, 2010), which are the percentages of students that must meet grade-level standards on the state tests in communication arts and mathematics for that particular school year. Educators refer to the percentage of students that meet grade-level standards as the percent proficient. Under No Child Left Behind, schools that do not meet Adequate Yearly Progress for two consecutive years face possible sanctions.

In addition to Adequate Yearly Progress, the Missouri Department of Elementary and Secondary Education issues each district an Annual Performance Report. State officials use the Annual Performance Report, which consists of 14 indicators, to determine the accreditation of a district. Although some of the indicators used on the Annual Performance Report are the same as those used for Adequate Yearly Progress, it is separate from No Child Left Behind. Missouri's Annual Performance Report developers chose to make 8 of its 14 indicators the same as those found in No Child Left

Behind. These 8 common indicators encompass the results of the Missouri Assessment Program (MAP) state test results as well as attendance and graduation rates. One of the Annual Performance Report indicators that differs from No Child Left Behind involves comparing the change in minority achievement in a district with the change in state majority achievement to measure if the achievement gap has been closed (Missouri Department of Elementary and Secondary Education [MO DESE], 2009e).

As might be expected, the shift to accountability for all student groups has made it difficult for schools to meet Adequate Yearly Progress. Consequently, school leaders have been looking for programs that will help raise overall achievement as well as the achievement of their subgroups. Since so much of the accountability under No Child Left Behind hinges on the results of state tests, these programs must positively effect achievement on state tests. Study Island, the program used by the district in this study to help meet Adequate Yearly Progress, is a program that claims to raise achievement on state tests.

Study Island is a web-based assessment program that students, teachers, and parents in this study were able to use both at school and at home by accessing it through the Internet. The Study Island website, www.studyisland.com, enables students to login and practice answering questions aligned to grade-level standards. By answering questions correctly, students earn opportunities to play games.

Study Island uses multiple-choice questions to determine what students know, as do most state tests, and developers aligned questions to grade-level standards. Therefore, the Study Island questions are similar to what students will encounter on state tests. Similarly, when students use Study Island they receive immediate feedback and are engaging in practice that should help them score better on the state tests. Study Island has the ability to inform educators about student achievement well in advance of the state tests.

The district in this study began using Study Island during the 2008-2009 school year for the specific purpose of raising achievement on state tests. This study will show whether Study Island was able to significantly effect the achievement of subgroups, as well as the overall population, on the state tests. Educators at the district and school level need to know whether Study Island helped them reach their goals so they can make an informed decision on whether they should continue to use it. Currently, there are not a wealth of independent studies available that have researched the effectiveness of Study Island on student achievement especially, when it comes to longitudinal studies.

Background of the Study

This study was conducted in a large district in Missouri, and included results from the 10 elementary and 5 middle schools in the district. All of the elementary schools in the study serve kindergarten through fifth grade, while all of the middle schools serve sixth through eighth grade. Table 1 provides enrollment and demographic information for the district in this study. Appendix A contains enrollment and demographic information on each individual school in the study.

The impact of Study Island on student achievement, as measured by the MAP, was determined in this study. State officials with the Department of Elementary and Secondary Education created the MAP to measure the Adequate Yearly Progress, as outlined in No Child Left Behind, of each school. Since MAP data is available for students in grades 3-8 in communication arts and mathematics, it served as the student achievement measure for analysis in this study. For the district in this study, students in grades 3-8 used Study Island during 2008-2009, both at school and at home, through the Internet. However, prior to 2008-2009, Study Island was not available for students and teachers in the district.

Table 1
School District Enrollment and Demographics 2009

Population	Number	Percent
Asian	464	2.6
Black	1,083	6.1
Hispanic	390	2.2
Indian	36	0.2
White	15,659	88.8
Total	17,632	100.0

Note. Adapted from Missouri Department of Elementary and Secondary Education School Profile Student Demographics. Retrieved February 22, 2010, from http://dese.mo.gov/planning/profile/.

All of the elementary and middle school teachers and students had access to Study Island during the 2008-2009 school year. The principal of each school developed a Study Island usage plan and set expectations with teachers. At each school in the study, teachers selected and assigned students to sessions aligned to Missouri Grade Level Expectations in communication arts and mathematics. Each session consisted of students answering a set of questions aligned to a particular standard. The teacher determined the number of questions per session and had the ability to vary the number of questions from one session to the next.

In addition to assigning students sessions, teachers also set goals for students regarding the percent of questions they should answer correctly within each session. For motivation purposes, teachers rewarded students with incentives and recognition, such as blue ribbons, for meeting the session goal. In some instances, teachers also set entire classroom goals, and some administrators set overall school goals. These goals encouraged students of all ability levels to increase their achievement and scaffold overall achievement to a level that would satisfy Adequate Yearly Progress.

Administrators, teachers, and parents monitored student progress by tracking the number of questions answered correctly and which grade-level standards these questions addressed. Similarly, administrators were able to access school-wide results as well as the results of individual classrooms and students. Likewise, teachers could view their overall classroom results and those of the individual students. In addition, parents could view their child's results, and an individual child could see his or her own results.

Significance of the Study

The results of this study will increase the understanding of educators and quantify the effect of Study Island on student achievement. Furthermore, the quantitative analysis of achievement data provides teachers and administrators evidence from which to base future decisions related to the use of Study Island. Moreover, if Study Island can increase the achievement of subgroup populations, it could help districts in Missouri meet Adequate Yearly Progress and Annual Performance Report achievement indicator.

In addition to the quantitative results identified in this study, Study Island also supported the Missouri School Improvement Program. The Missouri State Board of Education and the Department of Elementary and Secondary Education created the Missouri School Improvement Program to promote school improvement within each district and across the state. In fact, Study Island could address several items outlined in the Fourth Cycle Missouri School Improvement Program Standards and Indicators Manual (2006). State officials created the manual to guide school districts through the improvement process. The developers of the Instructional Design and Practices section of the manual suggested schools use a variety of assessment data both longitudinally and disaggregated by demographics (MO DESE, 2006). Ultimately, teachers must use assessment information to plan instruction for students. Not surprisingly, the district in this study chose Study Island because it supports the assessment data recommendations in the manual.

This study will not provide evidence that Study Island was the sole factor that effected student achievement. However, when reviewing 2008-2009 and 2009-2010 Study Island usage data, such as the amount of time spent on Study Island and the number of questions answered by the schools in this study, evidence exists that the use of Study Island represented a significant change in the practice of teachers as well as opportunities for students. Certainly, for most teachers and students, the use of Study Island was a greater change in practice and opportunity from one year to the next than the other possible variables.

Problem Statement

This quantitative study investigated whether the use of Study Island significantly effected overall and subgroup achievement. To that end, statistical tests determined whether a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores in communication arts and mathematics at the elementary and middle school levels. Results from the 2010 MAP tests became available prior to the submission of this study and therefore, the overall data was included in chapter 4 however, no statistical tests were performed since they were not part of the approved study parameters. Fortunately, MAP data works well for this study because each time a student completes a state test, they receive a scale score and an achievement level. The scale scores range in value from 450 to 910 across grade-levels and content areas. The achievement level cut points were identified and applied to the scale scores for each grade-level and content area so that a designation of Below Basic, Basic, Proficient, or Advanced could be applied to a student along with his or her scale score (MO DESE, 2009a).

Two sources, the Missouri Department of Elementary and Secondary Education and Study Island, provided access to the data necessary for this study. Annually, the Missouri Department of Elementary and Secondary Education officials provide each district with aggregated and disaggregated results of the annual MAP tests. Conveniently, Study Island personnel provide member districts with a database that reflects information specific to that district and the individual schools. Password-protected logins allowed district officials to access both of these databases.

Independent variable. The independent variable in this study was the use of Study Island. During 2007-2008, students in this district did not use Study Island; however, during 2008-2009 Study Island was used by students in all elementary and middle schools.

Dependent variable. The dependent variable in this study was student achievement in communication arts and mathematics at the elementary and middle school levels as measured by the MAP scale scores from 2008 and 2009.

Hypotheses

Null hypothesis #1. There is no significant difference between the 2008 MAP average communication arts scale scores and the 2009 MAP average communication arts scale scores at the elementary level.

Null hypothesis #2. There is no significant difference between the 2008 MAP average mathematics scale scores and the 2009 MAP average mathematics scale scores at the elementary level.

Null hypothesis #3. There is no significant difference between the 2008 MAP average communication arts scale scores and the 2009 MAP average communication arts scale scores at the middle school level.

Null hypothesis #4. There is no significant difference between the 2008 MAP average mathematics scale scores and the 2009 MAP average mathematics scale scores at the middle school level.

Alternative hypothesis #1. There is a significant difference between the 2008 MAP average communication arts scale scores and the 2009 MAP average communication arts scale scores at the elementary level.

Alternative hypothesis #2. There is a significant difference between the 2008 MAP average mathematics scale scores and the 2009 MAP average mathematics scale scores at the elementary level.

Alternative hypothesis #3. There is a significant difference between the 2008 MAP average communication arts scale scores and the 2009 MAP average communication arts scale scores at the middle school level.

Alternative hypothesis #4. There is a significant difference between the 2008 MAP average mathematics scale scores and the 2009 MAP average mathematics scale scores at the middle school level.

Definition of Terms

Adequate yearly progress. Adequate Yearly Progress is the minimum levels of improvement, the percentage of students scoring at grade-level, which schools and districts must achieve within periods specified in the law (DOE, 2004).

Formative assessment. Formative assessment is all the activities undertaken by students and teachers that provide information to be used as feedback, which leads to modified teaching and learning based on student needs (Black & Wiliam, 1998a).

Missouri Assessment Program. The Missouri Department of Elementary and Secondary Education assessments in communication arts and mathematics at the elementary, middle, and high school levels for determining Adequate Yearly Progress and state accreditation (MO DESE, 2009a).

No Child Left Behind Act. The No Child Left Behind Act reauthorized the Elementary and Secondary Education Act of 1965, which effected education from kindergarten through high school. The act required schools to help all students meet state defined learning goals by 2014 (DOE, 2004).

Study Island. Study Island is a web-based assessment program that provides access to multiple-choice assessment items aligned to state standards. The program provides users with the opportunity to play games when questions are answered correctly. It also allows teachers, administrators, and parents access to results (Study Island, 2010).

Limitations

Sample demographics. The study was completed in one district located in Missouri. Therefore, the results may be biased due to specific circumstances within that school community. The results may not be accurate when applied to other districts with different demographics.

Timeframe of study. This study only contained one year of data during which Study Island was used. In order to determine the impact of a program at least three to five years of data are necessary. However, with the increased demands of NCLB, programs must demonstrate almost immediate impact if they are to receive funding the next year.

Study Island usage. 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.

Missouri Assessment Program. The only achievement data used in the study came from two years of MAP testing. Other measures of achievement would help to triangulate data and further support conclusions.

Data analysis. This study used *z* tests and *t* tests with a 95% confidence interval to determine whether Study Island had a significant impact on student achievement. There are other data analysis methods that can be used to measure achievement.

Subgroups. The district in this study was accountable in seven different subgroups. The number of students that comprised each of the subgroups varied. Some of the subgroups had over 1,000 students while others had around 30. Random samples were taken from each subgroup and analyzed using a *z* test or *t* test.

Other variables. The study does not attempt to measure other variables that can effect student achievement such as curriculum, instruction, professional development, parent involvement, or leadership.

Conclusion

Educators know that many different variables effect student achievement on state tests. The design of this research project allows school officials to determine the effect of Study Island on student achievement. In fact, the study determined whether there was a statistically significant difference between the 2008 and 2009 MAP scale scores in communication arts and mathematics at the elementary and middle school levels. The null hypotheses stated that no significant difference existed between the 2008 and 2009 MAP scale scores in communication arts and mathematics at the elementary and middle school levels. To gain further insight, the MAP scale scores were disaggregated by subject, level, and subgroup.

Across the nation, increasing student achievement on state tests is an important issue for all public schools. The importance goes beyond making Adequate Yearly Progress; rather, it serves to cement the existence and value of public education. In

Missouri, increasing student achievement on state tests can help schools and districts meet federal and state achievement targets and indicators. When school officials provide teachers and students with a program that increases student achievement, it supports their quest to meet achievement goals. School officials can use the results of this study to help determine whether Study Island should continue to be used by teachers and students.

Chapter 2 contains the review of related literature and research related to the issues investigated in this study. The literature review discusses achievement studies performed on Study Island and state tests as well as research-based school improvement areas related to this study. In addition, chapter 2 summarizes the impact of formative assessment and No Child Left Behind on student achievement. Furthermore, to help the reader understand why schools are under pressure to meet Adequate Yearly Progress and why they might use a program like Study Island, an overview provides No Child Left Behind accountability specifics.

Chapter Two: Review of the Literature

Background

Lawmakers at the federal level designed No Child Left Behind to improve student achievement and close achievement gaps. President George W. Bush signed the law in January of 2002 and stated, "Too many of our neediest children are being left behind" (DOE, 2004, p. 1). Four pillars form the basis of No Child Left Behind: accountability for results, an emphasis on doing what works based on scientific research, expanded parental options, and expanded local control and flexibility (DOE, 2004).

No Child Left Behind set the goal of every child making the grade on state-defined education standards by the end of the 2013-2014 school year. "Making the grade" means meeting the grade-level curricular standards outlined by the state. In response to No Child Left Behind, education officials in each state developed their own assessment system to measure student achievement and their progress towards this goal on an annual basis. Since 2002, each state has reported the percentage of students making the grade, often referred to as scoring *Proficient*, in communication arts and mathematics to the federal government for determining Adequate Yearly Progress (DOE, 2004).

In Missouri, as is the case in many other states, the state tests used to determine Adequate Yearly Progress for No Child Left Behind are summative assessments of student achievement. Some common examples of summative assessments are state tests, unit tests, chapter tests, final exams, scores used for school accountability, and report cards. Summative assessments are referred to as "assessments *of* learning" and reflect the use of test data to monitor the progress of students and schools (Stiggins & Chappuis, 2008; Stiggins, Arter, Chappuis & Chappuis, 2005). To make an analogy, a summative assessment, or a standardized test, is like an autopsy. When an autopsy occurs, the person is already dead; the only thing left to do is determine what killed them. Likewise, when a summative assessment occurs, the instruction is over, and the only thing left to do is figure out what the student did or did not learn.

Conversely, educators and researchers have discovered another way to measure student achievement, formative assessment. Formative assessments measure achievement several times prior to a summative assessment, while there is still time for further instruction to occur. Furthermore, formative assessment activities may occur inside or outside the classroom and include more locally created tasks that check for immediate understanding such as teacher-made tests, curriculum-embedded tests, exit slips, oral questioning, and/or a variety of other performance activities. Rick Stiggins (2001, 2004, 2007) used the terminology "assessment *for* learning" to reflect the use of assessment for acquiring useful data to inform instructional practice. To continue the analogy, a formative assessment is more akin to a check-up than an autopsy. The purpose of a check-up is to diagnose a problem and treat it. Such is the case with formative assessment - it diagnoses what a student is having trouble learning, and allows time for re-teaching of those concepts to occur prior to a summative assessment.

As discussed in chapter 1, No Child Left Behind has an accountability component that requires schools to improve student achievement on an annual basis. Many schools successful in improving student achievement have turned to research and school improvement models that call for the use of best practices. Moreover, many of the research-based best practices in education called for the balanced use of formative and summative assessments. In addition, they advocated for the use of practices associated with the Correlates of Effective Schools (Lezotte, 1991), as well as the school-level, teacher-level, and student-level factors Robert Marzano (2003) identified through his research.

Student achievement can be both the easiest thing to measure and the most difficult. Schools are required to administer summative assessments for state and federal reporting purposes, and measuring student achievement for these purposes is relatively simple. However, knowing and understanding the antecedents that factored into the summative assessment results are much more complex. An endless number of variables can influence the achievement of students on a summative assessment. Due to the increased accountability placed on public schools by No Child Left Behind, school officials are investing money in programs that help them control the variables that effect student achievement on state tests. Study Island is one of those programs.

Study Island

Entrepreneurs Cameron Chalmers, a computer scientist, and David Muzzo, an economist and marketer, founded Study Island in 2000. The founders wanted to develop an online educational program that would engage learners of all ability levels. Chalmers and Muzzo developed Study Island in concert with the Department of Education in Ohio, where educators were revising their state standards, and consequently, launched Study Island in Ohio late in 2000. Today, teachers and students in nearly half of schools in Ohio use Study Island (Archipelago Learning, 2010).

Since launching Study Island, the founders have continued to expand and refine the product. Writers at *Inc.* magazine recognized Study Island as one of the top small businesses of 2006, and by the end of the 2008-2009 school year, 19,200 schools and 8.4

million students across all 50 states and the District of Columbia were using Study Island. In addition, Study Island earned a designation in *District Administration's* Readers' Choice Top 100 Products for 2008, won *Business Week's* prestigious STEVIE AWARD in 2009 for best customer service by a software company, and was named by *Dallas Business Journal* as one of the best places to work. Due to the success of Study Island, the founders acquired other online educational products, such as Education City, and continued to develop other products, such as Northstar Learning. In early 2009, Chalmers and Muzzo brought all of their products under one umbrella by introducing a new corporate name and brand identity, Archipelago Learning (Archipelago Learning, 2010).

Chalmers and Muzzo contracted Magnolia Consulting, an independent consulting firm specializing in educational evaluation, to summarize the results of pre-existing case studies and provide evidence of the effect of Study Island on student achievement. In fact, they asked the researchers at Magnolia Consulting to add new data and analyses to illustrate the effectiveness of their product. In October 2008, Magnolia Consulting published a compilation of the case studies they performed using data on Study Island. They conducted a retrospective study using quasi-experimental frameworks that varied by study. Although the case studies varied, each one included at least one of the following:

- Comparisons of student achievement before and after Study Island use.
- Comparisons between schools using Study Island and local and/or state norms.
- Comparisons of changes in proficiency between schools using Study Island and other schools in the district or region not using Study Island.

- Trends in growth of student achievement over time at Study Island schools (Watts, 2008).

Specifically, the compilation of case studies, which included case studies from 11 different states, addressed the following overarching evaluation questions for each case study:

> 1) Was there significant growth in student achievement over time after the students have used Study Island? 2) Was there a significant difference in student achievement between schools using Study Island and schools not using Study Island? (Watts, 2008, p. 2).

The report stated that Study Island effected student achievement in a positive manner, particularly in math. However, the author pointed out that the nature of the case studies precludes one from making causal claims that Study Island was the sole factor that effected student achievement (Watts, 2008).

The researchers at Magnolia Consulting used metrics such as percent proficient, score growth, average percent proficient, and change in percent proficient to illustrate the effectiveness of Study Island. Unfortunately, the study only involved one-year comparisons; thus, it did not include any longitudinal data. In fact, the author of the Magnolia Consulting case study likely ran into the same limitations that other studies, such as the Center on Education Policy, encountered when analyzing No Child Left Behind student achievement data. These limitations included the absences of data past percentage proficient and breaks in the data. Magnolia Consulting and the Center on Education Policy (2007) focused on state test scores and analyzed the percentage scoring proficient with limited access to mean scale scores and standard deviations. The Center on Education Policy study results are discussed in the second to last section, Student Achievement, of this chapter.

School Improvement

A school in need of improvement has a considerable amount of research to draw upon to help shape administrator and teacher actions. This section features some of the most prominent research, which has synthesized and outlined where school officials should focus their improvement actions. Over the past 50 years, researchers such as James Coleman et al. (1966) and Christopher Jencks have questioned the impact of schools. Likewise, officials with the federal government published reports aimed at igniting educational efforts in the United States. For example, in 1983, James J. Harvey, author of *A Nation at Risk*, claimed that if the United States felt that its current state of educational mediocrity had been imposed by another nation, it would be viewed as an act of war. If, in fact, educational mediocrity is a war, the following section would provide strategies to help end it.

Effective schools movement. According to *Equity in Educational Opportunity*, a report published by James Coleman and associates in July of 1966, the main determining factor of student achievement is family background, not schools. The authors went on to say, regardless of the quality of the instruction, that factors such as poverty and the lack of education of the parents could not be overcome. Furthermore, they stated that schools account for only about 10% of the variance in student achievement; the other 90% is accounted for by the background characteristics of the student. Simply put, they asserted that schools do not make a difference in predicting a student's achievement (Lezotte, 2001).

Christopher Jencks and his colleagues published another report, *Inequity: A Reassessment of the Effects of Family and Schooling in America*, in 1972. Their findings supported those made by the Coleman report and go on to assert the following:

- Schools do little to lessen the gap between rich students and poor students.
- Schools do little to lessen the gap between more and less able students.
- Student achievement is primarily a function of one factor – the background of the student.
- Little evidence exists that education reform can improve a school's influence on student achievement.

The reports by Coleman et al. (1966) and Jencks et al. (1972) dealt a blow to society's optimism about the effectiveness of public education. Regrettably, they strongly suggested that any reform initiatives designed to overcome or reverse the inevitable outcome of demographic barriers were bound to fail.

The assertion that schools do not matter drew the attention of several researchers, including Ron Edmonds (1982) and Lawrence Lezotte (1991) who believed schools do matter. They set out to find and study examples of schools that disproved the claims of Coleman et al. (1966) and Jencks et al. (1972). The research of these three men, known as the Effective Schools Movement, formed the basis of the Correlates of Effective Schools.

