Comparing the Reading Performance of High-Achieving Adolescents: Computers-Based Testing Versus Paper/Pencil

Linda P. Eno
Seton Hall University

Follow this and additional works at: http://scholarship.shu.edu/dissertations

Part of the Curriculum and Instruction Commons, and the Elementary and Middle and Secondary Education Administration Commons

Recommended Citation
http://scholarship.shu.edu/dissertations/1764
COMPARING THE READING PERFORMANCE OF HIGH-ACHIEVING ADOLESCENTS: COMPUTER-BASED TESTING VERSUS PAPER/PENCIL

by

Linda Peet Eno

Dissertation Committee:

Martin Finkelstein, Ph. D., Mentor

Joseph Stetar, Ph.D.

Gerald Babo, Ed.D.

in partial fulfillment of the requirement for the degree of

DOCTOR OF EDUCATION

Seton Hall University

2011
Abstract

Literacy is moving into the digital context. Many of the literacy tasks associated with higher education, the workplace, and civic life now take place in the digital world. Literacy in high school, however, languishes in the text world. This study compared the text literacy of a group of high-achieving 10th-grade students, to their digital literacy. Participants took two standards-based critical reading tests: one paper and pencil (PPT) and one on the computer (CBT). The students also took a norm-referenced text assessment, the Preliminary Scholastic Achievement Test (PSAT).

The analysis compared the mean scores of the standards-based tests using a paired-samples t test with mode (text vs. digital) as the independent variable, and critical reading as the dependent variable. One-way ANOVA was then used to disaggregate the scores within each mode by a set of seven contextual factors: school attended, gender, preferred mode of reading, time spent leisure reading, time spent on communication technology, order of testing, and prior instruction and assessment in digital reading. One-way ANOVA was used with factor as independent variable and critical reading as dependent variable.

The students, on average, performed better on the CBT than on the PPT, supporting previous research that found high-achieving students to have a positive mode effect from digital context. In all contexts, school attended and amount of leisure reading were associated with significant differences in scores, supporting literature that has shown positive academic influence being derived from higher levels of parent education and increased leisure reading. Gender was associated with significant differences on both of the standards-based tests, but not on the norm-referenced PSAT,
which seemed to speak to the motivational differences between boys and girls. The findings from this study will help school leaders as they seek to instruct and assess all students in the skills of 21st Century literacy.
Copyright © 2011

Linda P. Eno
Dedication

This work is dedicated to the family, friends, and colleagues, who supported me through the past two and a half years, making this degree possible.

To:

my husband and best friend, Peter,
for never losing sight of what really matters.

my parents, Bill and Elizabeth,
for always being proud of me.

my adult children, Ivan and Natasha,
for keeping me humble about myself and unendingly proud of them.

my friends, Meredith, Margaret, and Eileen,
for being to me what only other women can be.

my Advisor, Dr. Martin Finkelstein,
for sticking with me and not wavering in his commitment to good product.

Cohort 13 (Michael, FaithAnn, Tom, Bob, Mark et al),
for being there with inspiration, humor, hard work, and chardonnay, as needed.

and

Dr. Richard Bastian, Melissa Borodunovich, Lisa Catanzaro, and Maria Ferrara of Monmouth University
for their statistical support and insights and encouragement over the long haul.

Thank you.
# TABLE OF CONTENTS

Abstract .............................................................................................................................. ii

List of Figures .................................................................................................................... x

List of Tables .................................................................................................................... xi

## CHAPTER I

**STATEMENT OF THE PROBLEM** ................................................................................. 1

**Introduction** ................................................................................................................ 1

  The Importance of Reading in the 21st Century ..................................................... 3

  The History of Current Reading Policy ............................................................... 4

  Digital Reading ........................................................................................................ 6

  Adolescents’ Functional Literacies ................................................................... 8

**Problem Statement** .............................................................................................. 10

**Purpose of the Study** ............................................................................................. 11