In 1982, Edmonds was the first to identify the Correlates of Effective Schools. The correlates stemmed from similarities found among schools that exhibited evidence of high student achievement despite barriers such as poverty and lack of parental education.

Lezotte updated and refined the original correlates and current literature lists them as the following (Lezotte, 1991):

- Instructional leadership
- Clear and focused mission
- Safe and orderly environment
- Climate of high expectations
- Frequent monitoring of student progress
- Positive home-school relationships
- Opportunity to learn and student time on task

In the remainder of this section, the author will focus on three of the seven correlates, frequent monitoring of student progress, opportunity to learn and student time on task, and positive home-school relationships. The researcher chose to focus on these three correlates because Study Island has features that directly support these areas.

What works in schools. In the book *What Works in Schools: Translating Research into Action,* Robert Marzano (2003) organized the results of 35 years of research into three general factors that influence student achievement: school-level factors, teacher-level factors, and student-level factors. Each of the three level factors contained major components outlined below.

School-Level Factors

- A guaranteed and viable curriculum
- Challenging goals and effective feedback
- Parent and community involvement
- Safe and orderly environment

- Collegiality and professionalism

Teacher-Level Factors

- Instructional strategies
- Classroom management
- Classroom curriculum design

Student-Level Factors

- Home environment
- Learned intelligence and background knowledge
- Student motivation

As expected, there is considerable overlap when comparing the factors that Marzano (2003) outlined in his research and those of the correlates of effective schools. Study Island supported the three correlates emphasized in this chapter, frequent monitoring of student progress, opportunity to learn and student time on task, and positive home-school relationships. In addition, these three correlates connect with the school, teacher, and student-level factors Marzano outlined. This lends even more credibility to the assertion that these areas contain the leverage necessary to improve student achievement when schools make them a priority.

Frequent Monitoring of Student Progress

In his paper, *Correlates of Effective Schools: The First and Second Generation*, Lezotte (1991) stated that after what he termed as the "first generation" of frequent monitoring of student progress is accomplished, schools need to advance into the "second generation" of frequent monitoring of student progress. Lezotte explained that during the second generation the use of technology would permit teachers to do a better job of monitoring the progress of students. This same technology will allow students to monitor their own learning and make adjustments. Moreover, the use of technology to administer formative assessments, provide immediate feedback, and display correct solutions are a few of the available tools for assuring student learning (Lezotte, 1991).

Study Island fits into the second generation of monitoring that Lezotte described. Study Island allows administrators, teachers, and parents to monitor student learning. Additionally, it allows students to see whether they are progressing and provides incentives for them to achieve higher levels of learning. Students receive immediate feedback on the questions they attempt. In fact, this same information is immediately available to teachers and parents. Traditional tasks that teachers perform, such as grading homework and tests, disappear and therefore, students and teachers receive the results immediately.

Goal setting. Goal setting supports frequent monitoring of student progress, and becomes even more powerful when the goals are specific and measurable. Students need to personalize goals to fit their needs, therefore they become more likely to be achieved (Marzano, Pickering, & Pollock, 2001). In addition, research supports students actively tracking their own performance and monitoring their own progress (Marzano, 2003; Trammel, Schloss, & Alper, 1994). Students need to monitor the goals they set on a frequent basis to evaluate their progress. Fortunately, Study Island allows students to see their progress on specific standards. Similarly, teachers have access to this information making it easy for students and teachers to set and monitor goals aligned to these standards.

It is beneficial for an entire classroom, building, and district to set goals. Mike Schmoker (1999) noted that setting academic goals for the school as a whole has a powerful coalescing effect on teachers and administrators: "Goals themselves lead not only to success but also to the effectiveness and cohesion of a team" (p. 24). Study Island allows stakeholders to track data at the classroom, building, and district level. The district in this study monitored data on a monthly basis by analyzing reports at each level.

Feedback. Feedback is another area that connects to frequent monitoring of student progress. Feedback is information that provides learners with an understanding of how they are doing, have done, or might do in the future to enhance what their knowledge and achievement (Callingham, 2008; Crowie, 2005). In addition to this definition, a myriad of literature exists on feedback related to its impact on student achievement (e.g., Brookhart, 2008; Crooks, 1998; Kulhavy, 1977; Mory, 2004; Shute, 2008). Study Island incorporates elements of effective feedback in terms of providing immediate and specific feedback to students.

Researcher James Hattie (1992) reviewed close to 8,000 studies and concluded that the most powerful single modification that enhances achievement is feedback. Hattie went on to say that, the simplest prescription for improving education must be dollops of feedback. In the research Marzano conducted, he determined that providing descriptive feedback is the most significant strategy to increase student learning (Marzano, Pickering, & Pollock, 2001). Descriptive feedback provides students with information regarding what areas in which they are doing well, and provides the next step in classroom learning.

While few would challenge the premise that feedback is essential for learning, not all feedback is effective. In fact, some forms of feedback can have negative effects on student achievement (Hattie, 1992). Two separate reviews of research on feedback (Bangert-Drowns, Kulik, & Kulik, 1991; Fuchs & Fuchs, 1986) found that a substantial number of studies showed negative effects of feedback. In these instances, feedback actually harmed learning. Table 2 reflects the specific results of these two reviews of the research and shows that when students received feedback on whether they were right or wrong, with no explanation, it decreased student achievement. Amazingly, in this case, no feedback at all would have been better. Conversely, students provided feedback using a rule as the basis of the evaluation increased their achievement dramatically. Study Island has the ability to offer students an explanation regarding why an answer is right or wrong.

The feedback must be timely and specific in order to have the biggest impact on student achievement. Marzano (2003) asserted the following about providing feedback:

- Feedback should be corrective in nature.
- Feedback should be timely.
- Feedback should be specific to a criterion.
- Students can effectively provide some of their own feedback.

When teachers have an entire class or several classes of assignments to grade it is difficult to provide timely, corrective, and criterion specific feedback. When teachers do not promptly return assignments to students with specific suggestions on how to get better, it will not have a positive impact on student achievement. Study Island provides students immediate feedback on specific standards.

Table 2
Findings on the Effects of Different Types of Feedback

Source	Characteristics of Feedback from Classroom Assessments	Number of Studies	Effect Size	Percentile Gain or Loss in Student Achievement
Bangert-Drown, Kulik, Kulik, & Morgan (1991)	Right/wrong	6	-.08	-3
	Provide correct answer	39	.22	8.5
	Criteria understood by students vs. not understood	30	.41	16
	Explain	9	.53	20
	Repeat until correct	4	.53	20
Fuchs & Fuchs (1986)	Displaying results graphically	89	.70	26
	Evaluation (interpretation) by rule	49	.91	32

Note, From *Classroom Assessment and Grading That Works*, by R. J. Marzano, 2006, Alexandria, VA: Association for Supervision and Curriculum Development.

Opportunity to Learn and Time on Task

The instructional practices of teachers have a profound impact on the opportunity to learn and time on task of students. As expected, each day that students are in class is an opportunity to learn, and the more engaged they are, the better it will be for student learning. In effective schools, teachers spend a good deal of time delivering instruction on the essential learning objectives. For a high percentage of this time, students are actively engaged in teacher directed, large group activities (Lezotte, 2001).

Lezotte (1991) suggested creating an “interdisciplinary curriculum” to teach the necessary skills in the least amount of time, making decisions about what is most important, and letting go of the rest, called “organized abandonment.” Given the amount of curricular objectives teachers are expected to cover in a school year, it is imperative that they focus on monitoring student progress on the most important standards, sometimes referred to as “power” standards (Ainsworth, 2003). Power standards are curricular standards identified as more important because they are vital for the next level of learning. Study Island provides students and teachers with the opportunity to engage in extra practice on these power standards outside of the classroom.

The ability to use Study Island inside and outside of the classroom allows educators to provide intervention and enrichment opportunities for students. Specifically, teachers can intervene with students who are struggling to master certain standards by giving them more opportunities and time to practice those standards using Study Island. Likewise, teachers can enrich students who have already mastered the standards by giving them opportunities and time to move on to other standards. Teachers and schools that provide interventions for struggling students and enrichment opportunities for students who are excelling will experience gains in overall student achievement as well as the achievement of subgroups.

Positive Home-School Relationships

In order for students to be engaged in learning outside of school, a positive home-school relationship needs to be established. In effective schools, parents understand and support the basic mission of the school and opportunities are available for them to play important roles in helping the school to achieve its mission (Lezotte, 2001). Unfortunately, students who most desperately need extra support often come from low socioeconomic households. The parents of these households are often unable to support a home-school relationship and lack the educational background to help their children with their academic endeavors outside of school. Therefore, it is important for teachers to make web-based learning opportunities available for students when they are outside of school. When students can access learning opportunities, such as Study Island, from home, then engagement and learning occur no matter what level of education of the parents.

Study Island promotes parent involvement as well as strengthens the home-school connection (Study Island, 2010). Moreover, schools in this study sent username and password information home to parents and provided parents with information about the program and its use from home. Some schools held "Study Island Nights" where they demonstrated the program and illustrated how it helps improve scores on the state tests.

Assessment Research

Over the past decade, formative assessment and summative assessment have become buzzwords in education; however, the terms formative and summative are not new. In 1967, Michael Scriven coined the terms formative and summative when explaining the differences between formative evaluations and summative evaluations. Furthermore, Scriven emphasized the information each type of evaluation provided and its use. The notion of formative assessment was later incorporated into the practice of Mastery Learning (Bloom, Hastings, & Madaus, 1971). The authors of Mastery Learning called for the administration of a formative assessment upon the completion of instruction for a particular unit. Then, for students who had not mastered objectives, the teacher used diagnostic information from the assessment to provide further instruction that targeted the specific needs of the student. Additionally, in 1989, D. Royce Sadler determined that in order for an assessment to be "formative" it must (a) come to hold a concept of quality roughly similar to that of the teacher, (b) be able to compare the current level of performance with the standard, (c) be able to take action to close the gap (Shepard, 2005).

More recently, Black and Wiliam (1998b) have lobbied to expand the definition of formative assessment beyond that which Bloom described. They provided examples of how to use assessments formatively, even if it did not occur during day-to-day instruction with immediate feedback to follow. When teachers analyze where students are in their learning and provide specific, focused feedback on performance and ways to improve it, traditional tests and homework become formative assessments. Black and Wiliam make the following recommendations:

- Frequent short tests are better than infrequent long ones.
- New learning should be tested within about a week of first exposure.
- Be mindful of quality test items and work with other teachers and outside sources to collect good ones.

Researchers have not provided a single definition that encapsulates formative assessment. One research article might use the term *formative assessment,* while other research articles use the terms *classroom assessment* or *instructional assessment*. Although these terms differ, they provide similar definitions and examples. For example, Popham (2008) defined formative assessment as a series of evidence-collecting and decision-making events for both teachers and students in order to help students learn. For the purpose of this literature review, the term *assessment* refers to all those activities undertaken by teachers and students that provide information used as feedback to modify teaching and learning activities. Assessment becomes *formative* when the evidence is actually used to adapt teaching to meet student needs (Black & Wiliam, 1998b).

Formative assessment encompasses the bulk of the research presented in this study; however, educators also use summative assessment to effect student achievement. Currently, assessment practices in the United States are largely summative, especially those connected to No Child Left Behind. Summative assessments occur at the conclusion of an instructional period, with results typically released months later. Therefore, the results of summative assessments do not have as much influence on student learning. Summative assessments are a good tool for monitoring student achievement. However, if the goal is to improve learning, a more formative approach is needed (Stiggins, 2007).

Assessment and student achievement. In 1998, Paul Black and Dylan Wiliam published their work on formative assessment titled *Assessment and Classroom Learning*. Black and Wiliam reviewed 250 journal articles and reports to determine whether classroom-based formative assessment increased academic achievement (Black & Wiliam, 1998a). Amazingly, the researchers found that typical effect sizes of the formative assessment experiments were between 0.4 and 0.7. This means that an effect size gain of 0.7 in the recent international comparative studies in mathematics would have raised the score of a nation in the middle of 41 countries, such as the United States, to one of the top five. This illustrated the profound impact that formative assessment has on student achievement.

As the conclusions of Black and Wiliam (1998) gradually spread into faculty lounges, test publishers began to re-label many of their tests as "formative," perpetuating the buzz surrounding formative assessment (Popham, 2006). In reality, testing companies simply renamed existing products, many of which were summative in nature, to capitalize on the clamoring for formative assessments created within the educational community. Unfortunately, educators believed that student achievement would increase simply by administering these formative assessments. The assessments marked as formative by testing companies were not the same as the formative assessments found in the research of Black and Wiliam.

According to Black and Wiliam (1998a), in most of the studies they reviewed, they observed another common trend that low achievers benefited more from formative assessments than students with higher achievement did. This finding is significant because it implies that formative assessment practices can help close the achievement gap, while also raising overall achievement. Educators know they need to provide help to low-achieving students however, they often struggle to find the right interventions. The effect size gains attributed to formative assessment that were reported by Black and Wiliam are larger than most of those found in other educational interventions. Therefore, educators were quick to try anything related to formative assessment in an effort to help low achievers. In general, assessment experts concur that when everyday classroom instruction includes formative assessment, student achievement increases (e.g. Boston, 2002; Chappuis & Stiggins, 2002; Crooks, 1988; Stiggins, 1998).

The effect of summative assessment practices on student achievement is lower when compared with the effect of formative assessment practices (Crooks, 1988). In fact, summative assessments, in particular high-stakes accountability assessments, tend to be instructionally insensitive and may even undermine efforts to improve student learning (Popham, 2007; 2009). Moreover, summative assessments used for accountability purposes can have a negative effect on low-achieving students (Harlen & Deakin, 2002). When looking at No Child Left Behind subgroups, a literature review by Solorzano (2008) suggested that high-stakes accountability tests do not accurately gauge achievement of English-language learners and may actually widen the gap because of punitive consequences, such as unequal retention and graduation rates. Unfortunately, summative assessments may also reduce intrinsic motivation, increase test anxiety, lower self-efficacy, cause poor relationships among students, and reduce the use and effectiveness of teacher feedback (Crooks, 1988). All of this research points to the fact that students who do not typically perform well on high-stakes test are already aware of this fact, and when forced to take them, the less beneficial it is for their learning. Study Island provides struggling students with a low-risk and high-reward opportunity to engage in assessment practice that is not detrimental to their self-efficacy.

Formative assessment generalizations and practices. In 2006, Robert Marzano's *Classroom Assessment and Grading That Works* provided an overview of current research on formative assessments. From this research, Marzano asserted four generalizations:

- Students should gain a clear picture of their progress on learning goals and understand how to improve when provided feedback on classroom assessments.

- Feedback on classroom assessment should encourage students to improve.
- Classroom assessment should be formative in nature.
- Formative classroom assessment should be frequent (Marzano, 2006, p. 3).

The conclusions Marzano reached reiterate the need for formative assessment and feedback. In addition, they called for formative assessment to occur frequently. Table 3 displays a study conducted by Bangert-Drowns, Kulik, and Kulik (1991), which shows the effect size of achievement gains over a 15-week period. As the frequency of assessments increased, so did the effect size and percentile-point gain. However, it is clear that achievement gains leveled at certain points. Therefore, Table 3 should not compel educators to conduct 30 assessments over 15 weeks; instead, it should illustrate the profound impact of frequent formative assessment (Marzano, 2006).

Within the classroom, teachers use formative assessment practices that enhance student understanding. There are endless examples, some described in this section, of practices that teachers often use in the classroom. Black and Wiliam (1998b) encouraged teachers to use questioning and classroom discussion as an opportunity to increase the knowledge of students and improve understanding. Nevertheless, they caution teachers to ask thoughtful, reflective questions rather than simple, factual ones and then give students adequate time to respond. In order to involve everyone, Black and Wiliam suggested strategies such as the following:

- Invite students to discuss their thinking about a question or topic in pairs or a small group, then ask a representative to share the thinking with the larger group.
- Present several possible answers to a question, then ask students to vote on them.
- Ask all students to write down an answer, then read a selected few aloud.

Teachers might also assess the understanding of students in the following ways:

- Have students write their understanding of vocabulary or concepts before and after instruction.
- Ask students to summarize the main ideas they have taken away from a lecture, discussion, or assigned reading.
- Have students complete a few problems or questions at the end of instruction and check answers.
- Interview students individually or in groups about their thinking as they solve problems.
- Assign brief, in-class writing assignments about the topic.

Formative assessment and classroom instruction. Educators have learned that involving students in assessment causes assessment to become instruction (Davies, 2000). "In classrooms that use assessment to support learning, teachers continually adapt instruction to meet student needs" (Leahy, Lyon, Thompson, & Wiliam, 2005, p. 19). Moreover, when assessment supports learning, teachers allow the data to drive the instruction and by linking formative assessment and classroom instruction teachers are able to produce increased learning. Black and Wiliam (1998) found that students that had opportunities to be with teachers that used formative assessment as an instructional practice learned in approximately six or seven months, instead of a year (Black, Harrison, Lee, Marshall, & Wiliam, 2004).

Table 3
Gain Associated with Number of Assessments Over 15 Weeks

Number of Assessments	Effect Size	Percentile-Point Gain
0	0	0
1	.34	13.5
5	.53	20.0
10	.60	22.5
15	.66	24.5
20	.71	26.0
25	.78	28.5
30	.80	29.0

Note, From *Classroom Assessment and Grading That Works*, by R. J. Marzano, 2006, Alexandria, VA: Association for Supervision and Curriculum Development.

Formative assessment and student motivation. In addition to increasing student achievement, Black and Wiliam (1998a) suggest that formative assessment benefits student motivation. In fact, research suggests that when students partner with teachers in the assessment process, they take more responsibility for their own learning (Rieg, 2007). Students gain a feeling of empowerment when they help determine the criteria that teachers will use to judge their work (Brookhart, 1997). To that end, in order to improve student motivation, assessments must provide students frequent opportunities to improve their work so that mistakes and errors are considered part of the learning process (Cauley, Pannozzo, Abrams, McMillan, & Camou-Linkroum, 2006). According to Crooks (1988), effective education gives just as much attention to intrinsic interest and motivation as it does to cognitive outcomes. Study Island provides unlimited opportunities for students to answer questions, while motivating students to strive for getting the answer correct by rewarding them with opportunities to play games.

Formative assessment and feedback. Feedback has long been regarded as a key component of the assessment process and a critical piece of the student achievement puzzle (Callingham, 2008; Cauley et al., 2006; Hattie & Temperley, 2007; Shepard, 2000; Stiggins, 2004). Study Island provides formative and immediate feedback to students as well as real-time information to teachers, administrators, and parents. In fact, Study Island developers suggest these features are keys to accelerating a struggling student to proficiency and monitoring progress to allow information to guide instruction (Study Island, 2010). According to the *Review of Educational Research*, when students are provided with feedback about their learning and engage in self-assessment, their achievement gains are profound, especially in lower achieving students (Tschannen-Moran, Hoy, & Hoy, 1998). Study Island can provide formative feedback to students and the district in this study began using it in hopes that it would increase the achievement of struggling students.

The formative feedback and information Study Island provides is important because state tests are summative in nature, meaning they occur at the end of the school year. Unfortunately, the results of state tests are often not available to stakeholders until the beginning of the following school year. The ability of Study Island to provide administrators, teachers, students, and parents with timely information on student achievement aligned to specific state standards throughout the school year is an important feature.

Formative assessment and alignment. In order for formative assessment to effect student achievement, it must address the individual needs of the student and be aligned to specific learning objectives. The learning objectives should correspond to state and national standards. Regrettably, in many instances, meeting the individual needs of each student is not possible if all students are working simultaneously on the same assignments and trying to meet the same learning objectives (Crooks, 1998). In order to challenge students, it is important for assessment to align with their individual needs (Brimijoin, Marquissee, & Tomlinson, 2003). Research indicates that teachers and students should work together to develop targeted learning objectives that provide a progression of student learning, and that the formative assessments should be aligned to those objectives (Ayala, Shavelson, Ruiz-Primo, Brandon, Yin, Furtak, 2008; Stiggins & Chappuis, 2008; Valencia, 2008; Wiley, 2008).

Study Island has customized versions of their resource that align to the specific standards of each state (Study Island, 2010). It provides teachers the ability to assign specific standards to students for them to practice. In addition, students can also choose specific standards for practice. Prior to answering any questions on a specific standard, students are engaged in a brief lesson or tutorial. At the conclusion of the lesson or tutorial, students answer a series of multiple-choice questions. Fortunately, Study Island contains a large item bank so students rarely receive the same question twice and the placement of the correct answer varies to help promote learning the concept instead of guessing or memorizing the answers (Study Island, 2010). The alignment features of Study Island make it a valuable tool that educators can use when formatively assessing students.

Formative assessment and technology. Current advancements in technology, such as computer-based software, allow teachers and students to more effectively and efficiently assess and track achievement. Technology provides immediate feedback to students regarding their performance on formative assessments, which can improve student achievement (Epstein, Lazarus, Calvano, Matthews, Hendel, Epstein, 2002; Kulik & Kulik, 1988). As expected, the number of formative assessment programs delivered through software or the Internet has grown rapidly and, likewise, so have the capabilities of these programs.