**Research Questions** ............................................................................................... 11

**Hypotheses** ............................................................................................................ 12

**Significance of the Study** ....................................................................................... 12

**Conceptual Framework** ....................................................................................... 14

**Definition of Terms** ............................................................................................ 15

**Summary Organization of the Study** ................................................................. 16

## CHAPTER II

**REVIEW OF THE LITERATURE** .............................................................................. 18

**Introduction** ........................................................................................................... 18

**Functional Literacy** ............................................................................................ 20

**Reading Achievement** ....................................................................................... 24

  Reading Achievement Before A Nation at Risk ................................................... 24
CHAPTER III
METHODOLOGY .................................................................................................... 48

Introduction ............................................................................................................. 48
Study Design ........................................................................................................... 49
Instruments .............................................................................................................. 53
Individual Hypotheses ............................................................................................. 58
Data Analysis ........................................................................................................... 59
  Hypothesis 1 ............................................................................................ 61
  Hypothesis 2 ............................................................................................ 62
  Hypothesis 3 ............................................................................................ 63
  Hypothesis 4 ............................................................................................ 64
  Hypothesis 5 ............................................................................................ 64
Delimitations ............................................................................................................ 66

CHAPTER IV
ANALYSIS OF THE DATA ..................................................................................... 68

Introduction ............................................................................................................. 68
Description of the Sample ......................................................................................... 69
Findings for Hypothesis 1 ....................................................................................... 75
  Preliminary Findings ....................................................................................... 75
  Summary Findings ......................................................................................... 78
Findings for Hypothesis 2 ....................................................................................... 78
  Preliminary Findings ....................................................................................... 78
List of Figures

Figure 1 Trend Data for Reading Assessments, NAEP, Grade 12,
1992-2009 ................................................................. p 32
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Population and Sample of 10th Graders by School</td>
<td>69</td>
</tr>
<tr>
<td>Table 2</td>
<td>Population and Sample of Females by School</td>
<td>70</td>
</tr>
<tr>
<td>Table 3</td>
<td>Distribution of Sample by Independent Variables</td>
<td>72</td>
</tr>
<tr>
<td>Table 4</td>
<td>The Distribution of Text Literacy and Digital Literacy Scores before Transformation</td>
<td>74</td>
</tr>
<tr>
<td>Table 5</td>
<td>The Distribution of Text Literacy and Digital Literacy Scores after Transformation</td>
<td>74</td>
</tr>
<tr>
<td>Table 6</td>
<td>The Distribution of PPT and CBT Scores</td>
<td>75</td>
</tr>
<tr>
<td>Table 7</td>
<td>Mean Difference Between PPT and CBT Scores, Two-tailed T Test</td>
<td>76</td>
</tr>
<tr>
<td>Table 8</td>
<td>A Comparison of Respondents Quartile Positions on PPT and CBT Distributions</td>
<td>77</td>
</tr>
<tr>
<td>Table 9</td>
<td>Mean PPT-CBT Difference by School Attended</td>
<td>79</td>
</tr>
<tr>
<td>Table 10</td>
<td>Mean PPT-CBT Difference by School Attended, ANOVA</td>
<td>79</td>
</tr>
<tr>
<td>Table 11</td>
<td>Mean PPT-CBT Score Differences by Gender</td>
<td>80</td>
</tr>
<tr>
<td>Table 12</td>
<td>Mean PPT-CBT Score Differences by Gender, ANOVA</td>
<td>80</td>
</tr>
<tr>
<td>Table 13</td>
<td>Mean PPT-CBT Difference Scores by Preferred Mode for Reading</td>
<td>81</td>
</tr>
<tr>
<td>Table 14</td>
<td>Mean PPT-CBT Difference Scores by Preferred Mode for Reading, ANOVA</td>
<td>82</td>
</tr>
<tr>
<td>Table 15</td>
<td>Mean PPT-CBT Difference Scores by Preferred Time Spent Leisure Reading</td>
<td>83</td>
</tr>
<tr>
<td>Table 16</td>
<td>Mean PPT-CBT Score Difference by Amount of Time Spent Leisure Reading, ANOVA</td>
<td>83</td>
</tr>
<tr>
<td>Table 17</td>
<td>Mean PPT-CBT Score Difference by Amount of Time Spent on Communication Technologies</td>
<td>84</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Table 18</td>
<td>Mean PPT-CBT Score Difference by Amount of Time Spent on Communication Technologies, ANOVA</td>
<td>85</td>
</tr>
<tr>
<td>Table 19</td>
<td>Mean PPT-CBT Score Difference by Order of Testing</td>
<td>86</td>
</tr>
<tr>
<td>Table 20</td>
<td>Mean PPT-CBT Score Difference by Order of Testing, ANOVA</td>
<td>86</td>
</tr>
<tr>
<td>Table 21</td>
<td>The Effect of School Attended on PPT and CBT Scores</td>
<td>88</td>
</tr>
<tr>
<td>Table 22</td>
<td>The Effect of School Attended on PPT and CBT Scores, Tukey HSD</td>
<td>89</td>
</tr>
<tr>
<td>Table 23</td>
<td>The Effect of Gender on PPT and CBT Scores</td>
<td>91</td>
</tr>
<tr>
<td>Table 24</td>
<td>The Effect of Gender on PPT and CBT Scores, ANOVA</td>
<td>91</td>
</tr>
<tr>
<td>Table 25</td>
<td>The Effect of Preferred Mode of Reading on PPT and CBT Scores</td>
<td>93</td>
</tr>
<tr>
<td>Table 26</td>
<td>The Effect of Preferred Mode of Reading on PPT and CBT Scores, ANOVA</td>
<td>93</td>
</tr>
<tr>
<td>Table 27</td>
<td>The Effect of Amount of Time Spent Leisure Reading on PPT and CBT Scores</td>
<td>95</td>
</tr>
<tr>
<td>Table 28</td>
<td>The Effect of Amount of Time Spent Leisure Reading on PPT and CBT Scores, ANOVA</td>
<td>96</td>
</tr>
<tr>
<td>Table 29</td>
<td>The Effect of Amount of Time Spent on Communication Technologies on PPT and CBT Scores</td>
<td>97</td>
</tr>
<tr>
<td>Table 30</td>
<td>The Effect of the Amount of Time Spent on Communication Technologies on PPT and CBT Scores, ANOVA</td>
<td>98</td>
</tr>
<tr>
<td>Table 31</td>
<td>The Effect of the Order of Testing on PPT and CBT Scores</td>
<td>99</td>
</tr>
<tr>
<td>Table 32</td>
<td>The Effect of the Order of Testing on PPT and CBT Scores, ANOVA</td>
<td>100</td>
</tr>
<tr>
<td>Table 33</td>
<td>Mean Text Reading Scores PSAT and PPT</td>
<td>102</td>
</tr>
<tr>
<td>Table 34</td>
<td>A Comparison of Text Literacy Constructs: Correlation</td>
<td>103</td>
</tr>
<tr>
<td>Table 35</td>
<td>The Effect of School Attended on PSAT Critical Reading Scores</td>
<td>105</td>
</tr>
<tr>
<td>Table 36</td>
<td>The Effect of School Attended on PSAT Critical Reading Scores, Games Howell</td>
<td>106</td>
</tr>
</tbody>
</table>
Table 37  The Effect of Gender on PSAT Critical Reading Scores ............... 107
Table 38  The Effect of Gender on PSAT Critical Reading Scores, ANOVA .... 107
Table 39  The Effect of Preferred Mode of Reading on PSAT Critical Reading Scores .................................................................................. 108
Table 40  The Effect of Preferred Mode of Reading on PSAT Critical Reading Scores, ANOVA ................................................................. 109
Table 41  The Effect of Amount of Time Spent Leisure Reading on PSAT Critical Reading Scores ......................................................................... 110
Table 42  The Effect of Amount of Time Spent Leisure Reading on PSAT Critical Reading Scores, ANOVA ................................................................. 110
Table 43  The Effect of Amount of Time Spent Communication Technologies on PSAT Critical Reading Scores ................................................................. 111
Table 44  The Effect of Amount of Time Spent Communication Technologies on PSAT Critical Reading Scores, ANOVA ................................................................. 112
Table 45  Summary: Contextual Factor Effects by Mode ........................................ 114
CHAPTER I