Teachers and students currently have access to formative assessment programs that do far more than tell them whether an answer is right or wrong. They can score written (typed) responses and essays in the blink of an eye, as well as provide specific feedback on what the next steps in the learning process should be (Landauer, Lochbaum, & Dooley, 2009). However, as teachers gain access to new assessment tools, they must find ways of using the richer information acquired with those tools to impact instruction and increase student achievement. Otherwise, the new tools will continue to serve the traditional summative assessment purposes without significantly effecting student achievement (Even, 2005).

Study Island is a resource that integrates technology and formative assessment. Students can use it on a frequent basis, and there are features that motivate students to improve. For example, each time questions are correctly answered by students, they earn the right to play games. Likewise, teachers can set a desired level of mastery on a specific standard and when students reach that level, they earn blue ribbons. Although teachers could certainly use work completed while using Study Island as grades, this is not the intended purpose. Instead, Study Island is a program where students can practice and get better without fear of negatively effecting their grade. All of these Study Island attributes are research-based formative assessment practices.

No Child Left Behind

With the passage of No Child Left Behind, Congress set in motion a sweeping overhaul of federal efforts to support elementary and secondary education in the United States. As stated earlier, the federal law includes the four pillars described as accountability for results, an emphasis on doing what works based on scientific research, expanded parental options, and expanded local control and flexibility (DOE, 2004). The following sections provide a general overview of the components that make up the four pillars.

Accountability for results. Through the administration of state tests, No Child Left Behind identifies schools that it terms are “in need of improvement” (DOE, 2004). These schools must involve administrators, teachers, parents, and outside experts to develop a plan for improvement. This plan must involve spending money on teacher professional development. In addition to setting expectations for state test results, No Child Left Behind also put in place minimum qualifications for teachers.

No Child Left Behind required all states to set annual performance targets that each school must meet. Annual performance targets are the percentage of students that need to score proficient on the state tests in communication arts and mathematics for that particular school year. Oddly enough, the annual performance targets can vary by state and subject however, they must culminate in the ultimate goal of 100% proficiency by 2014. Table 4 illustrates the annual performance targets for the state of Missouri. When a school meets the annual performance target in both communication arts and mathematics in every subgroup for which they are accountable, as well as the overall population, they meet Adequate Yearly Progress.

Table 4
Missouri's Adequate Yearly Progress Annual Performance Targets

Year	Communication Arts	Mathematics
2002	18.4	8.3
2003	19.4	9.3
2004	20.4	10.3
2005	26.6	17.5
2006	34.7	26.6
2007	42.9	35.8
2008	51.0	45.0
2009	59.2	54.1
2010	67.4	63.3
2011	75.5	72.5
2012	83.7	81.7
2013	91.8	90.8
2014	100.0	100.0

Note. Adapted from Missouri Department of Elementary and Secondary Education School Profile Student Demographics. Retrieved February 22, 2010, from http://dese.mo.gov/planning/profile/.

Research-based actions. No Child Left Behind aims to support programs and practices that are research-based. An emphasis is on those programs that support learning in the early years. An example of a research-based program that No Child Left Behind supports is the Reading First program. Other examples are on the What Works Clearinghouse (http://ies.ed.gov/ncee/wwc/) website created by the United States Department of Education.

Expanded parental options. No Child Left Behind calls for parents to receive detailed report cards on schools and districts. These provisions ensure that parents have important and timely information about the schools their children attend. Parents can transfer their child to a higher-performing school within the district or a charter school, if the school their child currently attends is in need of improvement. In addition, if a school progresses into further levels of needing improvement, they are required to offer supplemental educational services, such as free tutoring (DOE, 2004).

The results of the state tests administered on an annual basis have a profound impact on districts and schools. In fact, schools that receive Title I funds face more intense sanctions for not meeting Adequate Yearly Progress than schools that do not receive Title I funds. Title I funds come from the federal government, and schools have the option of whether or not to accept these funds. Socioeconomic status is used to determine whether a school is eligible to receive Title I funds (DOE, 2004).

Title I schools that do not meet Adequate Yearly Progress are required to provide supplemental educational services. Supplemental educational services include offering free tutoring opportunities to low-income students outside of the school day, such as after school, before school, or during summer school (DOE, 2010). Conversely, non-Title I schools, schools that do not receive federal money, can be schools in need of improvement; however, since they do not receive federal money, the sanctions for not meeting Adequate Yearly Progress are less intense and usually involve revising the school improvement plan. These schools in need of improvement are looking for programs to help improve student achievement and it is of particular interest to research whether money spent on Study Island can support increasing student achievement on the state tests.

Expanded local control and flexibility. No Child Left Behind gives states and districts flexibility regarding how they use their federal funding. This includes how they spend their professional development money. Similarly, the federal law allows flexibility in how schools and districts retain and attract highly qualified teachers. This includes alternative routes to certification and merit pay plans. These features of No Child Left Behind provide states, districts, and schools with the ability to address their unique needs and challenges (DOE, 2004).

Missouri Annual Performance Report

The No Child Left Behind section in this chapter provided an overview of how districts and schools are accountable under the federal law. In Missouri, in addition to federal accountability, officials use the results of state tests to evaluate schools and districts at the state level. The evaluation at the state level is for accreditation purposes and the indicators used are known collectively as the Annual Performance Report. In fact, Missouri state officials issue each district an Annual Performance Report that includes 14 total indicators. In addition to MAP achievement indicators, the report includes attendance, graduation rate, and college and career readiness indicators. If a district becomes unaccredited by the state of Missouri, the parents and students may be given the opportunity to attend different public schools in adjoining counties (MO DESE, 2009e).

The MAP Performance Index calculates the movement of students throughout all of the MAP achievement levels. The MAP Performance Index takes student assessment information from all of the schools within a district and aggregates it together to determine if the district has met Annual Performance Report achievement indicators (MO DESE, 2009e). Instead of setting annual performance targets like No Child Left Behind, the Annual Performance Report measures overall student achievement improvement within each achievement level (Below Basic, Basic, Proficient, Advanced) compared to the previous year and evaluates districts accordingly.

Student Achievement

Has student achievement increased and has the achievement gap decreased since No Child Left Behind? This is a fair question since it was the stated purpose of why federal lawmakers designed No Child Left Behind. Researchers at the Center on Education Policy, an independent non-profit organization, set out to answer this question when they conducted the most comprehensive study of trends in state test scores since No Child Left Behind took effect in 2002. The study included state test data from all 50 states in both communication arts and mathematics.

The comprehensive study of No Child Left Behind state test scores occurred under the supervision of five experts in educational policy and testing. The researchers constructed the study to serve two main purposes, informational and educational. The informational purpose sought to answer whether student achievement has increased and the achievement gap between different subgroups had decreased since No Child Left Behind. The educational purpose targeted policymakers, and others, in an effort to explain what could, and could not, be known about student achievement based on the available data. This research team, based on the data that states provided, reached five main conclusions (CEP, 2007).

- Since No Child Left Behind, most states with three or more years of comparable test data increased student achievement in reading and math.
- There is more evidence of achievement gaps between groups of students narrowing since 2002 than evidence of gaps widening.
- In 9 of the 13 states with sufficient data to determine pre- and post-No Child Left Behind trends, average yearly gains in test scores were greater after No Child Left Behind took effect than before.
- It is very difficult, if not impossible, to determine the extent to which these trends in test results have occurred because of No Child Left Behind. Since 2002, state, school district, and school officials have simultaneously implemented many different but interconnected policies to raise achievement.
- Although No Child Left Behind emphasizes public reporting of state test data, before 2002 the data necessary to reach definitive conclusions about achievement were sometimes hard to find, unavailable, or had holes or discrepancies.

The conclusions stated in the previous paragraph are important to consider for two reasons as they relate to this dissertation study. First, it is more difficult than imagined to find states where comparable data exists for three or more years. Second, the type of available achievement data such as percent proficient, scale scores, and standard deviations vary by state. Therefore, it is difficult for researchers to conduct rigorous studies that include longitudinal achievement data. The rest of this section presents barriers researchers face when attempting to measure the effect of No Child Left Behind on student achievement and compares them to researching the effect of Study Island on MAP student achievement.

Quality and limitations of state tests. Given the immense weight that No Child Left Behind places on state tests, it might seem logical that the data needed to draw firm conclusions about student achievement were readily available and easily interpreted. In reality, attempts to conduct rigorous studies of No Child Left Behind state test score trends across the nation collapse due to missing, limited, inconsistent, and breaks in the data. In fact, only 13 states had data to enable a comparison of achievement trends, within that particular state, prior to and since No Child Left Behind (CEP, 2007). Technical issues with testing providers, continual revisions to state tests, and overworked state education departments have led to the incongruence of data. In addition, the data necessary to do in-depth studies of achievement trends, such as mean scale scores and standard deviation, are not available in many states. Fortunately, for this study the MAP provides educators with percent proficient and scale score data.

Prior to the implementation of No Child Left Behind, states were not required to publically report any type of student achievement data such as percent proficient or mean scale scores. Therefore, it should come as no surprise that No Child Left Behind ignited significant changes in and expansion of state testing programs. In an effort to comply with the new federal law, states changed several aspects of their testing programs, such as administering different tests, altering proficiency levels or cut scores, and changing the scoring scales. Consequently, all these changes have rendered it almost impossible to compare pre- and post-No Child Left Behind data, as illustrated by 37 states having breaks in their achievement data since 2002 (CEP, 2007). While the implementation of No Child Left Behind has made it somewhat easier to measure achievement, further steps are necessary to help researchers draw accurate conclusions about student achievement trends.

Even if No Child Left Behind required more rigorous achievement data and breaks in achievement data did not occur, barriers to interpreting student achievement data would still exist. Interpretation and evaluation of test score trends is complicated even in states that can provide data on percentage proficient, mean scale scores, and standard deviations. For example, there is a certain degree of distortion in state test results that comes from the way they are created, administered, and scored (CEP, 2007). Aside from the obvious breaks in data that may occur, there can still be minor manipulations of tests through a series of small changes made by test administrators, none of which individually raise a concern. However, in sum, these subtle decisions can effect the comparability of results from year to year. Some examples of subtle manipulations that may have a large effect include providing multiple forms of a test, weighting test questions, embedding field test items, changing scoring procedures, and re-using test questions.

To a lesser extent, when compared to other states, changes in the MAP mirror those mentioned in this section. Since the implementation of the MAP in 2002, a switch from grade-span to grade-level tests occurred, thus, introducing more and different state tests. In addition, changes occurred in the number of achievement levels, scale scores, and proficiency cut scores. Fortunately, for this study, the MAP state tests at the elementary and middle school levels have remained virtually unchanged for the past five years, 2006 to 2010 (MO DESE, 2010b). Nevertheless, due to budget constraints the 2011 MAP state tests will undergo changes that may cause another break in comparable data (VanDeZande, 2010).

Many studies rely on state test scores as the primary measure of student achievement. However, achievement and test scores are not the same. Although state tests are valid, reliable, and objective, they are incomplete measures of learning (CEP, 2007). For the most part, state tests are comprised of multiple-choice questions, which require students to select the best answer given four choices. Some state tests use constructed response questions and performance events that allow students to write out and explain their answers. However, state tests do not allow students to express their knowledge authentically or creatively. In addition, state tests occur on a particular day and, therefore, lack the ability to allow students an extended period to demonstrate what they know and can do. This said, despite all of the challenges associated with measuring achievement through state tests, they are still the best indicator available to draw inferences about student learning. Hence, the decision to use MAP scale scores in this study to measure the effect of Study Island on student achievement.

Achievement gap trends. Before No Child Left Behind, students with disabilities and limited English proficiency students were exempt from standardized testing or given different tests than others (National Research Council, 1997). In an effort to address the achievement gap, a primary purpose of No Child Left Behind was to highlight differences in the achievement of student subgroups. The law required states, districts, and schools to disaggregate test scores by subgroup and report them to the public. This forced school officials to look beyond their overall average and reflect on how all students were achieving on the state tests. Today, educators are more aware and feel increased pressure to address the needs of student subgroups that have traditionally not scored well on state tests.

Among states with enough data to identify trends by subgroup, the number of states in which achievement gaps among subgroups decreased since 2002 far exceeds the number of states in which these gaps increased. In fact, all of the states that decreased the achievement gaps of subgroups also increased overall achievement. More specifically, 14 of the 38 states with the necessary data showed evidence of African-American reading gaps narrowing, while no state showed evidence of the gap widening. In mathematics, 12 states closed the gap while one widened. Similar results were noted for the Hispanic and low-income subgroups. Data on achievement gap trends for students with disabilities and Limited English Proficiency students were not reliable enough to support solid conclusions (CEP, 2007).

The district in this study began using Study Island to increase overall and subgroup achievement on the MAP in hopes of meeting Adequate Yearly Progress in communication arts and mathematics. Certainly, the emphasis No Child Left Behind places on subgroup achievement led the district to explore programs designed to raise performance on state tests. In previous years, the district and schools had been successful in raising overall achievement; however, certain subgroups continued to struggle. Just as many states have shown progress in closing achievement gaps since No Child Left

Behind, the district in this study has also narrowed achievement gaps since No Child Left Behind.

Achievement levels. No Child Left Behind uses the percentage of students scoring at or above the proficient level on state tests to measure student achievement. However, the federal law does not define a proficient, or on grade-level, performance. Instead, No Child Left Behind required officials in each state to set their own proficiency standards and measure student progress with their own state tests. According to Gary Phillips, American Institutes for Research Vice-President, this is a fundamental flaw with No Child Left Behind because it allows states to report high levels of achievement by setting low standards (American Institutes for Research, 2010). Since the criteria for a proficient performance varies from state to state, it is inadvisable to use percent proficient to compare student achievement among states.

Massachusetts has been recognized nationally for the high proficiency standards officials set. On the other hand, until just recently, Tennessee was a state where officials had set low proficiency standards. In fact, the bar for eighth grade math proficiency in Massachusetts was two full standard deviations above the proficiency bar of Tennessee. Shockingly, this gap represents more than four grade levels difference between proficient eighth graders in the two states (Sparks, 2010). With state standards that vary this widely, it is easy to see why comparing No Child Left Behind results among states is inadvisable.

Missouri, like Massachusetts, has standards that are much higher than Tennessee. In 2009, Missouri had a lower percentage of students that scored proficient on their state tests when compared to Tennessee. For example, Tennessee had over 80% of students score proficient or better on state tests (Roberts, 2010), while not even 50% of students in Missouri scored proficient (MO DESE, 2010). Certainly, this does not mean that Missouri students are not as intelligent as their peers in Tennessee are. In fact, when comparing the percentage of students scoring proficient or above from both states on the 2009 National Assessment of Educational Progress fourth grade mathematics test, the data shows that Missouri had a higher percentage of students scoring proficient compared to Tennessee (National Assessment of Educational Progress, 2010). The National Assessment of Educational Progress is a national assessment with a consistent proficiency level applied across all states.

The federal government supports the annual use of the National Assessment of Educational Progress in addition to the state tests required by No Child Left Behind. Moreover, the National Assessment of Educational Progress, also known as "the Nation's Report Card," is the only national assessment of what students in the United States know and can do in various subject areas. In addition, it provides a measure of student achievement independent of state tests and differs from state tests in the content, question type, and rigor. Although not all students across the nation take the National Assessment of Educational Progress, a representative sample of students participate in each state. Therefore, it yields both national and state-level results by grade and by subgroup (NAEP, 2010).

Overall, National Assessment of Educational Progress result trends since 2002 show a less positive picture of student achievement than No Child Left Behind state test results. In fact, a low correlation existed between achievement gains on state tests and gains on the National Assessment of Educational Progress. However, the National

Assessment of Educational Progress results are not the "gold standard" and should not negate state test results. Instead, they provide an additional information about achievement. While the National Assessment of Educational Progress provides a useful independent measure, it also has limitations such as lack of alignment to state standards, less motivation for students to do well, and testing a changing population of students. Although far from perfect, state tests are still the best available measure of student learning in relationship to the curriculum (CEP, 2007).

Study Island offers the ability for students to take benchmark assessments in communication arts and mathematics several times throughout the year. The benchmark assessments predict the achievement level the student will earn on the state tests. In order to predict state test achievement levels, the rigor of the benchmark assessments must align with the state tests. Therefore, Study Island must set the achievement levels of their benchmark assessments to match the state in which the product is used.

Educators criticize the percent proficient measure used by No Child Left Behind because it only provides a picture of one level of achievement, thereby failing to reflect information about student achievement above or below that level. This makes it difficult for both educators and state officials to determine whether there has been achievement growth from one year to the next. For example, a school could increase the percent of students scoring just below proficient (Basic) and decrease the percent of students scoring well below proficient (Below Basic), meaning that the achievement of several students was raised from Below Basic to Basic. However, since proficient or above is the only measure, it would not appear that achievement increased in the school. Nevertheless, examination of data within both of the achievement levels below proficient would support that achievement had in fact increased. Conversely, a school could have a decrease in the percent of students scoring well above proficient (Advanced) and still have the same percent of students scoring at or above proficient. This means that fewer students scored Advanced however, they were still able to earn an achievement level of Proficient. In this example, the data in the achievement level well above proficient would suggest that achievement has actually decreased. Both of these examples illustrate the need to look at achievement levels below and above proficient, as well as within proficient, to obtain a clear picture of student achievement and achievement gaps (MO DESE, 2009a).

Scale scores, standard deviations, and effect sizes. There are inherent disadvantages when states are only required to report percent proficient data. If states were required to report data such as mean scale scores, standard deviations, and effect sizes, researchers could more easily determine achievement trends. Unfortunately, No Child Left Behind does not require states to collect mean scale scores and standard deviations. Mean scale scores occur on an interval (numerical) scale and permit more rigorous quantitative analysis than a simple determination of whether a student falls into the proficient or non-proficient category (Bluman, 2008). Standard deviations are a measure of how spread out or close together test scores are and exist for any set of data. If test scores are spread out, the standard deviation value is higher than if test scores are close together (Bluman, 2008). Some state education departments, such as the Missouri Department of Elementary and Secondary Education, collect mean scale scores and standard deviation data. The ability to analyze MAP average scale scores to determine the effect of Study Island makes for a more credible study as opposed to analyzing percent proficient data.

Data such as mean scale scores and standard deviations are used to compute effect sizes. Effect sizes are computed by subtracting the year 1 mean test score from the year 2 mean test score and dividing by the average standard deviation of the two years. An effect size of 0 indicates no change in the average score while an effect size of +1 indicates a shift upward of 1 standard deviation from the previous year's mean test score (Bluman, 2008). Since No Child Left Behind only requires states to report in terms of percent proficient, only 30 states had both percent proficient and effect size data (CEP, 2007).

Although data such as mean scale scores, standard deviations, and effect sizes can help researchers determine achievement trends, there are still drawbacks. For example, effect sizes do not take into account the relative difficulty of tests and standards from state to state. Thus, an easy test could yield a large effect size, while a more difficult test could produce a small effect size. Therefore, it is important to control for test rigor from one year to the next when using test scores to evaluate the impact of a resource or program. In this study, the MAP state tests used in 2008 and 2009 were of equal rigor (MO DESE, 2008 & MO DESE, 2009d).

Conclusion

Lawmakers designed No Child Left Behind to improve student achievement and close achievement gaps. Therefore, it is of interest to determine whether student achievement has increased and achievement gaps have decreased since No Child Left Behind. However, it is difficult to determine the exact impact of No Child Left Behind.

There has definitely been an increased focus on raising state test scores and publicly reporting the results. However, from a technical standpoint, the majority of state tests have limitations, and this does not make it easy to determine whether student achievement has increased. That said, state tests are still the best measures of student achievement currently available to researchers, and examination of the data does point to more instances of increased achievement and narrowing of achievement gaps.

The accountability portion of No Child Left Behind, known as Adequate Yearly Progress, has raised the awareness of districts and schools, mainly due to the sanctions applied when they do not meet the annual performance targets. Therefore, educators have been searching for school improvement models and programs that can help them keep pace with the annual performance targets and meet Adequate Yearly Progress. The charge placed on public schools by No Child Left Behind, 100% proficient in communication arts and mathematics by 2014, is in stark contrast to the findings of the 1966 *Equality of Educational Opportunity* report by James Coleman. In the report, Coleman concluded that only 10% of the variance in student achievement attributed to schools, while the other 90% attributed to student background characteristics. The Correlates of Effective Schools, introduced a decade and a half after the Coleman report, challenged the assertion that schools did not matter and illustrated examples of schools with high achievement that overcame barriers of poverty and lack of parental education. It would be hard to imagine a federal law such as No Child Left Behind if the assertions in the Coleman report had gone unchallenged.

The research on school improvement and increasing student achievement contains common themes among the various researchers and models. Among the commonalities are frequent monitoring of student progress, opportunity to learn and time on task, and positive home-school environment. An integral part of any improvement initiative or model that targets student achievement will involve both formative and summative assessments. A deep understanding of formative assessment, coupled with programs that support improvement efforts, will have the greatest probability of increasing student achievement.

This chapter illustrated the barriers and limitations to quantitatively analyzing the impact of No Child Left Behind on student achievement and compared them to determining the effect of Study Island on MAP student achievement. It is more realistic, and probably more beneficial, to conduct quantitative research on the impact of No Child Left Behind and the use of Study Island at a local level, such as at one school or several schools in the same district. That is precisely what occurred in this study. Chapter 3 outlines the methodology and instrumentation used in this study.