STATEMENT OF THE PROBLEM

Introduction

Reading is going digital. Adolescents are increasingly likely to face high-stakes events that require them to read sophisticated content in the digital environment. Many states are moving their high-stakes exit exams for high school online, as it is cost effective, more flexible, and provides more timely feedback. Colleges and universities are delivering more courses in a blended or strictly online format for the same reasons. Graduate Record Exams (GREs), Graduate Management Admission Tests (GMATs), and Law School Admissions Tests (LSATs) all reside exclusively in the online world. Additionally, the literacy lives of adolescents have been transacted almost exclusively in the digital environment for the past decade. But the instruction and assessment of reading in public schools, driven by high-stakes testing, continues to reside primarily in the paper-and-pencil world. Thus, there is an increasingly wide gap between the digital literacies that students practice and need, and reading as it is taught and assessed in schools.

Bean (1999) suggested that functional literacies are those that serve purposes in our lives that are other than academic. Drawing on the work of social cultural theorists, Bean described those purposes as extending beyond the cognitive to include the social. Reading achievement, on the other hand, can be understood as the institutional instruction and assessment of reading.

Functional literacy and reading achievement have existed as distinct but related constructs for well over a century. It is common for educational research to attempt to
quantify one or the other of these constructs for the purpose of showing change over time. But changing populations, definitions, and expectations have consistently confounded the interpretation of longitudinal data associated with reading. Adding digital literacy to both sides of the equation has complicated things further.

Little is known about how students read digitally or how their functional literacies in the digital world affect their reading achievement. What is known about how students read in the digital world is largely derived from comparability studies: studies that attempt to design digital equivalents to standardized reading tests. The goal of comparability testing is to have students earn the same score in each "mode", hence proving equality of the instruments. This goal presumes that reading is a single construct, and that variability in results indicates instrument inconsistencies. Based on the inconsistency of findings to date, a different approach would be to suggest that digital literacy and text literacy are fundamentally different constructs.

Building on that premise, this study sought to extend the body of knowledge about students' digital reading, in a direction recommended by Pommerich (2004) and Russell, Goldberg and O'Connor (2003). Their recommendation, based on the inconsistency of findings to date, is for educational programs to test their own students. By testing their own students, districts can change the goal from establishing instrument comparability to learning more about each student's abilities within the different modes.

The use of a high-achieving population builds on the summative work of Peak (2005) and the 2002 work of Clariana and Wallace. Peak found that, when testing involved reading longer passages, students tended to perform less well on computer tests than they performed on paper/pencil. Clariana and Wallace found that students who had
above-average content mastery (in an area of content other than reading) benefited more from computer-based testing than students with lower levels of mastery. Synthesis of these findings suggest the possibility that a group of students, homogenous at above-average reading levels, might actually benefit from reading and assessment in the digital environment.

With their focus on high-stakes testing, American high schools have had little time or energy to measure or incorporate students' functional literacies into the reading curriculum in any significant ways. It is critical, however, that high schools ensure that all students graduate ready to read in the digital world. Additionally, high schools would do well to better inform themselves about the functional literacies of their students, with an eye to leveraging their skills and motivation in this realm.