Chapter Three: Methodology

Research Overview

This quantitative study investigated the effect of Study Island on student achievement. The study used student achievement data from 10 elementary and 5 middle schools from the same district in St. Charles, Missouri. The superintendent of the school district granted permission for the study (Appendix B). This permission included access to databases of MAP test scores and Study Island usage data.

As with any study, there are limitations associated with this study. For example, this study includes one year of Study Island usage, and a longer study is necessary for conclusive results. The demographics are unique to the district in this study and differed among schools within the study. Similarly, the use of Study Island differed among schools. Therefore, even districts with similar demographics could experience different results. However, since student data from all parts of the district had an equal chance of random selection, the results of this study have a greater probability of being applicable to other districts of similar size and demographics. Regardless of the results discussed in chapter 4, this study will not provide conclusive evidence that Study Island was the sole factor that effected student achievement.

Research Design and Perspective

The students in this study began using Study Island in 2008-2009 to increase MAP scores. The null hypothesis stated that no statistically significant difference existed between 2008 MAP average scale scores and 2009 MAP average scale scores. Therefore, statistical analysis used z tests and t tests about the mean for two samples (2008 MAP scale scores and 2009 MAP scale scores) with a 95% confidence interval.

Ultimately, the *z* tests and *t* tests evaluated randomly selected populations of the total population and seven subgroups at the elementary and middle school levels in communication arts and mathematics.

A quantitative analysis of MAP test scores occurred since the developers of the Study Island promoted the ability of the program to improve student scores on state tests (Study Island, 2010). More specifically, Study Island aligned to Missouri state standards in communication arts and mathematics and increases student achievement on the MAP state tests. For the purpose of this study, the MAP scale scores represented the following subjects and grade-levels: elementary communication arts, elementary mathematics, middle school communication arts, and middle school mathematics. Within these four areas, *z* tests and *t* tests evaluated random samples from the total population and seven subgroup populations. The groups used in this study were School Total, Asian/Pacific Islander, Black, Hispanic, White, Free and Reduced Lunch, Individual Education Program, and Limited English Proficiency. This organization allowed for the analysis of data from 32 separate *z* tests and *t* tests to determine whether a significant difference existed between 2008 MAP average scale scores and 2009 MAP average scale scores.

Instrumentation

The Department of Elementary and Secondary Education Web Application login enabled the researcher access to MAP scale scores and student demographic information. Similarly, the district granted the researcher access to the Study Island database maintained for each school within the district. The Study Island database confirmed students in the study used Study Island during the 2008-2009 school year.

The Department of Elementary and Secondary Education Web Application login allowed for the creation of four separate Excel files titled elementary communication arts, elementary mathematics, middle school communication arts, and middle school mathematics. As the name implies, each Excel file contained student demographic information and 2008 and 2009 MAP scale scores corresponding to the grade-level and subject designated in the title. Next, using the Excel filter, eight spreadsheets, within each of the original Excel files, reflected the achievement data specific to the total population and seven subgroups. The eight spreadsheets of student information and MAP scale scores, within each of the four original Excel files, received titles that corresponded to the following groups School Total, Asian/Pacific Islander, Black, Hispanic, White, Free and Reduced Lunch, Individual Education Program, and Limited English Proficiency.

The researcher used Research Randomizer (www.randomizer.org) to generate random samples from each of the spreadsheets. Altogether, *z* tests and *t* tests with a 95% confidence interval, evaluated 32 random samples of student scale scores from the 2008 and 2009 MAP state tests.

Missouri Assessment Program reliability and validity. The Missouri Department of Elementary and Secondary Education officials created the MAP state tests in response to the 1993 Outstanding Schools Act that called for all states to implement an assessment system that measured challenging academic standards (MO DESE, 2009a). As a result, the MAP state tests consisted of grade-span tests in communication arts, mathematics, and science. Grade-span tests occur at one particular grade-level in elementary and again at middle school. Table 5 provides an implementation timeline for grade-span assessments in the state of Missouri. In response to changes within No Child Left Behind, grade-span tests ended in 2006 and the Department of Elementary and Secondary Education officials contracted CTB McGraw-Hill to expand the testing program to include grade-level tests. This meant that MAP grade-level tests for students in grades 3-8 were implemented beginning with the 2005-2006 school year. Table 6 provides an implementation timeline for MAP grade-level tests (MO DESE, 2001).

Table 5
Missouri Assessment Program Grade-Span Assessment Timeline

Year	Event
1996	Show-Me Standards Approved
1996	Frameworks for Curriculum Development Published
1998	First Operational Administration of Math (4, 8, 10)
1999	First Operational Administration of Communication Arts (3, 7, 11)
2005	Last Year of Grade Span Missouri Assessment Program

Note. Adapted from Missouri Department of Elementary and Secondary Education Missouri Assessment Program Technical Report, 2009. Retrieved March 19, 2010, from http://dese.mo.gov/divimprove/assess/tech/index.html

Table 6
Missouri Assessment Program Grade-Level Assessment Timeline

Year	Event
2004	Grade Level Expectations Published
2005	Communication Arts & Math Field Testing and Standards Setting
2006	First Operational Administration Communication Arts & Math
2008	Version 2.0 Grade Level Expectations Published
2008	Last Operational Administration of High School Missouri Assessment Program
2009	Last Operational Administration of Version 1.0 Grade Level Expectations

Note. Adapted from Missouri Department of Elementary and Secondary Education Missouri Assessment Program Technical Report, 2009. Retrieved March 19, 2010, from http://dese.mo.gov/divimprove/assess/tech/index.html

Educators and personnel within the Department of Elementary and Secondary Education and CTB McGraw-Hill developed the MAP tests for grades 3-8 in accordance with accepted standards and criteria. In addition, the developers of the MAP tests designed them to provide information about what individual students know and can do relative to the Show-Me Standards. For accountability purposes, educators and policy makers may appropriately use MAP results for groups of students to judge the effectiveness of educational programs and services offered at the local level (MO DESE, 2001).

When evaluating tests, such as the MAP tests, it is important to investigate their validity and reliability. Validity is how meaningful the results are for their intended purposes and reliability is the dependability of the results (MO DESE, 2001). Knowing how test results are used, the developers of the MAP tests took steps to ensure validity. First, the Department of Elementary and Secondary Education collaborated with a reputable assessment design company, CTB McGraw-Hill, and followed industry standards. For example, as the following passage from *Standards for Educational and Psychological Testing*, explains standards for validity and reliability (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 1999).

> Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing system. This includes evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all examinees (p. 17).

Second, all items appearing on the MAP tests were scrutinized to make certain they measured the intended standard. Third, evidence exists that the MAP tests impact teacher beliefs and practices (MO DESE, 2001).

High-quality tests address reliability by consistently delivering dependable results. Interestingly, reliability is a necessary but not a sufficient condition of validity. Therefore, CTB McGraw-Hill and the Department of Elementary and Secondary Education officials have put stringent procedures in place to ensure the reliability of the MAP tests especially, when it comes to scoring MAP items. In fact, score dependability or reliability can be qualified and reported as a number ranging from 0 to 1 and the higher the coefficient, the more dependable the score. The closer the value of the reliability coefficient is to 1, the more consistent the scores, where 1 refers to a perfectly consistent test. As a rule of thumb, reliability coefficients that are equal to or greater than 0.8 are acceptable for tests of moderate lengths (MO DESE, 2001).

MAP scale scores have been found to have high reliability coefficients and can give stakeholders, such as researchers, confidence in the results. Tables 7 and 8 provide reliability coefficients for the 2009 MAP grade-level tests. Moreover, the MAP scale score reliability coefficients are comparable to those associated with tests such as the Stanford Achievement Test for ninth graders, Advanced Placement Examinations, and the American College Test (MO DESE, 2001). The scale score represents the achievement level of the student, where higher scale scores represent higher levels of achievement on the test and lower scale scores represent lower levels of achievement (MO DESE, 2009a).

Table 7

2009 Communication Arts Missouri Assessment Program Grade-Level Assessments

Grade	Number of Items	Number of Score Points	Reliability Coefficient
3	57	67	0.90
4	55	63	0.92
5	55	62	0.92
6	55	62	0.90
7	61	72	0.92
8	61	68	0.91

Note. Adapted from Missouri Department of Elementary and Secondary Education Missouri Assessment Program Technical Report, 2009. Retrieved March 19, 2010, from http://dese.mo.gov/divimprove/assess/tech/index.html

Table 8

2009 Math Missouri Assessment Program Grade-Level Assessments

Grade	Number of Items	Number of Score Points	Reliability Coefficient
3	60	67	0.92
4	65	77	0.92
5	62	69	0.91
6	61	68	0.92
7	62	69	0.92
8	64	76	0.93

Note. Adapted from Missouri Department of Elementary and Secondary Education Missouri Assessment Program Technical Report, 2009. Retrieved March 19, 2010, from http://dese.mo.gov/divimprove/assess/tech/index.html

Population and Sampling Procedures

Since 2005-2006, all students in grades 3-8 have taken the communication arts and mathematics MAP grade-level tests. The purpose of the study was to determine whether a significant difference existed between the MAP average scale scores from 2007-2008 and the MAP average scale scores from 2008-2009. Since 2008-2009 was the first year students used Study Island across the district at the elementary and middle school levels, it makes sense to choose these two years. Since the district in the study does not have a transient student population, it was easy to identify students in the district that have a MAP scale score from 2007-2008 and 2008-2009 in communication arts and mathematics. To manage the data for this study, a database containing student demographics and MAP scale scores existed that included all elementary and middle school students with both a 2008 and 2009 MAP scale score in communication arts and mathematics.

Determining the eligible sample population. When looking at all of the elementary and middle school students during the 2008-2009 school year, students without a MAP scale score from the previous year were excluded from the study. This was necessary because there was no way to measure the impact of Study Island on student achievement with just one MAP scale score. This meant that, even though third grade students took the MAP tests in 2008-2009, since this was their first and only MAP scale score, they were not eligible for the study.

After eliminating students without 2008 and 2009 MAP scale scores, the creation of four separate databases began for elementary communication arts, elementary mathematics, middle school communication arts, and middle school mathematics. Then, each of the four databases were disaggregated into eight groups that included School Total, Asian/Pacific Islander, Black, Hispanic, White, Free and Reduced Lunch, Individualized Education Program, and Limited English Proficiency. Ultimately, 32 eligible populations existed once each of the four databases (elementary communication arts, elementary mathematics, middle school communication arts, middle school mathematics) contained the disaggregated total population and seven subgroup populations.

Determining the actual sample population. Research Randomizer created the random sample populations from each of the 32 eligible populations. For larger eligible populations, 50 students comprised the random sample population, and for smaller eligible populations, 15 students comprised the random sample population. For accuracy, students within each of the 32 random sample populations cross-referenced against the Study Island database to ensure they had used the program during 2008-2009. If a student within the random sample population had not used Study Island, a replacement occurred. No replacements were necessary in this study because it was highly unlikely that a student did not ever use Study Island. For example, if a teacher only encouraged students to use Study Island from home then students without Internet access may have never used Study Island.

Research Design and Procedure

This quantitative study sought to determine whether a significant difference existed between 2008 and 2009 MAP average scale scores in communication arts and mathematics at the elementary and middle school levels. At each level, elementary and middle, the schools in this study varied in student enrollment and demographic characteristics. Similarly, Study Island usage varied by level and among schools. However, a similarity existed among each level and schools in that students used Study Island for the first time in 2008-2009.

During 2008-2009, teachers, students, and parents began using Study Island in various ways. Teachers allowed students to use Study Island on classroom computers during designated times. In fact, a small number of teachers in each school had a classroom computer for each student. However, most classrooms only had two or three student computers and in those cases, the teacher arranged to visit a computer lab. In addition to using computers, teachers with interactive whiteboards used Study Island during whole group instruction while those without interactive whiteboards printed and copied Study Island material for students. Regardless of how the teacher chose to use Study Island, each student had a username and password that enabled them to use Study Island anywhere they had access to a computer and the Internet connection. Furthermore, parents received information about Study Island and were encouraged to provide opportunities to us it at home.

In April of 2008-2009, the schools in this study administered the annual MAP state tests in communication arts and mathematics to all students in grades 3-8 as required by No Child Left Behind. As expected, the results of the MAP tests taken during the 2008-2009 school year became available in August of 2009. Once the results were available, a database, constructed in Excel, consisted of elementary and middle school student demographic information and MAP scale scores from 2007-2008 and 2008-2009. The 2008-2009 Study Island database helped serve as a reference point to check student usage.

Preparing the data for analysis. The researcher took the Excel database of student demographic information and MAP scale scores from 2008 and 2009 and eliminated any student that did not have a MAP scale score for 2008 and 2009. The researcher divided the one original database into two separate databases, communication arts and mathematics. Next, two more databases, elementary and middle school, emerged by splitting the communication arts and mathematics databases. The elementary database contained students that were in grades 4 and 5 during the 2008-2009 school year and the middle school database contained students that were in grades 6-8. Overall, four databases represented elementary communication arts, elementary mathematics, middle school communication arts, and middle school mathematics.

For each of the four databases, Excel filters used the student demographic columns to sort and create eight separate spreadsheets that corresponded to the groups School Total, Asian/Pacific Islander, Black, Hispanic, White, Free and Reduced Lunch, Individualized Education Program, and Limited English Proficiency. Then, the eight spreadsheets, all with corresponding titles, contained MAP scale scores from the 2008 and 2009 state tests. Now, since each of the four databases contained eight spreadsheets, there existed 32 eligible populations. Next, Research Randomizer generated random numbers for each of the 32 eligible populations and ultimately, random sample populations formed from each of the 32 eligible populations.

Data Analysis

In this quantitative study, the goal was to determine the effect of Study Island on student achievement by determining whether a statistically significant difference existed between 2008 and 2009 MAP average scale scores. This determination occurred for the total population and seven subgroups contained in each of the four databases of elementary communication arts, elementary mathematics, middle school communication arts, and middle school mathematics. The 32 random samples were evaluated using *z* tests or *t* tests, with a 95% confidence interval.

The statistical analysis of the 32 random sample populations occurred within the Excel databases using the Data Analysis features Descriptive Statistics, *z* test for Two Sample Means, *F* Test for Two Sample Variances, *t* Test for Two Sample Means Assuming Equal Variances, and *t* Test for Two Sample Means Assuming Unequal Variances. The Descriptive Statistics of each actual sample supplied data such as the mean, median, mode, and standard deviation. Next, a *z* test for Two Sample Means evaluated random sample populations that included 50 MAP scale scores. However, not all of the actual sample populations included 50 MAP scale scores. For these instances, a *t* Test for Two Sample Means Assuming Equal Variances or a *t* Test for Two Sample Means Assuming Unequal Variances evaluated random sample populations that included 15 MAP scale scores. As expected, prior to using a *t* test, the *F* Test for Two Sample Variances determined whether the assumed variance was equal or unequal. The null hypothesis stated that no significant difference existed between the 2008 and 2009 MAP average scale scores.

Prior to the submission of this study, 2010 MAP data became available and the overall results in communication arts and mathematics were included in chapter 4 however, none of the 2010 MAP scale scores were statistically analyzed since it was not part of the approved study. In addition, the district in this study continued using Study Island for a second consecutive year during 2009-2010. Therefore, Study Island usage data, such as time spent using Study Island and percent of questions answered correctly, were included for both 2008-2009 and 2009-2010 in chapter 4.

Conclusion

This study occurred within a large school district located in St. Charles, Missouri. The district included 10 elementary and 5 middle schools and during the 2008-2009 school year students began using Study Island. For perspective, demographics of the school district represent a population that is 88.8% White, 6.1% Black, Asian 2.6%, and Hispanic 2.2%, with 13.4% of the population eligible for free and reduced lunch.

This quantitative study investigated the effect of Study Island on student achievement, as measured by MAP scale scores. Therefore, the analysis of MAP scale scores from 2008 and 2009 determined whether a statistically significant difference existed in terms of student achievement. The null hypothesis stated that no significant difference existed between the 2008 and 2009 MAP average scale scores.

The demographic characteristics of the population and one year of Study Island use limited the reliability of the results. The study provided a thorough analysis of data from the total population and seven subgroups within the areas of elementary communication arts, elementary mathematics, middle school communication arts, and middle school mathematics. Chapter 4 provides discussion and results from the 32 descriptive statistics, z tests, and t tests performed on each level, subject, and group.

Chapter Four: Results

Overview

No Child Left Behind set the expectation that all public schools meet Adequate Yearly Progress on an annual basis. In order for a school to meet Adequate Yearly Progress, student achievement must meet the annual performance targets set at the state level. Schools that do not meet the annual performance targets in the same subject area, such as communication arts or mathematics, for two consecutive years face sanctions. Consequently, this system of accountability and sanctions has prompted school districts to invest time and money into assessment programs to help them meet annual performance targets. For example, the school district in this study chose to begin using an assessment program called Study Island during the 2008-2009 school year for the purpose of increasing student achievement and meeting Adequate Yearly Progress.

This was a quantitative study focused on determining whether a statistically significant difference existed between 2008 MAP average scale scores, the year prior to the school district using Study Island, and 2009 MAP average scale scores, the year the school district began using Study Island. Within the district in this study, data supported that there was a statistically significant difference between 2008 MAP average scale scores and 2009 MAP average scale scores. Moreover, the researcher attributed the significant difference in MAP average scale scores to the use of Study Island. School officials may use the results of this research to understand the relationship between Study Island and student achievement.

Data Analysis

To evaluate the null hypotheses, it was necessary to use MAP scale score data from 2008 and 2009 at the elementary and middle school levels for communication arts and mathematics. I disaggregated the MAP scale scores by level, subject, and subgroup to promote an in-depth analysis. Descriptive statistics for each level and subject provided data on the total population, School Total, and the seven subgroup populations of Asian/Pacific Islander, Black, Hispanic, White, Free and Reduced Lunch, Individualized Education Program, and Limited English Proficiency. A 95% confidence interval accompanied the *z* tests and *t* tests for the difference of means on random samples from each of the eight groups. Most of the random samples from the total population and seven subgroups included 50 scale scores; however, for smaller subgroups, 15 scale scores comprised the random samples.

Elementary communication arts. The null hypothesis stated that there was no significant difference between the 2008 MAP communication arts average scale scores and the 2009 MAP communication arts average scale scores at the elementary level. I rejected the null hypothesis for the total population and the seven subgroups at the elementary level in communication arts because the *z* values of all of the randomly sampled groups fell into the critical regions on a bell curve. Therefore, the data supported the hypothesis that a significant difference existed between 2008 and 2009 MAP average scale scores. I used a *t* test instead of a *z* test to evaluate the Hispanic and Limited English Proficiency subgroups because fewer than 30 MAP scale scores comprised the random samples. The following sections contain the specific results of the descriptive statistics along with the results of the *z* tests and *t* tests for each group.