The Importance of Reading in the 21st Century

Peter Drucker first coined the phrase “knowledge worker” in 1959—a phrase that continues to echo in the literature (Gomez, 2007; Achieve, 2005). Technology and the increasingly global economy have created a job market that requires more sophisticated skills in many areas, but especially literacy. At the same time that literacy demands are rising, opportunities for employment and economic advancement for those with less than a college degree are shrinking. Data presented at the 2005 Governor’s National Education Summit on High Schools indicated that 60% of jobs available at that time required some level of post-high school education, and that that number was expected to rise. A 2007 survey by the National Endowment for the Arts (NEA) had responses from 430 companies. Those companies placed reading as the number one job skill, even for jobs requiring only a high school diploma. Stedman and Kaestle (1986) pointed out that
the level of functional literacy being demanded by the workplace might be overwhelming education’s ability to prepare its students.

Failure to produce a literate populace has economic consequences for both the individual and the nation. Data from the National Education Summit (Achieve, 2005) indicated that, over a lifetime, the median income of high school graduates is 42% higher than that of nongraduates. A college diploma produces earnings that are 67% higher than a high school diploma. Additionally, civic engagement is requiring increasingly sophisticated literacy skills to make informed decisions in areas such as health care and personal finance. Without these skills, individuals and families risk being sidelined in the democratic process and on the path to economic opportunity (Achieve, 2005). The Alliance for Excellent Education (2007) reported that, of all aspects of literacy, reading is the most important.

The History of Current Reading Policy

The 1983 report, A Nation at Risk was commissioned by then-President Ronald Reagan, in response to perceptions of declining student achievement. This report delivered a scathing indictment of public education at all levels. One of the allegations made by A Nation at Risk was that both the functional literacy and the reading achievement of adolescents were in serious decline over the decades of the 1960’s and 1970’s. The report asserted that 13% of adolescents were functionally illiterate. While the degree and duration of that decline (and its associated statistics) have subsequently been challenged, the school reform movement as it currently exists was effectively launched.
Five national educational summits have taken place in the intervening years, the first in 1989. The continuance of these summits has illustrated a sustained commitment by states across the nation to the recommendations made by A Nation at Risk. These recommendations included more stringent graduation requirements: rigorous, measurable standards in core areas; more time dedicated to teaching those standards; and testing student mastery of those standards. The use of core standards and high-stakes testing gained additional impetus through Goals 2000, No Child Left Behind (NCLB) (2001) and the more recent Race to the Top (RTT) (2011).

While high school standards have become more rigorous, their implementation has been weakened by their poor alignment with high school exit exams. High school exit exams have not traditionally required students to demonstrate mastery of the standards at the level they are written. These tests are also traditionally poorly aligned with university and workplace expectations. The 2005 Educational Summit, which focused on high school, cited this lack of articulation as a primary reason high school graduates are so often described as poorly prepared.

Peter Hart, Inc., surveying for Achieve in 2005, said, “Employers were questioned about their satisfaction with the job that high schools are doing in preparing graduates for a number of skills needed in the work force. Forty-one percent of employers say that they are dissatisfied with graduates’ ability to read and understand complicated materials” (p.6). In addition, the Achieve report cited 70% of college professors as being dissatisfied with the job public schools are doing in preparing students to read and comprehend complex materials. It estimated that institutions of higher learning spend between one and two billion dollars a year on remediation of students. Businesses incur
costs, as high as sixteen billion dollars annually, through the lost productivity and skills remediation associated with under-prepared students.

**Digital Reading**

Delivering high school literacy instruction through standards has been complicated by the rapid rise of technology. The computer, on and offline, is central to the dissemination, retrieval, and assembly of the knowledge required for literacy in the 21st Century. It would be difficult to conceptualize a relevant set of standards for high school literacy without referencing digital literacy (Balajthy, 2007). Digital literacy includes reading, searching, assimilating, filtering, and assembling knowledge in the online environment (Bulger, 2006). Latham (2008) described these skills as essential to the personal, academic, and professional success of 21st Century learners. Gomez (2007) added that the need for students to leave high school proficient in online reading becomes more pressing as higher education increases its use of blended learning and online databases, and as textbooks, manuals and high stakes testing are increasingly found in the online environment.

Yet, O'Brien and Scharber (2008) found that digital literacy does not typically reside within the language arts curriculum. Digital literacy standards are more likely to be an add-on, not embedded in specific subject standards. Digital literacy is often framed within technology standards that are designed to be integrated by all teachers, and hence owned by no one. The work by Latham and Gross (2008) confirmed that the delivery and assessment of the technology standards continues to vary widely from school to school, and assessment of online reading and response is rare.
There will be increasing demand, and opportunity, for students to read digitally, as more content is stored and delivered electronically (Dziuban, 2004). There are a host of reasons for the shift. Online textbooks are lighter, less expensive, and do not depreciate like paper text. Online courses are cost effective and adapt more readily to the range of learners now populating university classrooms. Digital databases eliminate the space requirements of text libraries and stay current without the high cost of replacement. In addition, moving online with high stakes testing reduces costs, adapts better to block scheduling, and provides timelier turnaround of test results.

How students perform when reading digitally should be an essential question in 21st Century literacy programs (The National Endowment of the Arts, 2007; Bauerlein, 2009), since digital reading differs from text reading in several ways. To describe one aspect of the differences, Bulger (2006) employed the term *affordance*, first introduced by Gibson in 1979: “Affordance refers to the characteristics of an object that make it useful. Paper documents, for example, allow us to write notes in the margins, spread multiple pages out on a desk to visualize the information, and visually manipulate the page order” (p.3). She cited a 2003 study of adults at work that showed that 89% of them edited documents in paper and pen format alongside the digital document—a clear nod to affordances. Peak (2005) pointed to mode effects such as scrolling, page turning, and legibility as additional factors that can decrease satisfaction with, and success in, digital reading.

There is a body of work that measures reading achievement in the digital environment compared to reading achievement in the paper/pencil environment. The studies are known as comparability studies, and Peak (2005) described most of the earlier
comparability studies as having been set up to evaluate the technology that was assessing the learning, with measurements of student performance and preference being by-products. While some more recent studies have maintained this focus, others have sought to explore factors associated with the student that may influence their success, or lack thereof, when testing in the digital realm—factors including gender, ethnicity, competitiveness, and level of achievement.

**Adolescents’ Functional Literacies**

Federal policy has acknowledged the primacy of technology by subsidizing its expansion in the public schools. Internet access has been subsidized for public schools through the E-Rate since 1998. The computer-to-student ratio has risen alongside this investment (Bauerlein, 2009). Schmar-Dobler (2003) provided Census Data from 2001 that described 98% of the K-12 classrooms as having access to the Internet. K-12 education has arrived at the place where access to technology is no longer the biggest barrier to high-level digital reading.

Students’ home environments, like 21st Century schools, are well-equipped to support their functional literacies. The 2003 Census provided data stating that 47.9% of all 12- to 17-year-olds in the United States had access to a home computer. By 2008, data from Nielsen’s National TV panel found that around 80.6% of homes in the US possessed a computer. In addition, today’s adolescents make use of an extensive array of communication technologies such as iPads and smart phones that broaden their information and social networks. These communication technologies form the foundation for the new functional literacies (i.e., tweeting, texting, blogging, and instant messaging) of adolescents in the millennium.
Research is exploring the interface between these communication technologies and the formal education of adolescents. Two compelling reasons for this exploration emerge from the literature. First, Bauerlein (2009) pointed out that these technologies have created such a fundamental shift in communication that they may be affecting adolescent processing. These technologies allow information to flow at an unprecedented rate—much of it reaching adolescents without ever being filtered, or even shared, by an adult. Second, they are so ubiquitous that they are affecting students' expectations, motivation, and leisure-time habits. The more time they spend in front of a screen, the less time they spend doing other things, including text reading.