School total. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the School Total population and the random sample taken from the School Total population. Table 9 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the School Total population. The descriptive statistics data for 2008 and 2009 is side by side in Table 9 to show how the random sample compared to the entire School Total population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire School Total population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 10 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 9
Descriptive Statistics for Elementary Communication Arts School Total

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	652.3	668.7	653.7	672.1
Median	658	673	655	674
Mode	663	675	656	672
Standard Deviation	38.3	27.4	33.3	30.3
Variance	1,470.0	751.2	1,109.6	920.4
Minimum	470	601	470	549
Maximum	721	724	774	840
Sum	31,963	32,771	1,560,386	1,604,470
Count (n-1)	49	49	2,387	2,387

Table 10
Quantitative Analysis for Elementary Communication Arts School Total

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	2.56
Alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Asian/Pacific Islander. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Asian/Pacific Islander population and the random sample taken from the Asian/Pacific Islander population. Table 11 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Asian/Pacific Islander population. The descriptive statistics data for 2008 and 2009 is side by side in Table 11 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 12 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 11
Descriptive Statistics for Elementary Communication Arts Asian/Pacific Islander

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	653.2	680.1	656.9	680.5
Median	655	682	656	683
Mode	660	695	660	695
Standard Deviation	31.4	28.2	33.7	27.4
Variance	990.6	797.3	1,138.1	755.2
Minimum	575	616	575	616
Maximum	728	748	774	748
Sum	32,008	33,327	40,072	41,514
Count (n-1)	49	49	61	61

Table 12
Quantitative Analysis for Elementary Communication Arts Asian/Pacific Islander

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	4.45
Alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Black. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Black population and the random sample taken from the Black population. Table 13 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Black population. The descriptive statistics data for 2008 and 2009 is side by side in Table 13 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 14 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 13
Descriptive Statistics for Elementary Communication Arts Black

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	631.0	656.4	641.6	659.4
Median	634	661	643	662
Mode	643	660	629	660
Standard Deviation	37.9	35.3	34.9	32.5
Variance	1,436.8	1,246.6	1,223.4	1,061.5
Minimum	538	584	530	565
Maximum	737	748	742	748
Sum	30,919	32,167	87,263	89,690
Count (n-1)	49	49	136	136

Table 14
Quantitative Analysis for Elementary Communication Arts Black

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	3.44
Alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Hispanic. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Hispanic population and the random sample taken from the Hispanic population. Table 15 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Hispanic population. The descriptive statistics data for 2008 and 2009 is side by side in Table 15 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a *t* test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. A *t* test was used on the random sample because it contained fewer than 30 MAP scale scores. Prior to performing the *t* test, I conducted an *F* test to determine whether the two sample variances were statistically equal or not. The results of the *F* test indicated that the variances were equal because the *F* test value fell inside the critical value. Therefore, I used a *t* test assuming equal variances on the random sample for this subgroup. Table 16 shows the results of the *t* test for the random sample. Since the *t* test value was larger than the critical value of 2.05, the *t* value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between

the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 15
Descriptive Statistics for Elementary Communication Arts Hispanic

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	644.8	669.0	643.0	660.6
Median	647	675	645	664
Mode	#N/A	657	648	644
Standard Deviation	25.7	26.9	31.6	29.4
Variance	661.2	724.6	1,001.9	867.4
Minimum	587	600	555	600
Maximum	682	708	700	738
Sum	9,028	9,366	27,653	28,407
Count (n-1)	14	14	43	43

Table 16
Quantitative Analysis for Elementary Communication Arts Hispanic

Statistical Test	Entire Population
Hypothesized Mean Difference	0
t	2.42
Alpha	0.05
T Critical two-tail	2.05
Confidence Interval	95%

White. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the White population and the random sample taken from the White population. Table 17 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire White population. The descriptive statistics data for 2008 and 2009 is side by side in Table 17 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 18 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 17
Descriptive Statistics for Elementary Communication Arts White

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	649.7	671.6	654.5	672.9
Median	649	671	656	675
Mode	674	647	668	672
Standard Deviation	29.6	25.6	33.0	30.0
Variance	881.4	660.2	1,093.5	903.1
Minimum	563	609	470	549
Maximum	707	727	774	840
Sum	31,838	32,912	1,402,805	1,442,188
Count (n-1)	49	49	2143	2143

Table 18
Quantitative Analysis for Elementary Communication Arts White

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	3.90
Alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Free and reduced lunch. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Free and Reduced Lunch population and the random sample taken from the Free and Reduced Lunch population. Table 19 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Free and Reduced Lunch population. The descriptive statistics data for 2008 and 2009 is side by side in Table 19 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 Missouri Assessment Program scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 20 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 19
Descriptive Statistics for Elementary Communication Arts Free and Reduced Lunch

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	638.6	654.5	640.0	658.3
Median	636	655	642	662
Mode	625	644	629	680
Standard Deviation	32.9	32.2	35.8	29.6
Variance	1,085.6	1,040.6	1,288.4	881.6
Minimum	562	591	470	565
Maximum	711	719	774	724
Sum	31,296	32,073	231,053	237,677
Count (n-1)	49	49	361	361

Table 20
Quantitative Analysis for Elementary Communication Arts Free and Reduced Lunch

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	2.40
Alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Individualized education program. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Individualized Education Program population and the random sample taken from the Individualized Education Program population. Table 21 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Individualized Education Program population. The descriptive statistics data for 2008 and 2009 is side by side in Table 21 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 22 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 21

Descriptive Statistics for Elementary Comm. Arts Individualized Education Program

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	631.4	647.7	626.0	648.8
Median	636	649	629	649.5
Mode	593	626	629	675
Standard Deviation	33.6	27.2	39.3	35.3
Variance	1,134.0	741.1	1,545.1	1,252.5
Minimum	547	590	470	561
Maximum	689	699	735	820
Sum	30,940	31,740	196,564	203,754
Count (n-1)	49	49	314	314

Table 22

Quantitative Analysis for Elementary Comm. Arts Individualized Education Program

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	2.63
Alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Limited English proficiency. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Limited English Proficiency population and the random sample taken from the Limited English Proficiency population. Table 23 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Limited English Proficiency population. The descriptive statistics data for 2008 and 2009 is side by side in table 23 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a *t* test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. A *t* test was used on the random sample because it contained fewer than 30 MAP scale scores. Prior to performing the *t* test, I conducted an *F* test to determine whether the two sample variances were statistically equal or not. The results of the *F* test indicated that the variances were equal because the *F* test value fell inside the critical value. Therefore, I used a *t* test assuming equal variances on the random sample for this subgroup. Table 24 shows the results of the *t* test for the random sample. Since the *t* test value was larger than the critical value of 2.05, the *t* value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 23
Descriptive Statistics for Elementary Communication Arts Limited English Proficiency

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	616.9	648.7	623.9	647.1
Median	617	646.5	629.5	645.5
Mode	616	#N/A	616	643
Standard Deviation	34.6	32.2	30.8	31.5
Variance	1,197.9	1,037.5	949.7	996.9
Minimum	555	601	555	600
Maximum	665	697	679	708
Sum	8,637	9,083	16,222	16,825
Count (n-1)	14	14	26	26

Table 24
Quantitative Analysis for Elementary Communication Arts Limited English Proficiency

Statistical Test	Entire Population
Hypothesized Mean Difference	0
t	2.52
Alpha	0.05
T Critical two-tail	2.05
Confidence Interval	95%

Elementary mathematics. The results in elementary mathematics perfectly mirrored those found in elementary communication arts. The null hypothesis stated that there was no significant difference between the 2008 MAP mathematics average scale scores and the 2009 MAP mathematics average scale scores at the elementary level. I rejected the null hypothesis for the total population and the seven subgroups at the elementary level in mathematics because the *z* values of all of the randomly sampled groups fell into the critical regions on a bell curve. Therefore, the data supported the hypothesis that a significant difference existed between 2008 and 2009 MAP average scale scores. I used a *t* test instead of a *z* test to evaluate the Hispanic and Limited English Proficiency subgroups because fewer than 30 MAP scale scores comprised the random samples. The following sections contain the specific results of the descriptive statistics along with the results of the *z* tests and *t* tests for each group.

School total. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the School Total population and the random sample taken from the School Total population. Table 25 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the School Total population. The descriptive statistics data for 2008 and 2009 is side by side in Table 25 to show how the random sample compared to the entire School Total population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire School Total population, I used a *z* test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 26 shows the results of the *z* test for the random sample. Since the *z* test value was larger than the critical value of 1.95, the *z* value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 25
Descriptive Statistics for Elementary Mathematics School Total

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	637.6	663.1	640.0	666.1
Median	645	667	641	665
Mode	631	638	638	666
Standard Deviation	40.6	34.7	32.2	35.8
Variance	1,651.1	1,210.9	1,042.1	1,284.5
Minimum	469	564	450	465
Maximum	712	754	805	830
Sum	31,246	32,495	1,529,689	1,592,175
Count (n-1)	49	49	2,390	2,390

Table 26
Quantitative Analysis for Elementary Mathematics School Total

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	3.33
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Asian/Pacific Islander. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Asian/Pacific Islander population and the random sample taken from the Asian/Pacific Islander population. Table 27 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Asian/Pacific Islander population. The descriptive statistics data for 2008 and 2009 is side by side in Table 27 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 28 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 27

Descriptive Statistics for Elementary Mathematics Asian/Pacific Islander

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	656.3	677.3	654.2	675.2
Median	652	676	651	673
Mode	626	673	626	673
Standard Deviation	31.2	31.6	35.4	32.4
Variance	977.2	999.1	1,257.2	1,050.1
Minimum	599	616	599	616
Maximum	735	765	780	765
Sum	32,161	33,192	39,908	41,192
Count (n-1)	49	49	61	61

Table 28

Quantitative Analysis for Elementary Mathematics Asian/Pacific Islander

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	3.31
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Black. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Black population and the random sample taken from the Black population. Table 29 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Black population. The descriptive statistics data for 2008 and 2009 is side by side in Table 29 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 30 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 29
Descriptive Statistics for Elementary Mathematics Black

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	625.4	652.9	623.4	646.4
Median	627	654	626	650
Mode	616	605	616	666
Standard Deviation	25.0	28.8	30.2	35.5
Variance	628.8	832.6	912.9	1,264.1
Minimum	568	587	556	520
Maximum	669	701	714	773
Sum	30,649	31,995	84,788	87,920
Count (n-1)	49	49	136	136

Table 30
Quantitative Analysis for Elementary Mathematics Black

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	5.02
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Hispanic. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Hispanic population and the random sample taken from the Hispanic population. Table 31 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Hispanic population. The descriptive statistics data for 2008 and 2009 is side by side in Table 31 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a *t* test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. A *t* test was used on the random sample because it contained fewer than 30 MAP scale scores. Prior to performing the *t* test, I conducted an *F* test to determine whether the two sample variances were statistically equal or not. The results of the *F* test indicated that the variances were equal because the *F* test value fell inside the critical value. Therefore, I used a *t* test assuming equal variances on the random sample for this subgroup. Table 32 shows the results of the *t* test for the random sample. Since the *t* test value was larger than the critical value of 2.05, the *t* value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 31

Descriptive Statistics for Elementary Mathematics Hispanic

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	623.7	652.2	624.4	651.7
Median	618	646.5	621	650.5
Mode	621	625	615	649
Standard Deviation	36.1	35.6	29.7	26.8
Variance	1,308.0	1,274.1	882.5	720.1
Minimum	559	585	556	585
Maximum	678	729	681	729
Sum	8,732	9,131	27,477	28,679
Count (n-1)	14	14	44	44

Table 32

Quantitative Analysis for Elementary Mathematics Hispanic

Statistical Test	Entire Population
Hypothesized Mean Difference	0
t	2.09
alpha	0.05
T Critical two-tail	2.05
Confidence Interval	95%

White. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the White population and the random sample taken from the White population. Table 33 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire White population. The descriptive statistics data for 2008 and 2009 is side by side in Table 33 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a *z* test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 34 shows the results of the *z* test for the random sample. Since the *z* test value was larger than the critical value of 1.95, the *z* value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 33

Descriptive Statistics for Elementary Mathematics White

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	639.3	672.0	641.0	667.4
Median	643	678	642	666
Mode	662	684	638	663
Standard Deviation	28.9	31.2	31.9	35.6
Variance	837.8	978.7	1,020.0	1,271.3
Minimum	563	603	450	465
Maximum	699	737	805	830
Sum	31,328	32,930	1,374,945	1,431,722
Count (n-1)	49	49	2145	2145

Table 34

Quantitative Analysis for Elementary Mathematics White

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	5.36
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Free and reduced lunch. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Free and Reduced Lunch population and the random sample taken from the Free and Reduced Lunch population. Table 35 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Free and Reduced Lunch population. The descriptive statistics data for 2008 and 2009 is side by side in Table 35 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 36 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 35

Descriptive Statistics for Elementary Mathematics Free and Reduced Lunch

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	631.0	655.9	626.3	651.1
Median	630	659	627	650
Mode	616	679	616	663
Standard Deviation	36.6	34.8	33.1	37.4
Variance	1,339.8	1,216.3	1,096.8	1,399.8
Minimum	556	568	450	465
Maximum	780	752	780	752
Sum	30,923	32,143	226,732	235,703
Count (n-1)	49	49	362	362

Table 36

Quantitative Analysis for Elementary Mathematics Free and Reduced Lunch

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	2.40
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Individualized education program. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Individualized Education Program population and the random sample taken from the Individualized Education Program population. Table 37 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Individualized Education Program population. The descriptive statistics data for 2008 and 2009 is side by side in Table 37 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 38 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 37
Descriptive Statistics for Elementary Mathematics Individualized Education Program

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	628.0	650.3	618.7	642.8
Median	627	655	619	643
Mode	621	662	621	660
Standard Deviation	33.8	41.2	35.4	39.6
Variance	1,147.8	1,698.9	1,256.3	1,574.0
Minimum	558	568	450	465
Maximum	696	747	714	779
Sum	30,774	31,865	194,899	202,505
Count (n-1)	49	49	315	315

Table 38
Quantitative Analysis for Elementary Mathematics Individualized Education Program

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	2.63
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Limited English proficiency. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Limited English Proficiency population and the random sample taken from the Limited English Proficiency population. Table 39 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Limited English Proficiency population. The descriptive statistics data for 2008 and 2009 is side by side in Table 39 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a *t* test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. A *t* test was used on the random sample because it contained fewer than 30 MAP scale scores. Prior to performing the *t* test, I conducted an *F* test to determine whether the two sample variances were statistically equal or not. The results of the *F* test indicated that the variances were equal because the *F* test value fell inside the critical value. Therefore, I used a *t* test assuming equal variances on the random sample for this subgroup. Table 40 shows the results of the *t* test for the random sample. Since the *t* test value was larger than the critical value of 2.05, the *t* value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 39

Descriptive Statistics for Elementary Mathematics Limited English Proficiency

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	607.9	638.6	612.3	642.1
Median	607.5	634.5	615	640
Mode	#N/A	649	615	649
Standard Deviation	38.5	37.4	30.7	30.7
Variance	1,483.1	1,401.3	942.9	945.0
Minimum	556	585	556	585
Maximum	701	728	701	728
Sum	8,511	8,941	16,533	17,337
Count (n-1)	14	14	27	27

Table 40

Quantitative Analysis for Elementary Mathematics Limited English Proficiency

Statistical Test	Entire Population
Hypothesized Mean Difference	0
t	2.52
alpha	0.05
t Critical two-tail	2.05
Confidence Interval	95%

Middle school communication arts. The results of the middle school communication arts data was the exact opposite of the results found at the elementary level. The null hypothesis stated that there was no significant difference between the 2008 MAP communication arts average scale scores and the 2009 MAP communication arts average scale scores at the middle school level. I did not reject the null hypothesis for the total population and the seven subgroups at the middle school level in communication arts because the *z* values of all of the randomly sampled groups did not fall into the critical regions on a bell curve. Therefore, the data did not support the hypothesis that a significant difference existed between 2008 and 2009 MAP average scale scores. I used a *t* test instead of a *z* test to evaluate the Limited English Proficiency subgroup because fewer than 30 MAP scale scores comprised the random samples. The following sections contain the specific results of the descriptive statistics along with the results of the *z* tests and *t* tests for each group.

School total. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the School Total population and the random sample taken from the School Total population. Table 41 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the School Total population. The descriptive statistics data for 2008 and 2009 is side by side in Table 41 to show how the random sample compared to the entire School Total population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire School Total population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 42 shows the results of the z test for the random sample. Since the z test value was smaller than the critical value of 1.95, the z value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 41
Descriptive Statistics for Middle School Communication Arts School Total

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	680.5	688.4	677.7	686.4
Median	678	685	678	688
Mode	676	685	677	688
Standard Deviation	24.1	30.7	29.5	30.9
Variance	584.5	943.8	870.4	959.2
Minimum	626	635	485	505
Maximum	758	756	808	865
Sum	33,348	33,735	2,504,290	2,536,615
Count (n-1)	49	49	3695	3695

Table 42
Quantitative Analysis for Middle School Communication Arts School Total

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	1.41
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Asian/Pacific Islander. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Asian/Pacific Islander population and the random sample taken from the Asian/Pacific Islander population. Table 43 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the Asian/Pacific Islander population. The descriptive statistics data for 2008 and 2009 is side by side in Table 43 to show how the random sample compared to the entire Asian/Pacific Islander population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire Asian/Pacific Islander population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 44 shows the results of the z test for the random sample. Since the z test value was smaller than the critical value of 1.95, the z value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 43
Descriptive Statistics for Middle School Communication Arts Asian/Pacific Islander

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	676.0	687.1	679.2	689.8
Median	672	691	685	694
Mode	688	694	688	706
Standard Deviation	37.5	36.9	34.1	33.1
Variance	1,410.7	1,364.7	1,165.6	1,099.3
Minimum	565	598	565	598
Maximum	758	798	762	798
Sum	33,124	33,672	64,527	65,536
Count (n-1)	49	49	95	95

Table 44
Quantitative Analysis for Middle School Communication Arts Asian/Pacific Islander

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	1.48
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Black. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Black population and the random sample taken from the Black population. Table 45 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the Black population. The descriptive statistics data for 2008 and 2009 is side by side in Table 45 to show how the random sample compared to the entire Black population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire Black population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 46 shows the results of the z test for the random sample. Since the z test value was smaller than the critical value of 1.95, the z value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 45

Descriptive Statistics for Middle School Communication Arts Black

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	670.6	676.4	661.2	672.2
Median	671	674	663	672
Mode	663	679	665	672
Standard Deviation	28.8	30.9	29.2	28.0
Variance	832.2	955.8	857.9	787.7
Minimum	611	604	505	569
Maximum	744	760	744	760
Sum	32,863	33,148	139,518	141,844
Count (n-1)	49	49	211	211

Table 46

Quantitative Analysis for Middle School Communication Arts Black

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	0.96
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Hispanic. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Hispanic population and the random sample taken from the Hispanic population. Table 47 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the Hispanic population. The descriptive statistics data for 2008 and 2009 is side by side in Table 47 to show how the random sample compared to the entire Hispanic population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire Hispanic population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 48 shows the results of the z test for the random sample. Since the z test value was smaller than the critical value of 1.95, the z value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 47

Descriptive Statistics for Middle School Communication Arts Hispanic

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	664.0	672.7	660.5	669.1
Median	667	672	662	671
Mode	673	648	635	715
Standard Deviation	27.3	27.4	28.3	30.7
Variance	746.1	753.1	801.2	945.5
Minimum	594	628	594	593
Maximum	728	743	728	743
Sum	32,536	32,967	42,277	42,823
Count (n-1)	49	49	64	64

Table 48

Quantitative Analysis for Middle School Communication Arts Hispanic

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	1.59
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

White. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the White population and the random sample taken from the White population. Table 49 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the White population. The descriptive statistics data for 2008 and 2009 is side by side in Table 49 to show how the random sample compared to the entire White population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire White population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 50 shows the results of the z test for the random sample. Since the z test value was smaller than the critical value of 1.95, the z value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 49
Descriptive Statistics for Middle School Communication Arts White

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	681.9	689.8	679.1	687.7
Median	680	689	680	689
Mode	700	703	677	688
Standard Deviation	27.9	26.7	28.8	30.6
Variance	780.2	714.5	831.2	937.1
Minimum	594	634	485	505
Maximum	768	752	808	865
Sum	33,415	33,803	2,250,786	2,279,155
Count (n-1)	49	49	3314	3314

Table 50
Quantitative Analysis for Middle School Communication Arts White

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	1.43
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Free and reduced lunch. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Free and Reduced Lunch population and the random sample taken from the Free and Reduced Lunch population. Table 51 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the Free and Reduced Lunch population. The descriptive statistics data for 2008 and 2009 is side by side in Table 51 to show how the random sample compared to the entire Free and Reduced Lunch population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire Free and Reduced Lunch population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 52 shows the results of the z test for the random sample. Since the z test value was smaller than the critical value of 1.95, the z value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 51

Descriptive Statistics for Middle School Communication Arts Free and Reduced Lunch

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	658.6	662.9	661.2	669.0
Median	658	666	661	671
Mode	655	659	664	675
Standard Deviation	28.7	40.7	29.8	32.7
Variance	824.6	1,657.2	893.5	1,074.2
Minimum	581	530	505	530
Maximum	713	731	762	763
Sum	32,273	32,484	298,224	301,747
Count (n-1)	49	49	451	451

Table 52

Quantitative Analysis for Middle School Communication Arts Free and Reduced Lunch

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	0.60
Alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Individualized education program. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Individualized Education Program population and the random sample taken from the Individualized Education Program population. Table 53 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the Individualized Education Program population. The descriptive statistics data for 2008 and 2009 is side by side in Table 53 to show how the random sample compared to the entire Individualized Education Program population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire Individualized Education Program population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 54 shows the results of the z test for the random sample. Since the z test value was smaller than the critical value of 1.95, the z value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 53

Descriptive Statistics for Middle School Comm. Arts Individualized Education Program

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	642.9	647.9	639.4	647.0
Median	646	646	642	648
Mode	660	621	647	642
Standard Deviation	29.2	27.6	37.3	36.9
Variance	853.2	763.9	1,394.2	1,367.4
Minimum	533	590	485	505
Maximum	707	707	725	731
Sum	31,503	31,749	215,498	218,056
Count (n-1)	49	49	337	337

Table 54

Quantitative Analysis for Middle School Comm. Arts Individualized Education Program

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	0.87
Alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Limited English proficiency. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Limited English Proficiency population and the random sample taken from the Limited English Proficiency population. Table 55 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Limited English Proficiency population. The descriptive statistics data for 2008 and 2009 is side by side in Table 55 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a *t* test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. A *t* test was used on the random sample because it contained fewer than 30 MAP scale scores. Prior to performing the *t* test, I conducted an *F* test to determine whether the two sample variances were statistically equal or not. The results of the *F* test indicated that the variances were equal because the *F* test value fell inside the critical value. Therefore, I used a *t* test assuming equal variances on the random sample for this subgroup. Table 56 shows the results of the *t* test for the random sample. Since the *t* test value was smaller than the critical value of 2.05, the *t* value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 55
Descriptive Statistics for Middle School Communication Arts Limited English Proficiency

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	642.5	654.0	640.9	652.7
Median	637.5	657.5	640.5	651
Mode	#N/A	663	618	651
Standard Deviation	19.7	28.4	20.9	20.6
Variance	388.2	807.6	437.5	427.3
Minimum	618	593	594	593
Maximum	682	706	682	706
Sum	8,995	9,156	24,356	24,806
Count (n-1)	14	14	38	38

Table 56
Quantitative Analysis for Middle School Comm. Arts Limited English Proficiency

Statistical Test	Entire Population
Hypothesized Mean Difference	0
t	1.24
Alpha	0.05
T Critical two-tail	2.05
Confidence Interval	95%

Middle school mathematics. The results of the middle school mathematics data were mixed when compared to elementary mathematics. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores in mathematics at the middle school level. I rejected the null hypothesis for the randomly sampled middle school mathematics subgroups of Asian/Pacific Islander, Hispanic, Free and Reduced Lunch, and Individualized Education Program because the *z* values fell into the critical regions on a bell curve. Therefore, the data supported the hypothesis that a significant difference existed between 2008 and 2009 MAP average scale scores for the subgroups Asian/Pacific Islander, Hispanic, Free and Reduced Lunch, and Individualized Education Program. Conversely, I did not reject the null hypothesis for the randomly sampled middle school mathematics total population, School Total, and the three subgroup populations of Black, White, and Limited English Proficiency because the *z* values never fell into the critical regions on a bell curve. The data supported the hypothesis that no significant difference existed between 2008 and 2009 MAP average scale scores in mathematics at the middle school level for the School Total and subgroups of Black, White, and Limited English Proficiency. A *t* test instead of a *z* test evaluated the Limited English Proficiency subgroup because fewer than 30 MAP scale scores were in the random sample. The following sections contain the specific results of the descriptive statistics along with the results of the *z* tests and *t* tests for each group.