Skill standards associated with online literacy is a step toward integrating adolescent functional literacies. Bean (1999) and Alverman (2008), however, resisted definitions that establish the flow of technology as being from school to student. These researchers encouraged schools to look for ways to bring the functional literacies that students practice in their social lives into the classroom, in order to promote engagement and deepen learning. These outcomes occur for students when they see school learning as connected to their lives. These authors further contended that inclusion of these functional literacies would provide teachers with greater insights into who their students are in the broader context of their lives. Connected, engaged learners perform at higher academic levels.

From implementing standards to building a curriculum that instructs, assesses, and engages 21st Century literacy learners, there are challenges and choices for each district. Education needs to continue the research that seeks some marriage between what our students currently know and are able to do, the technologies that increasingly
consume their time and attention, and the research-based evidence about what improves reading achievement. Policy and practice must work backwards from rigorous, relevant standards and individually appropriate outcome assessments, to design, deliver, and sustain 21st Century literacy for every student.

Problem Statement

In 1995, Elizabeth Birr Moje described adolescent literacy as a complex construct, influenced by home, peer group, and the multiple texts to which adolescents are exposed. Literacy of the millennium, and its assessment, has been further complicated by the proliferation of technology. Technology has driven many applications of functional literacy into the digital world, including social, civic, and workplace activities. But despite the rapid expansion of functional literacy into the digital world, reading in schools has languished in the text world.

Over the past 30 years, the focus on high-stakes testing has forced an equating of literacy with reading achievement (first norm-referenced and, more recently, standards-based). There is little opportunity to explore the discrete but related functional literacies of adolescents. Current policy is creating an ever-widening gap in schools’ knowledge about the literacy lives, and the functional literacy skills, of students.

The demands on literacy are rising, driven by globalization and technological advances. Literacy and adolescent learners are being rapidly and radically reshaped by technology. Digital literacy is fundamentally different from text literacy; it is not simply text literacy repackaged. In order to ensure that all students leave high school ready for the literacy challenges that lie ahead, schools must instruct and assess in both modes.
Purpose of the Study

The purpose of this study was to explore the distinct, but related, skills of text literacy and digital literacy. First, the study examined whether students’ reading achievement differs by mode (text versus digital). Second, the study explored whether factors affecting achievement in each mode are similar or different. Finally, the study sought to determine whether the factors that are significant for a norm-referenced assessment of text literacy are the same or different for a standards-based assessment.

Research Questions

The research questions are described below.

1. Are reading scores different, on average, on a standards-based assessment of text literacy than on a standards-based assessment of digital literacy, for high-achieving 10th grade students?

2. Are reading scores different, on average, on a standards-based assessment of text literacy than on a standards-based assessment of digital literacy, for high-achieving 10th grade students, when controlling for those contextual factors that the literature suggests may mediate the effect of mode, including: a) school attended, b) gender, c) reading preference (text or digital), d) amount of time spent leisure reading, e) amount of time spent on communication technologies, f) order of testing, and g) prior instruction and assessment in digital reading?

3. Within one literacy mode (text or digital), are reading scores significantly different, on average, for high-achieving, 10th grade students of different gender and reading habits, when controlling for the selected contextual factors above?
4. Within the arena of text literacy, are mean reading scores on a norm-referenced test (Preliminary Scholastic Achievement Test [PSAT]) related to mean reading scores on a standards-referenced test?

5. Within the arena of text literacy, do the set of selected contextual factors (v. supra) influence reading scores on a norm-referenced test (PSAT) in the same way as scores on a standards-referenced test, for high-achieving 10th grade students?

**Hypotheses**

This study was built on the conceptual framework that text literacy and digital literacy are fundamentally different constructs, due to differences in the skills associated with each, as well as the purposes students attach to each literacy and their associated motivation relative to each. This study tested the null hypotheses that correspond to the research questions above.