School total. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the School Total population and the random sample taken from the School Total population. Table 57 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the School Total population. The descriptive statistics data for 2008 and 2009 is side by side in Table 57 to show how the random sample compared to the entire School Total population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire School Total population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 58 shows the results of the z test for the random sample. Since the z test value was smaller than the critical value of 1.95, the z value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 57

Descriptive Statistics for Middle School Mathematics School Total

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	685.7	698.6	683.6	700.9
Median	687	699	686	703
Mode	685	695	675	710
Standard Deviation	38.4	39.6	36.5	36.3
Variance	1,475.5	1,569.7	1,338.3	1,324.2
Minimum	544	576	480	495
Maximum	753	766	845	857
Sum	33,604	34,236	2,526,948	2,590,825
Count (n-1)	49	49	3696	3696

Table 58

Quantitative Analysis for Middle School Mathematics School Total

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	1.63
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Asian/Pacific Islander. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Asian/Pacific Islander population and the random sample taken from the Asian/Pacific Islander population. Table 59 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Asian/Pacific Islander population. The descriptive statistics data for 2008 and 2009 is side by side in Table 59 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 60 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 59

Descriptive Statistics for Middle School Mathematics Asian/Pacific Islander

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	688.8	709.1	691.1	711.8
Median	684	705	691	709
Mode	699	702	685	730
Standard Deviation	38.3	38.3	45.4	41.8
Variance	1,473.0	1,473.5	2,068.4	1,748.7
Minimum	595	634	568	623
Maximum	792	798	845	857
Sum	33,756	34,748	67,046	69,052
Count (n-1)	49	49	97	97

Table 60

Quantitative Analysis for Middle School Mathematics Asian/Pacific Islander

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	2.61
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Black. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Black population and the random sample taken from the Black population. Table 61 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the Black population. The descriptive statistics data for 2008 and 2009 is side by side in Table 61 to show how the random sample compared to the entire Black population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire Black population, I used a *z* test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 62 shows the results of the *z* test for the random sample. Since the *z* test value was smaller than the critical value of 1.95, the *z* value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 61
Descriptive Statistics for Middle School Mathematics Black

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	662.0	675.5	658.7	675.4
Median	657	676	658	675
Mode	659	697	679	672
Standard Deviation	42.4	33.0	35.8	32.5
Variance	1,805.7	1,089.2	1,287.6	1,059.1
Minimum	524	595	524	545
Maximum	790	744	790	756
Sum	32,441	33,102	138,989	142,510
Count (n-1)	49	49	211	211

Table 62
Quantitative Analysis for Middle School Mathematics Black

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	1.75
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Hispanic. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Hispanic population and the random sample taken from the Hispanic population. Table 63 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Hispanic population. The descriptive statistics data for 2008 and 2009 is side by side in Table 63 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a *z* test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 64 shows the results of the *z* test for the random sample. Since the *z* test value was larger than the critical value of 1.95, the *z* value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 63

Descriptive Statistics for Middle School Mathematics Hispanic

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	663.4	682.3	660.9	679.4
Median	673	685	669	684
Mode	639	690	639	699
Standard Deviation	36.4	33.6	37.7	35.6
Variance	1,326.2	1,133.6	1,424.8	1,267.6
Minimum	555	612	555	579
Maximum	728	756	728	756
Sum	32,508	33,434	42,960	44,162
Count (n-1)	49	49	65	65

Table 64

Quantitative Analysis for Middle School Mathematics Hispanic

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	2.66
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

White. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the White population and the random sample taken from the White population. Table 65 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the White population. The descriptive statistics data for 2008 and 2009 is side by side in Table 65 to show how the random sample compared to the entire White population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire White population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 66 shows the results of the z test for the random sample. Since the z test value was smaller than the critical value of 1.95, the z value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 65
Descriptive Statistics for Middle School Mathematics White

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	684.5	697.7	685.5	702.7
Median	685	705	687	705
Mode	715	701	675	710
Standard Deviation	34.3	46.1	35.5	35.6
Variance	1,178.2	2,126.3	1,265.1	1,269.5
Minimum	562	495	480	495
Maximum	739	757	830	857
Sum	33,543	34,192	2,271,831	2,328,825
Count (n-1)	49	49	3314	3314

Table 66
Quantitative Analysis for Middle School Mathematics White

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	1.61
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Free and reduced lunch. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Free and Reduced Lunch population and the random sample taken from the Free and Reduced Lunch population. Table 67 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Free and Reduced Lunch population. The descriptive statistics data for 2008 and 2009 is side by side in Table 67 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 68 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 67

Descriptive Statistics for Middle School Mathematics Free and Reduced Lunch

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	657.4	676.4	661.8	679.0
Median	659	671	665	679
Mode	679	661	671	695
Standard Deviation	40.2	37.6	37.2	37.7
Variance	1,623.1	1,417.0	1,390.7	1,422.4
Minimum	559	596	495	545
Maximum	747	770	768	785
Sum	32,217	33,145	297,164	304,883
Count (n-1)	49	49	449	449

Table 68

Quantitative Analysis for Middle School Mathematics Free and Reduced Lunch

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	2.40
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Individualized education program. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Individualized Education Program population and the random sample taken from the Individualized Education Program population. Table 69 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Individualized Education Program population. The descriptive statistics data for 2008 and 2009 is side by side in Table 69 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a z test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. The null hypothesis stated that there was no significant difference between the 2008 MAP average scale scores and the 2009 MAP average scale scores. Table 70 shows the results of the z test for the random sample. Since the z test value was larger than the critical value of 1.95, the z value fell into the critical regions on a bell curve and, thus, I rejected the null hypothesis. I supported the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 69
Descriptive Statistics for Middle School Mathematics Individualized Education Program

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	644.1	660.9	642.2	659.5
Median	650	659	644	659
Mode	662	665	631	659
Standard Deviation	39.7	37.0	43.7	42.2
Variance	1,577.0	1,372.9	1,918.1	1,788.5
Minimum	559	579	480	495
Maximum	735	759	768	785
Sum	31,564	32,387	215,784	221,621
Count (n-1)	49	49	336	336

Table 70
Quantitative Analysis for Middle School Mathematics Individualized Education Program

Statistical Test	Random Sample
Hypothesized Mean Difference	0
z	2.16
alpha	0.05
Z Critical two-tail	1.95
Confidence Interval	95%

Limited English proficiency. The first step in analyzing whether the difference between 2008 MAP scale scores and 2009 MAP scale scores was statistically significant was to find the descriptive statistics of these two sets of data for both the Limited English Proficiency population and the random sample taken from the Limited English Proficiency population. Table 71 shows the statistics for both of these groups from 2008 and 2009. The 2008 Sample and 2009 Sample columns include data on the random samples and the 2008 Population and 2009 Population columns include data on the entire Limited English Proficiency population. The descriptive statistics data for 2008 and 2009 is side by side in Table 71 to show how the random sample compared to the entire subgroup population.

After finding the descriptive statistics for 2008 and 2009 MAP scale scores for the random sample and entire subgroup population, I used a *t* test for the difference of means using the hypothesized mean difference of zero to evaluate the random sample. A *t* test was used on the random sample because it contained fewer than 30 MAP scale scores. Prior to performing the *t* test, I conducted an *F* test to determine whether the two sample variances were statistically equal or not. The results of the *F* test indicated that the variances were equal because the *F* test value fell inside the critical value. Therefore, I used a *t* test assuming equal variances on the random sample for this subgroup. Table 72 shows the results of the *t* test for the random sample. Since the *t* test value was smaller than the critical value of 2.05, the *t* value did not fall into the critical regions on a bell curve and, thus, I did not reject the null hypothesis. I did not support the hypothesis that a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores for the random sample.

Table 71
Descriptive Statistics for Middle School Mathematics Limited English Proficiency

Description	2008 Sample	2009 Sample	2008 Population	2009 Population
Mean	649.0	667.5	643.9	668.7
Median	644.5	658	643	665.5
Mode	#N/A	#N/A	648	684
Standard Deviation	34.9	36.6	36.1	32.2
Variance	1,220.0	1,345.1	1,303.3	1,043.1
Minimum	596	616	555	612
Maximum	722	737	722	743
Sum	9,086	9,345	25,758	26,750
Count (n-1)	14	14	40	40

Table 72
Quantitative Analysis for Middle School Mathematics Limited English Proficiency

Statistical Test	Entire Population
Hypothesized Mean Difference	0
t	1.36
alpha	0.05
t Critical two-tail	2.05
Confidence Interval	95%

Study Island Data

The degree to which a program is used can often determine how effective it will be in achieving the desired results. This section provides data on how often Study Island was used across the district and by each school. Up to this point in chapter 4, all data and statistics presented in this study pertained to the 2007-2008 and 2008-2009 school years. However, by the time I wrote this chapter, data from 2009-2010 was also available. Therefore, I included 2009-2010 data because it was the second consecutive year the district used Study Island. In fact, during the 2009-2010 school year, Study Island usage was further expanded to include the high school level.

Number of sessions and time spent using Study Island. Study Island allowed users to monitor the number of sessions completed and the time spent using the resource. Data points were available on a district, building, classroom, and student level, as well as disaggregated by subject and grade level. The two categories discussed in this section were dependent on the decisions that administrators and teachers made in terms of how students would use Study Island.

Prior to using Study Island, each school submitted a plan detailing how they planned to use Study Island to meet the needs of their school. The needs of the schools varied and, therefore, the plans they submitted also varied. Similarly, another factor that differed widely amongst the schools in this study was student enrollment; however, many features of the school, such as the number of computer labs, are the same within each school. Unfortunately, the district was unable to provide in-depth Study Island training for teachers and, therefore, did not mandate or set specific expectations regarding usage. Tables 73 and 74 display the total number of sessions completed and the time spent using

Study Island in communication arts and mathematics from August through February during 2008-2009 and 2009-2010 across the district.

I chose to display the months of August through February in Tables 73 and 74, as well as other tables in this chapter, because the usage during these months will have the greatest impact on preparing students for the MAP state tests. Data from March, April, and May was not included because the MAP testing window begins in late March and continues throughout the month of April. By May, the state testing has concluded. In addition, during the month of March the elementary schools in this study have a three-week break, and the middle schools have a one-week break. For these reasons, I believe Study Island usage from March through May does not have the potential effect on student achievement that it would when used from August through February.

Table 73

District Comm. Arts Total Number of Sessions and Time Spent Using Study Island

Month	2008-09 Sessions	2009-10 Sessions	2008-09 Hours	2009-10 Hours
August	0	11,615	0	882
September	10,092	21,283	787	1,701
October	13,094	18,943	1,069	1,561
November	23,386	28,101	1,783	2,333
December	16,857	15,501	1,281	1,288
January	23,695	23,127	1,970	2,002
February	40,935	26,328	3,338	1,968

Table 74
District Mathematics Total Number of Sessions and Time Spent Using Study Island

Month	2008-09 Sessions	2009-10 Sessions	2008-09 Hours	2009-10 Hours
August	382	23,299	20	1,543
September	6,134	33,547	365	2,384
October	14,516	34,734	873	2,510
November	32,428	42,801	2,052	2,983
December	16,355	21,314	1,029	1,533
January	28,520	37,122	1,914	2,512
February	46,940	37,082	2,982	2,406

It is evident from Tables 73 and 74 that the number of sessions and time spent on Study Island across the district varied by month as well as by subject. Moreover, it should come as no surprise that similar differences in these two factors existed from school to school across the district according to how they used Study Island. To illustrate this point, I have included comparative data from the individual elementary and middle schools in the study for the month of February 2009. I chose the month of February because it is the last full month of instruction prior to the MAP state tests and likewise, as a school official in this district, it has been my experience that schools view February as the final month to intervene with students to effect their performance on the state tests. Therefore, teachers have students use Study Island more often during this month, as the data confirms. In addition, February is the month that Study Island developers suggest giving students the final benchmark assessment that predicts the achievement level of a student. Table 75 depicts the total sessions and time spent using Study Island during the month of February in 2008-2009 and 2009-2010 across all subject areas.

Table 75
February Study Island Total Number of Sessions and Time Spent

School	2008-09 Enrollment	2009-10 Enrollment	2008-09 Sessions	2009-10 Sessions	2008-09 Hours	2009-10 Hours
Elementary 1	942	993	7,823	4,702	518	292
Elementary 2	951	936	5,625	5,440	397	470
Elementary 3	925	891	2,783	11,076	162	695
Elementary 4	414	392	11,491	5,837	742	355
Elementary 5	911	912	2,354	5,593	160	380
Elementary 6	524	521	5,207	6,150	254	335
Elementary 7	711	679	4,000	1,520	256	111
Elementary 8	750	734	11,539	5,579	868	378
Elementary 9	813	843	2,072	4,394	132	322
Elementary 10	758	732	8,305	5,414	567	367
Middle 1	837	789	12,884	376	1,212	22
Middle 2	707	705	11,224	5,923	853	451
Middle 3	801	829	767	514	42	31
Middle 4	766	770	2,082	1,404	155	112
Middle 5	900	932	6,024	3,967	348	252

Although variables such as student enrollment differ from school to school, it does not appear that enrollment determines the number of sessions or hours logged by a school. For example, when looking at the 2008-09 Sessions column and the 2008-09

Hours column in Table 75 the reader can see that Elementary 4 and Elementary 8 have 11,491 and 11,539 sessions and 742 and 868 hours respectively. However, the 2008-09 Enrollment column numbers in Table 75 for Elementary 4 and Elementary 8 are 414 and 750 respectively. Even when attempting to use Table 75 to compare Elementary 4 and Elementary 8 to schools with similar enrollments it is tough to identify similarities. This example provides further support for the claim that use of Study Island varied drastically by school.

Study Island questions attempted and percent correct. Study Island allowed users to monitor the number of questions completed as well as the number and percent answered correctly. The data points were available on a district, building, classroom, and student level and were disaggregated by subject and grade level. As expected, the frequency with which students used Study Island effected both the number of questions completed as well as the number and percent answered correctly. Tables 76 and 77 display the number of questions correct, the total number completed, and the percent correct in communication arts and mathematics from August through February during 2008-2009 and 2009-2010 across the district.

Table 76

District Communication Arts Study Island Questions Correct, Total, and Percent

Month	2008-09 Correct	2008-09 Total	2008-09 Percent	2009-10 Correct	2009-10 Total	2009-10 Percent
August	0	0	0%	66,769	97,479	68.5%
September	57,157	86,431	66.1%	131,330	191,933	68.4%
October	79,185	118,514	66.8%	119,150	172,943	68.9%
November	135,622	201,817	67.2%	164,702	236,694	69.6%
December	98,929	144,738	68.4%	96,404	134,301	71.8%
January	146,379	210,373	69.6%	136,264	192,107	70.9%
February	255,341	365,777	69.8%	152,882	208,979	73.2%

Table 77

District Mathematics Study Island Questions Correct, Total, and Percent

Month	2008-09 Correct	2008-09 Total	2008-09 Percent	2009-10 Correct	2009-10 Total	2009-10 Percent
August	2,127	3,060	69.5%	146,060	196,648	74.3%
September	26,200	39,774	65.9%	204,782	281,697	72.7%
October	75,285	106,504	70.7%	200,436	284,725	70.4%
November	191,739	261,661	73.3%	255,654	348,314	73.4%
December	91,891	123,399	74.5%	124,676	169,195	73.7%
January	179,818	242,908	74.0%	208,998	280,143	74.6%
February	284,431	375,798	75.7%	203,903	265,915	76.7%

Similar discrepancies of these variables existed between schools across the district. To illustrate these differences, I have included comparative data from the individual elementary and middle schools for the month of February 2009. Table 78 depicts the number of questions correct, total number attempted, and percent correct during the month of February in 2008-2009 and February 2009-2010 across all subject areas.

Table 78
February Study Island Questions Correct, Total, and Percent

School	2008-09 Correct	2008-9 Total	2008-09 Percent	2009-10 Correct	2009-10 Total	2009-10 Percent
Elementary 1	46,117	60,686	76.0%	26,719	35,718	74.8%
Elementary 2	40,748	54,420	74.9%	36,906	48,901	75.5%
Elementary 3	16,098	21,976	73.3%	59,238	78,799	75.2%
Elementary 4	81,377	104,861	77.6%	36,464	46,570	78.3%
Elementary 5	13,184	18,819	70.1%	24,766	35,170	70.4%
Elementary 6	25,851	34,591	74.7%	32,609	43,492	75.0%
Elementary 7	18,237	24,631	74.0%	8,805	12,093	72.8%
Elementary 8	66,242	87,832	75.4%	31,569	39,937	79.0%
Elementary 9	11,379	14,846	76.6%	21,944	28,606	76.7%
Elementary 10	44,030	60,392	72.9%	32,829	44,256	74.2%
Middle 1	90,121	133,892	67.3%	1,327	2,350	56.5%
Middle 2	82,055	117,541	69.8%	39,087	53,588	72.9%
Middle 3	3,662	5,621	65.1%	3,000	4,282	70.1%
Middle 4	12,971	18,027	72.0%	8,013	11,888	67.4%
Middle 5	32,289	46,777	69.0%	24,266	31,916	76.0%

As previously discussed, variables such as student enrollment differ by school; however, these differences do not account for the vast inconsistencies displayed in Table 78. The data in Table 78 illustrates how Study Island usage differed between schools during one month in 2008-2009, the year the district began using Study Island. The same data from the same month in the 2009-2010 school year was available, and was included for comparison purposes. Table 78 shows that even during the second year of Study Island usage (2009-2010), differences across the district continued to exist. Furthermore, the data suggests that some of the individual schools adjusted their usage from one year to the next.

District Achievement Data

Since the district in this study began using Study Island to increase Adequate Yearly Progress achievement, I thought it would be appropriate to include this data. When reviewing Adequate Yearly Progress data from this school district over the past several years, I noticed several trends. These trends, absent of statistical analysis, are provided in overall district achievement at the elementary and middle school levels for communication arts and mathematics, both before and after the use of Study Island.

Beginning in 2008-2009, the year the district began using Study Island, the district had two consecutive years of making gains in all eight of their communication arts Adequate Yearly Progress groups. Remarkably, this is something that had never happened in the district since they began measuring Adequate Yearly Progress in 2002. Moreover, similar Adequate Yearly Progress group achievement improvements occurred in mathematics. The district increased achievement in seven out of eight mathematics groups for two consecutive years after implementing Study Island. Regardless of whether the increases are statistically significant, this upward trend is noteworthy considering it coincided with the use of Study Island.

To further illustrate these trends, I have included Tables 79 and 80, which show the percent of students scoring proficient in communication arts and mathematics in all eight Adequate Yearly Progress groups from 2006-2010. Although similar trends were evident in the data from 2002-2005, I did not include that data in the tables because it cannot be compared to data from 2006-2010. In 2006, the MAP tests in both communication arts and mathematics changed from grade-span tests to grade-level tests as required by the federal government as part of No Child Left Behind. In addition to this change, Missouri state officials adjusted the number of possible achievement levels from five to four, thereby making it difficult to compare data from 2002-2005 with data from 2006 and beyond.

Table 79
District Communication Arts Adequate Yearly Progress Percent Proficient

Subgroup	2006	2007	2008	2009	2010
School Total	51.8	52.7	53	60.7	65.7
Asian	56.6	54.2	53.5	66.8	70.6
Black	28.0	27.9	33.8	39.3	43.9
Hispanic	32.7	38.5	33.2	46.9	47.6
White	53.3	54.4	54.6	62.3	67.7
FRL	28.8	32.8	32.2	39.5	42.9
IEP	18.5	20.3	21.9	30.2	34.8
LEP	13.6	8.5	3.6	17.2	23.1

Table 80
District Mathematics Adequate Yearly Progress Percent Proficient

Subgroup	2006	2007	2008	2009	2010
School Total	54.9	56.3	58	60.5	66.3
Asian	65.4	69.9	69.8	67.8	74.3
Black	27.4	28.3	30.4	36.9	40.6
Hispanic	41.3	39.2	37.4	45.6	51.6
White	56.5	58	59.9	62.2	68.5
FRL	33.3	32.9	33.5	39.1	44.6
IEP	22.8	25.7	27.4	34.1	36.4
LEP	20.0	18.1	13.8	26.6	36.2

Conclusion

This was a quantitative study conducted to determine whether a relationship existed between the use of Study Island and student achievement. The null hypothesis stated that there was no significant difference in the 2008 MAP average scale scores and the 2009 MAP average scale scores. I used the 2008 and 2009 MAP scale scores because the students in this study used Study Island during 2008-2009 but not during 2007-2008. To allow for an in-depth study, the disaggregation of MAP scale scores occurred which created the four areas of elementary communication arts, elementary mathematics, middle school communication arts, and middle school mathematics. I performed statistical tests on random samples for each of these four areas had statistical tests performed on random samples taken from the total population and seven subgroup populations the district in the study was accountable for on their Adequate Yearly

Progress. The groups included School Total, Asian/Pacific Islander, Black, Hispanic, White, Free and Reduced Lunch, Individualized Education Program, and Limited English Proficiency. Altogether, 32 statistical tests were performed for this quantitative study that consisted of 26 z tests and 6 t tests on random samples using a 95% confidence interval to find the difference between the sample means.

The results of the 16 statistical tests performed at the elementary level on communication arts and mathematics random samples from the total population and seven subgroup populations produced identical results. The results of all the statistical tests led to rejecting the null hypothesis that there was no significant difference in the 2008 MAP average scale scores and the 2009 MAP average scale scores. Therefore, I supported the hypothesis that for the total population and seven subgroup populations in elementary communication arts and mathematics a significant difference existed in the 2008 MAP scale scores and 2009 MAP scale scores. Keep in mind, the results at the elementary level do not provide proof positive that Study Island was the reason for the significant difference; however, clear evidence exists that MAP scores increased noticeably from 2008 to 2009.

The results of the 16 statistical tests performed at the middle school level on communication arts and mathematics random samples from the total population and seven subgroup populations produced mixed results. In opposition to the elementary results, I did not reject the null hypothesis based on the statistical tests on random samples from the total population and seven subgroup populations in middle school communication arts. Therefore, I did not support that there was a significant difference in the 2008 MAP average scale scores and the 2009 MAP average scale scores in middle school communication arts. Interestingly, the mixed results came in the area of middle school mathematics where the total population and three subgroups (School Total, Black, White, Limited English Proficiency) were found to have no significant difference while four subgroups (Asian/Pacific Islander, Hispanic, Free and Reduced Lunch, Individualized Education Plan) did have a significant difference. The results at the middle school level, minus four subgroups in mathematics, were almost the exact opposite of the elementary level.

Chapter 5 discusses the implications of the results found in each of the statistical analyses conducted as part of this study. The reader will find inferences and summaries of the general themes that arose upon analysis of the data. In addition, I suggest what these results may mean for school districts in terms of using Study Island to raise student achievement. Finally, recommendations provide several areas in which schools may benefit from further research.

Chapter Five: Implications and Recommendations

Overview

No Child Left Behind requires all public schools and districts to meet Adequate Yearly Progress on an annual basis. Failure to meet Adequate Yearly Progress can occur for several reasons. Many of the strategies, programs, and resources that schools implement to obtain the desired result of meeting Adequate Yearly Progress are ineffective, or they may not be implemented with fidelity. School officials may use the results of this research to determine whether using Study Island could effect student achievement in the grade levels and subject areas where they struggle to meet Adequate Yearly Progress.

This quantitative study focused on determining the effect of Study Island on student achievement. To that end, I analyzed whether a significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores. The district in this study used Study Island for the first time during 2008-2009 (2009 MAP tests) and therefore, did not use Study Island during 2007-2008 (2008 MAP tests). Hence, the use of Study Island was the independent variable and student achievement, as measured by the MAP tests, was the dependent variable.

I disaggregated the 2008 and 2009 MAP scale scores by level (elementary and middle) and subject (communication arts and mathematics). For each level and subject, I analyzed random samples from the total population and each of the seven subgroup populations. The eight groups analyzed for each level and subject included School Total, Asian/Pacific Islander, Black, Hispanic, White, Free and Reduced Lunch, Individualized Education Program, and Limited English Proficiency.

Data Analysis

The objective of this study was to determine the effect of Study Island (independent variable) on student achievement (dependent variable). Therefore, I conducted a statistical analysis of the 2008 MAP average scale scores and the 2009 MAP average scale scores to determine whether a significant difference existed between the two data sets. As such, *z* tests and *t* tests evaluated (using a 95% confidence interval) the MAP scale scores of each random sample taken from the various groups. However, *t* tests were only used when the random sample contained fewer than 30 MAP scale scores. The null hypotheses stated that no significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores at either level. Ultimately, the data analysis of the *z* tests and *t* tests on the random samples allowed me to make a decision regarding whether the null hypotheses were rejected or not.

Elementary communication arts and mathematics. The *z* values and *t* values for the random samples of each group fell into the critical regions on a bell shaped curve (using a 95% confidence interval), thereby confirming that at the elementary level there was a significant difference in student achievement in communication arts and mathematics for the total population and each of the seven subgroup populations. Thus, the results at the elementary level strongly imply that some factor, or combination of factors, positively effected Adequate Yearly Progress student achievement in communication arts and mathematics from 2007-2008 (2008 MAP scale scores) to 2008-2009 (2009 MAP scale scores). I am suggesting that at the elementary level the use of Study Island contributed to this difference in student achievement. These results are not surprising when considering that the study Magnolia Consulting conducted for Study

Island primarily contained examples of how achievement had increased at the elementary level (Watts, 2008).

Middle school communication arts and mathematics. At the middle school level for communication arts, the statistical tests indicated that there was not a significant difference in MAP scale scores between 2007-2008 and 2008-2009 in either the total population or any of the seven subgroup populations. This led me to conclude that at the middle school level for communication arts, the use of Study Island did not have a significant impact on student achievement.

At the middle school level for mathematics, the results were mixed. The statistical tests indicated that there was not a significant difference in MAP scale scores between 2007-2008 and 2008-2009 in the total population, School Total, or the three subgroups of Black, White, and Limited English Proficiency. Therefore, I concluded that for those four groups, the use of Study Island did not have a significant impact on student achievement. On the other hand, for the subgroups of Asian/Pacific Islander, Hispanic, Free and Reduced Lunch, and Individualized Education Plan, the statistical analysis confirmed that there was a significant difference in MAP scale scores between 2007-2008 and 2008-2009 in the area of middle school mathematics. These results seem to indicate that there is a propensity for Study Island to increase mathematics achievement when compared to communication arts which is consistent with the findings of Magnolia Consulting (Watts, 2008).

Overall, communication arts and mathematics Adequate Yearly Progress achievement at the middle school level increased from 2007-2008 to 2008-2009. However, achievement was not effected significantly in communication arts and had mixed results in mathematics at the middle school level from 2007-2008 (2008 MAP scale scores) to 2008-2009 (2009 MAP scale scores). Similarly, the study that Magnolia Consulting conducted contained fewer results from the middle school level (Watts, 2008). As stated in the previous section, in education it is difficult to isolate one factor that is the sole cause for the impact on student achievement. While the middle school results are encouraging and warrant continued investigation, for some reason or reasons the results were not as significant when compared to the elementary level. In the Implications section I discuss in more depth the potential reasons for the results that were referenced in this section on Data Analysis.

Implications

The use of Study Island was a major change between 2007-2008 and 2008-2009, across elementary and middle schools in the district used for this study. After the implementation of Study Island, the district experienced more gains in overall and subgroup achievement than previously recorded. The results of the data analysis suggest that some factor or factors at the elementary level had a significant effect on communication arts and mathematics student achievement when comparing 2008 and 2009 MAP test scores. This is useful information that provides the data necessary to drive discussions among school officials to determine the factor or factors that had the biggest impact. Similarly, the conversation among middle school officials could also be driven by the data analysis however, since only 4 out of the 16 groups showed a significant difference in student achievement, the determinations might differ. The rest of this section will provide implications that are relevant to both elementary and middle school stakeholders.

Curriculum. A guaranteed and viable curriculum is one of the cornerstones of school-level factors that a district can provide to ensure opportunities for students to learn (Marzano, 2003). Simply put, curriculum is the program that outlines the learning experiences a student will encounter under the direction of the school (Oliva, 1982). Implementation of a new or revised curriculum, along with professional development for teachers, often has a positive impact on student achievement. This occurs because a new or revised curriculum provides teachers with the most updated standards assessed on state tests. Likewise, professional development provides teachers with the support they need to teach the updated curricular standards.

In this study, the same year the elementary students began using Study Island to practice communication arts and mathematics, the teachers received a revised communication arts curriculum and one day of professional development on its implementation. However, the same was not true in the area of mathematics. Moreover, it had been several years since the mathematics curriculum had been revised at the elementary level. Therefore, the fact that all eight Adequate Yearly Progress groups were significantly effected in both communication arts and mathematics seems to suggest that something other than just a revised curriculum helped to achieve these results. This difference lends support to my conclusion that Study Island is the factor that contributed to this difference in student achievement.

Similarly, the same year the middle school students began using Study Island, the teachers in both communication arts and mathematics received a new curriculum along with professional development. Teachers in these areas participated in one day of professional development prior to the beginning of school. Unfortunately, the student achievement results at the middle school level were not statistically significant..A drawback to implementing a new curriculum and a program, such as Study Island, at the same time is that learning the curriculum can consume teacher attention that otherwise could be directed towards better implementation of Study Island. In addition, it also makes it difficult to determine which variable, the new curriculum or Study Island, had the biggest effect on student achievement. Fortunately, Study Island aligns to state standards therefore, when students use the program they are engaged in content that supports the district curriculum that is developed from state standards.

Instruction. One of the most prized commodities for teachers is time, more specifically instructional time. This makes sense because research has shown that at the teacher-level, instructional strategies and classroom curriculum design greatly effect student achievement (Marzano, 2003). Moreover, the way teachers conduct classroom lessons will directly impact the opportunity to learn and student time on task (Lezotte, 1991).

The district in this study has a trend of communication arts MAP scores decreasing when students enter sixth grade, which is the year students enter middle school. This type of achievement dip is common in many districts when students transition into middle school. Moreover, the dip in MAP scores mirrors the decrease in communication arts and reading instructional time that occurs when students enter middle school. For example, students switch from a two-hour communication arts and literacy block in elementary school to a 50-minute communication arts class with no specific reading instruction in middle school. It would be difficult for any program to make a significant difference in improving achievement when this much instructional time with a teacher is lost from one year to the next.

In mathematics, the district in this study has a trend of MAP scores either remaining flat or decreasing when students enter sixth grade. Similar to the trend in communication arts, this trend in middle school mathematics MAP scores mirrors a decrease in time spent on mathematics when students enter middle school. However, the difference in time is not as drastic when compared to communication arts. Therefore, it is easier to hold achievement steady or increase it and likewise, it is more likely for a program to have a significant effect on student achievement when there is not as much loss of instructional time to overcome.

It is intriguing that half of the middle school mathematics subgroups (Asian/Pacific Islander, Hispanic, Free and Reduced Lunch, Individualized Education Program) showed a significant difference in MAP scores from 2008 to 2009 while the same could not be said for even one middle school communication arts subgroup. It would be interesting to further investigate and compare the ability of subgroups to increase achievement connected to reading literacy (communication arts) versus numerical literacy (mathematics). It might seem logical that language barriers that may exist in subgroups such as Asian/Pacific Islander and Hispanic make it more difficult to increase communication arts achievement since this content area is based on words while the content area of mathematics is based on numbers that might be more universally understood. In that same vein, it might be worthwhile to set up a study that determines whether socioeconomic status, as measured by Free and Reduced Lunch, and special education status, as measured by Individualized Education Program, impacts the ability to increase achievement in one area over another more difficult when considering communication arts and mathematics.

It is difficult to overcome certain barriers simply by implementing a new program. Nevertheless, programs can provide valuable support until better solutions arise that directly eliminate the barriers. Study Island can help address the decrease in instructional time when students enter middle school; however, it cannot eliminate this barrier or take the place of teacher instruction.

Assessment. In order to make instructional decisions and measure whether students have learned the curriculum, teachers must provide assessment opportunities for students to demonstrate what they know and can do. Effective assessment practices should be formative and frequent in nature while providing students feedback that encourages them to improve (Marzano, 2006). In fact, formative assessment practices can yield dramatic improvements in student achievement especially, for low achievers (Black & Wiliam, 1998a). However, in addition to short-term assessments, it is also necessary for teachers and students to engage in less frequent snapshots of learning which are summative in nature and often used for accountability purposes.

Each time a student uses Study Island, they are participating in a formative assessment opportunity. The student receives immediate feedback on their progress and motivation to improve which comes in the form of short games earned after correct answers. Best of all, using Study Island is a safe way for students to engage in the assessment process without fear of failure or worrying about being graded. For the first time in the educational careers of many students, especially low achievers, they may be having fun while participating in the assessment process. In addition, teachers are able to more effective monitor student learning since they are not burdened with traditional grading tasks and can focus their time and energy on analyzing results and planning instruction.

During 2008-2009 and 2009-2010, the district in this study had access to benchmark assessments through Study Island. Benchmark assessments are tests administered approximately three to four times throughout the year to provide teachers and administrators feedback on student achievement and progress. In addition, the Study Island benchmark assessments mimicked the MAP state tests in content and structure and were available for online administration. Administrators and teachers could disaggregate the benchmark assessment data by district, building, grade level, teacher, and student. More importantly, the benchmark assessments also yielded an estimated achievement level (Below Basic, Basic, Proficient, Advanced), which predicted how an individual student would perform on the MAP test.

The number of Study Island benchmark assessments administered to students varied by school during 2008-2009 and 2009-2010. For example, some buildings administered all four of the Study Island benchmark assessments, while others did not use them at all. The varied use of benchmark assessments could be a factor that effected the MAP scale scores used in this study. One of the reasons that schools used the benchmark assessments was to acquire data for driving intervention and tutoring opportunities for students whose achievement level was below proficient. It would be beneficial to determine whether this method of identifying students for interventions and tutoring was effective and accurate. Future studies should consider regulating the use of this option.

In fact, the use of benchmark assessments and the effect on student achievement could be the basis for an entire study.

Professional development. The state tests required by No Child Left Behind provide school officials with detailed student achievement data however, like most accountability policies it assumes that administrators and teachers know how to turn the data into information (Massell, 2001). In fact, educators have never lacked data however, the information extracted from the data has often been poor (Wayman, Stringfield, & Yakimowski, 2004). Districts and schools paralyzed by data are suffering from the DRIP syndrome, data rich-information poor. As with many educational initiatives where schools struggle, such as data analysis, lack of professional development for teachers exists (Newmann, Kings, & Youngs, 2000).

Massell (2001) found that even when districts provide data analysis professional development, it is usually only delivered to a handful of teachers which means the majority of the staff must informally learn from other personnel, as opposed to being directly supported by the expert. This is exactly what happened the same year Study Island was implemented in the district in this study. The administrators and a select group of teachers at each elementary and middle school received MAP data analysis professional development from the Director of Assessment. In previous years, although schools received data, the administrators and teachers did not receive this level of support. Unfortunately, the fact that only a small percentage of teachers directly participated in this opportunity, it may not have had enough impact to significantly effect the practice of most teachers and student achievement.

Study Island usage. Study Island supports the "second generation" of frequent monitoring of student progress Lezotte (1991) stated would involve technology that enables teachers to more effectively monitor student learning. In the district used for this study, thousands of students used Study Island during the 2008-2009 school year. Students accessed the program through the Internet by using a computer at school or at home. Teachers were able to view student data immediately and plan instruction accordingly.

Does usage matter? With most programs designed to effect student achievement, it is up to the administrators and teachers to figure out how they are ultimately used. While the companies that supply the program usually provide training or professional development, school officials within the district must determine how to use, or not use, the program on a daily basis. Therefore, even in the same district, usage can vary by building, classroom, or even from student to student. Variation is not bad in and of itself; however, when attempting to study the impact of a variable on a large scale, such as across an entire district, it can create challenges.

In education, it is often difficult to isolate one single variable that is the sole cause for a particular result especially, when several new or different variables are occurring simultaneously. However, by using data to investigate a particular variable it is possible to reach supportable conclusions. The data in chapter 4 illustrated how Study Island was used frequently and continuously throughout 2008-2009 by students and teachers. In this study, I believe the Study Island usage data provides evidence that Study Island had a far greater effect on student achievement than one day of communication arts and mathematics curriculum professional development or one day of data analysis support for administrators and selected teachers.

Since 2008-2009 was the first year each elementary and middle school used Study Island, it is reasonable to investigate whether the increase in overall communication arts and mathematics achievement on the MAP occurred because of the use of Study Island. Interestingly, the degree to which each school used Study Island differed significantly. The vast difference among schools in regards to Study Island usage may explain why there was not a uniform pattern of results especially, at the middle school level. Chapter 4 included data that illustrated the usage variance among schools. Nevertheless, even though Study Island usage varied by school, it was used enough to conclude that it deserved strong consideration for contributing to increased student achievement.

Study Island was a program that was added to each elementary and middle school during the 2008-2009 school year, but it did not take the place of an existing assessment program. This is an important fact to consider when reviewing the Study Island usage data presented in chapter 4. In fact, it is reasonable to conclude that the number of questions answered and amount of time on task would not have occurred without Study Island. Although there might be instances to the contrary, teachers did not stop giving their traditional unit tests, quizzes, and homework assignments when students began using Study Island. This is because it was not mandatory for all teachers and students to use Study Island, and in many cases it served as an intervention or extension opportunity for selected students. Given the magnitude of the Study Island usage data, I believe it is hard to dismiss the claim that it strongly effected student achievement, at least at the elementary level.

Subgroups. In this study, the School Total group included all students and in addition, each student was part of at least one of the seven subgroups. It is possible however, for a student to be a member of several subgroups. For example, a student in the School Total group could also be included in the White, Free and Reduced Lunch, and Individualized Education Program subgroups which means their MAP scores count for each of these four Adequate Yearly Progress groups. Therefore, when a student is able to increase their achievement on the MAP state tests it helps at least two Adequate Yearly Progress groups however, it could impact up to five (School Total, Race, Free and Reduced Lunch, Individualized Education Program, Limited English Proficiency).

All of the schools in this study have time set aside on a weekly basis for providing students with intervention and extension opportunities that could involve using Study Island. Likewise, each school in this study had a tutoring budget that supported targeting students who needed to improve their academic performance in communication arts and mathematics. It is common for schools to look at MAP scores to determine the students with the greatest need and provide extra tutoring for them. Thus, an effective tutoring program can impact subgroup achievement and the impact could vary by subject area depending on the time and effectiveness of the action in each area, communication arts and mathematics. By investigating these variables closer, it might explain the mixed results at the middle school level.

Recommendations

Student achievement in this district had increased in communication arts and mathematics prior to using Study Island. However, the consistency and magnitude of the increases after Study Island had been implemented (2009 and 2010) are noticeably greater than the increases prior to using Study Island. When triangulating the statistical results from this study with Study Island usage data and Adequate Yearly Progress data, I concluded that the district in this study can attribute the increase in MAP scores to the use of Study Island. Nevertheless, any number of other factors, including but not limited to the new curriculum, professional development, and data analysis support discussed in the previous sections, could have effected student achievement during this study, thereby skewing the results of the study. In the following sections, I identify and make recommendation on several areas related to the effect of Study Island use on MAP scores that school officials or future studies may still want to investigate.

Extending the timeframe of the study. One of the limitations of this study was the timeframe during which it was conducted. This study analyzed student achievement data after the elementary and middle schools had used Study Island for only one year, which is not enough time for a conclusive cause and effect relationship to be established. Investigating the impact of Study Island over a longer period is one way researchers could collect reliable data about the relationship between the use of Study Island and student achievement. In a new study, a recommended timeframe would involve three to five years of data as results are more reliable when the studies occur over a longer timeframe.

Case study and qualitative components. This was strictly a quantitative study focused on determining the effect of Study Island on student achievement in communication arts and mathematics, as measured by MAP state tests. Stakeholders could expand on this body of research by adding a qualitative component or components that compliment the quantitative components. Researchers could conduct a mixed methods study which might involve surveying administrators, teachers, students, and parents in an effort to isolate the specific usage strategies that yield the best results. Likewise, in a case study investigators may use a qualitative component to examine usage strategies in specific schools or classrooms and correlate them with student achievement.

Qualitative studies such as these might yield results of great value to the specific district, schools, classrooms, administrators, teachers, and/or students involved. For example, a study involving only one school and just a few classrooms might allow the researcher to carry out a deeper analysis using mixed methods. However, a drawback to a study with such a narrow sample population is that the results may not translate well to other schools and classrooms that have differing demographics (Appendix A) or circumstances. Although this study only included one district, it included several schools with different demographics and therefore, provided a better snapshot of what other districts could expect if they chose to use Study Island.

Control group. This study used 2008 and 2009 MAP scale scores to determine whether a significant difference existed in student achievement between those two years. During 2007-2008, no students in the district used Study Island however; in 2008-2009, almost all students used Study Island. In education, it is difficult to set up a control group of students, with representative demographics, who had not used Study Island during 2008-2009. In addition, it is difficult to justify using a program with the potential to help all students, but intentionally withhold it from a group of students. In a future study, however, a district might decide to use Study Island in one school or classroom and compare the gains in that school/classroom with gains in a school/classroom of similar demographics in the same district that did not use Study Island. School districts often use this approach to limit the expenditures on a program until it has proven successful.

Different statistical measures. This study analyzed only one variable, the use of Study Island. Future studies, on the other hand, might benefit from investigating more than one variable. For example, an investigation could involve the analysis of variables such as the number of Study Island questions answered correctly and MAP scale score data to find the Pearson Product-Moment Correlation Coefficient of the two data sets. The correlation coefficient of two variables shows the strength of the relationship between those two variables, as well as whether the relationship is positive or negative (Bluman, 2008). In addition, after finding the correlation coefficient, the study could include a regression analysis on the data to determine whether it would be possible to predict the MAP scale score of a student based on his or her number of Study Island questions answered correctly.

Disaggregate Study Island usage data. Since this study included achievement data from each elementary and middle school across the district, the Study Island usage data collected for this study was very broad. The usage data for this study did not go any further than the school and subject level. However, Study Island has the ability to disaggregate the usage data down to the individual student. Therefore, it is possible to report usage data more specifically such as by grade-level, classroom, and subgroup. A smaller, more targeted study would benefit from looking at usage data in a more specific manner.

Study Island also has the ability to disaggregate data by home versus school usage. This would be interesting to study since it could provide school officials with information on how Study Island effects instructional time during the school day and quantify the amount of out-of-school impact the program has generated. A future study might involve investigating the connections that are made between home and school through the use of Study Island.

Conclusion

When researching the relationship between the use of Study Island and student achievement, the researcher analyzed whether a statistically significant difference existed between the 2008 MAP average scale scores and the 2009 MAP average scale scores. A significant difference existed in the total population and all seven subgroup populations at the elementary level for communication arts and mathematics. Conversely, a significant difference did not exist in the total population and seven subgroup populations for middle school communication arts or for middle school mathematics in the total population and subgroup populations Black, White, and Limited English Proficiency. However, a significant difference existed for middle school mathematics in the subgroup populations Asian/Pacific Islander, Hispanic, Free and Reduced Lunch, and Individualized Education Program.

At the elementary level, there is strong evidence that the use of Study Island was a cause of the increase in scores in communication arts and mathematics. Study Island is appealing to elementary students because it allows them to use the computer and earn the opportunity to play games by answering questions correctly. As expected, no matter what the ethnicity, socioeconomic status, or disability level, elementary students are typically enthusiastic about school and excited about the opportunity to play games. Fortunately, Study Island offers a way to capitalize on factors that are appealing to elementary students while increasing their time on task in communication arts and mathematics.

At the middle school level, although overall achievement in communication arts increased, a significant effect was not present in the total population and seven subgroup populations. As discussed earlier, the district in this study has a trend of middle school communication arts achievement decreasing in sixth grade, as well as a significant decrease in communication arts and literacy instruction when students enter middle school. In addition, if students are below grade-level in their reading ability it will likely impact their ability to productively use Study Island and may discourage them from using it voluntarily. Although there is evidence that Study Island increased time on task in communication arts, as evidenced by the Study Island usage data in Tables 73 through 78, it was not enough to overcome some of the other variables and make a significant difference. Therefore, educators should keep in mind that while Study Island provides practice on communication arts standards, it is not designed to address reading and literacy components.

At the middle school level, overall achievement in mathematics increased and a significant impact occurred in four subgroups (Asian/Pacific Islander, Hispanic, Free and Reduced Lunch, Individualized Education Program). When students enter middle school, there is some decrease in mathematics instructional time however, not nearly as much when compared to communication arts. Similar to communication arts, there is evidence that Study Island increased time on task in mathematics. Given the subgroups significantly impacted, it is difficult to detect a pattern and offer potential explanations. Since mathematics is not as dependent on reading ability, when compared to communication arts, it may have been easier for students of all reading levels to productively use the mathematics component of Study Island when compared to using the communication arts component of Study Island. In this study, the two largest groups were the total population, School Total, and the subgroup White. Neither of these groups increased significantly however, four of the six smaller subgroups did see a significant increase. This may provide evidence that the smaller the group, the more likely Study Island will significantly effect MAP scores.

Many times, in the system of accountability that No Child Left Behind has created, educators are looking for a silver bullet to meet the challenges they face when it comes to meeting Adequate Yearly Progress on an annual basis. These silver bullets often come in the form of a program schools implement with the hope that the program will have a large enough impact on student achievement that the school will meet Adequate Yearly Progress. In reality, the fidelity and degree to which a program is used is just as critical as the content of the program. This is especially true when it comes to analyzing how the program effected student achievement. There is no reason to suspect that program implementation is the only factor that effects student achievement.

The possibility exists that the poor execution of factors such as curriculum, instruction, assessment, professional development, and the implementation of a program may play a larger role than is commonly suspected in causing poor student achievement. Rarely does the first instinct of an educator lead him or her to question his or her own practices and look internally at how he or she is addressing the factors listed previously. Instead, the first instinct of an educator is to look outside the classroom or school and focus on resources he or she feels are missing or factors that are out of the control of the school, such as parental support and involvement. Nevertheless, an extensive commitment to research these factors and reflect on current practices is necessary before school personnel begin to shift their understanding of research-based best practices and act accordingly.

Schools should be responsible for collaboratively examining their practices and connecting them to factors that research has shown to impact student achievement. These factors include areas such as curriculum, instruction, assessment, professional development, intervention, and leadership. Schools and districts are accountable for addressing concerns in these areas and have the ability to choose programs that will support their efforts.

The system of accountability that provides the framework for what schools and districts must demonstrate on an annual basis should promote and reward best practices. The quick fix required from year to year through No Child Left Behind perpetuates the continued search for the silver bullet. There are many things that school officials can do in an attempt to increase student achievement, but they must first identify the most pressing needs and formulate action plans to address them. To improve student achievement and productively use resources and programs, administrators and teachers must work interdependently while frequently monitoring progress and using data to drive their decisions and actions.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association, Inc.

Ainsworth, L. (2003). *Power standards: Identifying the standards that matter the most.* Englewood, CO: Advanced Learning Press.

American Institutes for Research (AIR). (2010, October 25). *Students learning expectations gap can be twice the size of national black-white achievement gap.* Retrieved January 7, 2011, from http://www.air.org/news/index.cfm?fa=viewContent&content_id=1022

Archipelago Learning. (2010). *Study island history and timeline.* Retrieved on October 24, 2010 from http://www.archipelagolearning.com/?action=history

Ayala, C. C., Shavelson, R. J., Ruiz-Primo, M. A., Brandon, P. R., Yin, Y., Furtak, E. M., et al. (2008). From formal embedded assessments to reflective lessons: The development of formative assessment studies. *Applied Measurement in Educaiton*, *21*(4), 315-334.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback on test-like events. *Review of Educational Research, 61(2), 213-238.*

Bangert-Drowns, R. L., Kulik, C C., & Kulik, J. A., (1991). Effects of classroom testing. *Journal of Educational Research, 85(2), 89-99.*

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86, 8-21.*

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5(1), 7-70.*

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80(2), 139-148.*

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.

Bluman, A. G. (2008). *Elementary statistics: a step by step approach*. New York, NY: McGraw-Hill.

Boston, C. (2002). *The concept of formative assessment*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Brimijoin, K., Marquissee, E., & Tomlinson, C. A. (2003). Using data to differentiate instruction. *Educational Leadership, 60*(5), 70-73.

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education, 10, 161-180.*

Brookhart, S. M. (2008). *How to give effective feedback to your students*. Alexandria, VA: Association for Supervision and Curriculum Development.

Callingham, R. (2008). Dialogue and feedback: Assessment in the primary mathematics classroom. *Australian Primary Mathematics Classroom, 13*(3), 18-21.

Cauley, K. M., Pannozzo, G., Abrams, L., McMillan, J., & Camou-Linkroum (2006). *The relationship between classroom assessment practices and student motivation and engagement*. Richmond, VA: Virginia Commonwealth University.

Center on Education Policy. (2007, May 31). *Answering the question that matters most: Has student achievement increased since no child left behind?* Retrieved November 7, 2010, http://www.cep-dc.org/index.cfm?DocumentTopicID=16

Chappuis, S. & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership, 60*(1), 40-43.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfield, F. D., & York, R. L. (1966). Equality of educational opportunity. Washington, DC: U. S. Government Printing Office.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

Crowie, B. (2005). Student commentary on classroom assessment in science: A socio-cultural interpretation. *International Journal of Science Education, 27*(2), 199-214.

Davies, A. (2000). *Making classroom assessment work*. Courtenay, BC, Canada: Connections Publishing.

Edmonds, R. R., (1982). *Programs of school improvement: An overview*. Washington, DC: National Institute of Education.

Education Week (2010). Achievement gap [Electronic Version]. *Education Week,* Retrieved January 8, 2011, from http://www.edweek.org/ew/issues/achievement-gap/

Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., et al. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record, 52*, 187-201.

Even, R. (2005). Using assessment to inform instructional decisions: How hard can it be? *Mathematics Education Research, 17*(3), 45-615.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta analysis. *Exceptional Children, 53(3), 199-208.*

Harlan, W., & Deakin, C. R., (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning. In *Research Evidence in Education Library.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Harvey, J. J. (1983). *A nation at risk: The imperative for educational reform.* Retrieved October 17, 2010, from http://www2.ed.gov/pubs/NatAtRisk/index.html.

Hattie, J. A. (1992). Measuring the effects of schooling. *Australian Journal of Education, 36*(1), *5-13.*

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), *81-112.*

Jencks, C., Smith, M. S., Ackland, H., Bane, M. J., Cohen, D., Grintlis, H., Heynes, B., & Michelson, S. (1972). *Inequity: A reassessment of the effects of family and schooling in America.* New York: Basic Books.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*(2), *211-232.*

Kulik, J. A. & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), *791-97.*

Landauer, T. K., Lochbaum, K. E., & Dooley, S. (2009). A new formative assessment technology for reading and writing. *Theory into Practice, 48*(1), *44-52.*

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment minute by minute, day by day. *Educational Leadership, 63*(3), *18-24.*

Lezotte, L. (1991). *Correlates of effective schools: The first and second generation*. Okemos, MI: Effective Schools Products, Ltd.

Lezotte, L. (2001). *Revolutionary and evolutionary: The effective schools movement*. Okemos, MI: Effective Schools Products, Ltd.

Marzano, R. J. (2003). *What works in schools: Translating research into action*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2006). *Classroom assessment and grading that works*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D. J., & Pollock, J. E., (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Massell, D. (2001). The theory and practice of using data to build capacity: State and local strategies and their effects. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (pp. 148–169). Chicago: University of Chicago Press.

Missouri Department of Elementary and Secondary Education (MO DESE). (2001). *MAP: score use, meaningfulness, and dependability.* Retrieved March 17, 2010, from http://dese.mo.gov/divimprove/fedprog/discretionarygrants/ReadingFirst/DMAP.pdf

Missouri Department of Elementary and Secondary Education (MO DESE). (2006). *Standards and indicators manual: accreditation standards for public school districts in missouri.*. Retrieved February 22, 2010, from http://dese.mo.gov/divimprove/sia/msip/Fourth%20Cycle%20Standards%20and%20Indicators.pdf

Missouri Department of Elementary and Secondary Education (MO DESE). (2008). *Technical report 2008*. Retrieved February 26, 2011, from http://www.dese.mo.gov/divimprove/assess/tech/documents/MAP-ATechncialReport2008.pdf

Missouri Department of Elementary and Secondary Education (MO DESE). (2009a). *Missouri assessment program: guide to interpreting results.* Retrieved February 22, 2010, from http://dese.mo.gov/divimprove/sia/dar/girspring2009.pdf.

Missouri Department of Elementary and Secondary Education (MO DESE). (2009b). *Missouri ayp summary 2009: improvement status of all districts, with levels of sanctions*. Retrieved February 22, 2010, from http://dese.mo.gov/divimprove/nclb/documents/Final_List_of_DistrictSchools_AYPStatus20091120.pdf

Missouri Department of Elementary and Secondary Education (MO DESE). (2009c). *School Profile*. Retrieved February 22, 2010, from http://dese.mo.gov/planning/profile/apr/ayp092088.html

Missouri Department of Elementary and Secondary Education (MO DESE). (2009d). *Technical report 2009*. Retrieved February 26, 2011, from http://www.dese.mo.gov/divimprove/assess/tech/documents/2009-Tech-Report.pdf

Missouri Department of Elementary and Secondary Education (MO DESE). (2009e). *Understanding your annual performance report*. Retrieved February 22, 2010, from http://www.dese.mo.gov/divimprove/sia/dar/understandingyourAPR.pdf

Missouri Department of Elementary and Secondary Education (MO DESE). (2010a). Final adequate yearly progress. Retrieved December 29, 2010, from http://dese.mo.gov/planning/profile/apr/ayp000000.html

Missouri Department of Elementary and Secondary Education (MO DESE). (2010b). *Technical report 2010*. Retrieved February 26, 2011, from http://www.dese.mo.gov/divimprove/assess/tech/documents/asmt-gl-2010-tech-report.pdf

Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.). *Handbook of research on educational communication and technology* (pp. 745-783). Mahwah, NJ: Lawrence Erlbaum Associates.

National Assessment of Educational Progress (NAEP). (2010). Mathematics Grade 4 State Results. Retrieved December 29, 2010, from http://nationsreportcard.gov/math_2009/gr4_state.asp.

National Research Council. (1997). *Educating one and all: Student with disabilities and standards-based reform.* McDonnell, L. M., & McLaughlin, M. J., eds. Washington, DC: National Academy Press.

Newmann, F., King, B.,&Youngs, P. (2000). Professional development that addresses school capacity: Lessons from urban elementary schools. *American Journal of Education, 108,* 259–299.

Oliva, P. F. (1982). *Developing the curriculum.* Boston, MA: Little, Brown and Company.

Popham, J. (2006). Phony formative assessments: Buyer beware. *Educational Leadership*, *64*(3), 86-87.

Popham, J. (2007). Instructional sensitivity: Educational accountability's dire deficit. *Phi Delta Kappan*, *89*(2), 149-155.

Popham, J. (2008). Formative assessment: Seven stepping-stones to success. *Principal Leadership*, 9(1), 16-20.

Popham, J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, *48*(1), 4-11.

Rieg, S. A. (2007). Classroom assessment strategies: What do students at-risk and teachers perceive as effective and useful? *Journal of Instructional Psychology*, *34*(4), 214-225.

Roberts, J. (2010, July 31). Tennessee student scores plunge in new proficiency test. *The Commercial Appeal*. Retrieved December 29, 2010, from http://www.commercialappeal.com/news/2010/jul/31/student-scores-plunge-in-new-test/

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119-144.

Schmoker, M. (1999). *Results: The key to continuous school improvement* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Scriven, M. (1967). *The methodology of evaluation*. Washington, DC: American Educational Research Association.

Shepard, L. A. (2000). *The role of classroom assessment in teaching and learning.* (CSE Technical Report). Los Angeles: National Center for Research and Evaluation, Standards, and Student Testing.

Shepard, L. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, *63*(3), 66-70.

Shute, V. J. (2008). Focus on formative assessment. *Review of Educational Research*, *78*(1), 153-189.

Solorzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English Language Learners. *Review of Educational Research*, *78*(2), 260-329.

Sparks, S. D. (2010, October 25). States set widely varying proficiency bars. *Education Week*. Retrieved January 3, 2011, from http://www.edweek.org/ew/articles/2010/10/25/10air.h30.html?tkn=PLMFm/CtDOcCdop6kpKUElRykvxm3RBFZAMV&cmp=clp-edweek

Stiggins, R. J. (1998). *Classroom assessment for student success.* Student Assessment Series. Annapolis, MD: NEA Professional Library.

Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment*. Educational Measurement: Issues and Practice, 20*(3), 5-14.

Stiggins, R. J. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan, 86*(1), 22-27.

Stiggins, R. J. (2007). Assessment through the student's eyes. *Educational Leadership, 64*(8), 22-26.

Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2005). *Classroom assessment for student learning: Doing it right-using it well* . Portland, OR: Assessment Training Institute.

Stiggins, R. J. & Chappuis, J. (2008). Enhancing student learning. *District Administration, 44*(1), 42-44.

Study Island. (2010). *Missouri: GLE and common core standards mastery and map preparation.* Retrieved on February 20, 2010 from http://www.studyisland.com/demoAsk.cfm?myState=MO

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202-248.

Trammel, D. L., Schloss, P. J., & Alper, S. (1994). Using self-recording and graphing to increase completion of homework assignments. *Journal of Learning Disabilities*, 27(2), 75-81.

United States Department of Education (DOE), Office of the Deputy Secretary, *No child left behind: A toolkit for teachers.* Washington, D.C., 2004.

United States Department of Education (DOE). (2010). *No child left behind (NCLB).* Retreived on February 20, 2010, from http://www.ed.gov/nclb/overview/intro/parents/parentfacts.html

Valencia, S. W. (2008). *Understanding assessment: Putting together the puzzle. Beyond the Book.* Retrieved from http://www.beyond-the-book.com/strategies/strategies_100307.html

VanDeZande, J. (2010, October 5). Re: 2010-11 Missouri assessment program (map) – suspension of performance events and writing prompts assessments [Electronic mailing list message]. Retreived from http://www.dese.mo.gov/divimprove/assess/documents/Performance_Event_and_Writing_Prompt_Suspension.pdf

Watts, J. (2008). *Case study research summaries of study island.* Louisa, VA: Magnolia Consulting.

Wayman, J. C., Stringfield, S., & Yakimowski, M. (2004). *Software enabling school improvement through the analysis of student data* (Report No. 67). Retrieved December 14, 2004, from Center for Social Organization of Schools Web site: http://www.csos.jhu.edu/crespar/techReports/Report67.pdf

Wiley, C. R. H. (2008). Traditional teacher tests. In T. L. Good (Ed.), *21st century education: A reference handbook: Vol 1*. (pp. 431-442). Los Angeles: SAGE.

Appendix A

Elementary and Middle School Enrollment and Demographics 2009

Table B1

Elementary School Enrollment and Demographics 2009

School	Enrollment	% FRL	% White	% Black	% Asian	% Hispanic
ES 1	942	14.7	83.2	7.3	5.8	3.6
ES 2	951	11.7	94.4	2.3	1.9	1.3
ES 3	925	35.9	81.5	12.6	2.4	3.1
ES 4	414	8.2	97.8	0.7	0.7	0.5
ES 5	911	19.5	89.1	6.7	2.2	1.9
ES 6	524	12.2	92.6	4.4	1.3	1.5
ES 7	711	39.0	74.4	13.9	6.3	4.8
ES 8	750	6.7	89.3	4.4	3.2	3.1
ES 9	813	6.6	90.4	4.6	3.8	1.2
ES 10	758	12.0	89.3	3.0	4.6	3.0

Note. Adapted from Missouri Department of Elementary and Secondary Education School Profile Student Demographics. Retrieved February 22, 2010, from http://dese.mo.gov/planning/profile/.

Table B2

Middle School Enrollment and Demographics 2009

School	Enrollment	% FRL	% White	% Black	% Asian	% Hispanic
MS 1	837	18.0	82.7	8.6	4.5	4.1
MS 2	707	14.6	90.2	6.9	1.7	1.1
MS 3	801	5.5	93.5	3.4	1.9	1.1
MS 4	766	16.4	88.5	6.5	2.9	2.0
MS 5	900	14.9	89.1	7.0	2.1	1.4

Note. Adapted from Missouri Department of Elementary and Secondary Education School Profile Student Demographics. Retrieved February 22, 2010, from http://dese.mo.gov/planning/profile/.

Appendix B

Travis Bracht
3409 East Lime Kiln
Saint Charles, Missouri, 63301

September 9, 2009

Dear Mr. Bracht,

I am writing to grant your request for permission to use elementary and middle school Study Island and MAP data from the Francis Howell School District for your doctoral dissertation titled <u>The Affect of Study Island on Student Achievement</u>. I understand you are completing this project through Lindenwood University and that this study will take place during the fall 2009 semester and part of the spring 2010 semester. As part of the study you will have access to subgroup achievement data in areas such as ethnicity, socioeconomic status, special needs and special programs.
Please contact me with any questions you may have.

Sincerely,

Renée Schuster
Superintendent

Vitae

Travis Bracht is currently the Director of Student Learning for the Francis Howell School District, in St. Charles, Missouri and supervises K-12 curriculum and instruction. Other administrative experiences include serving as Francis Howell's Director of Assessment and Program Development, Associate Principal at Francis Howell High School, and Dean of Students at Francis Howell North High School. Teaching experience includes three years as a science teacher at Francis Howell North High School and four years as a science teacher at Hazelwood Central High School.

Educational studies have resulted in the anticipation of an Educational Doctorate degree awarded upon graduation in May 2011 from Lindenwood University. A Master's in Educational Administration from Lindenwood University was earned in 2002 as well as a Bachelor's in Education from the University of Missouri-Columbia in 1998.