**Significance of the Study**

High school students, particularly the college bound, are increasingly likely to face high-stakes events that require them to read sophisticated content in the digital environment. Yet, high schools do little to instruct their students in digital reading, and even less to assess their mastery of this skill before graduation. One method of testing adolescents as digital readers is comparability testing, which is testing done to demonstrate the comparability of instruments across modes. In this type of testing, when the scores are not the same, the differences are suppressed through a process called *equating*. So, while this study has been guided by comparability testing, it differs in that
it assumes that student scores will differ across modes and seeks to learn from these differences.

The research findings related to digital reading have been inconsistent, but this study sought to build on three findings. First, that students, on average, have performed less well on computer tests when longer reading passages were included. (Peak, 2005). All assessments in this study included multiple paragraph readings. Second, that the negative mode effect may not be present for students who possess a high level of reading competency, as academic competency has been shown in several cases to mitigate mode effect (Clariana & Wallace, 2002). The sample for this study is homogenous for high achievement. And, finally, that results are likely to be test-specific and difficult to generalize, leading to the need for programs to do their own testing (Russell et al., 2003).

Russell et al. pointed out that, with norm-referenced testing on the decline and criterion-referenced (or standards-based) testing on the rise, there will be greater opportunity to move away from uniform conditions for all students during testing. Russell et al. raised the notion that text and digital literacy are different skill sets, for which students may have varying degrees of aptitude and affinity. They suggested that schools test for the purpose of determining each student’s optimal literacy, and then, much like accommodations, allow each student’s success to be maximized by testing within that literacy. This study is significant for administrators looking for models, as they seek to assess the skills and motivation of their own students as digital readers. Additionally, it adds to the body of work that is advising policy makers and practitioners about the needs and abilities of adolescents as digital readers.
Conceptual Framework

This study is aligned with two conceptual frameworks. The first is that there are different, but related, literacies. Stedman and Kaestle (1985, 1986) conceptualized this relationship by placing functional literacy on a horizontal axis and reading achievement on a vertical axis and examining the contextually changing intersection. This model worked well for this study, as digital literacy is a predominantly functional literacy, used widely but not taught or measured in schools. Text literacy on the other hand, can be viewed as synonymous with reading achievement.

The second concept is that there is a growing gap between the functional literacies practiced by adolescents and demanded by the real world, and reading achievement as currently defined and assessed by high schools (Bauerlein, 2009; Alverman, 2008). Adolescents’ functional literacies now reside primarily in the digital world, while reading achievement (instruction and assessment) continues to exist almost exclusively in the paper/pencil world. The high-stakes reading tests required of public education today largely drives this. As long as these tests reside in the paper/pencil world, so will reading instruction and assessment.

There is not a universal interpretation of the impact of this gap. Alverman (2008) saw it as posing a threat to student motivation and a risk to the relevance of the high school literacy curriculum. Bauerlein (2009) saw it as necessary to stave off the “dumbing down” of American students being brought about by their obsession with communication technologies. More research is needed to better understand the relationship between adolescents’ functional literacies and their reading achievement.
Definition of Terms

Adolescent: For the purpose of this study, the term adolescent is understood to refer to students in grades nine through 12. The age range, on average, was between 14 years and 18 years old. Features of adolescence that come to bear on literacy include, but are not limited to, emerging identity issues, increased agency, and escalating academic demands in content area reading and writing (Bean, 1999).

Communication technologies: Socially motivated, multimodal ways of exchanging information (e.g., language, imagery, sounds) including, but not limited to, texts, instant messages, Facebook walls, blogs, wikis, and webpages (Alverman, 2008). This definition does not mean to imply that these technologies do not have application in academic settings. Rather, it reflects the research that finds that students’ primary functional interface with these technologies is social.

Digital literacy: The necessary skills to complete academic tasks using the computer. These include, but are not limited to, reading, searching, assimilating, filtering, and assembling knowledge (Bulger, 2006). This skill set can be taught and assessed, and some of the variables in this study were derived from this skill set. This definition does mean to imply that these skills do not have application outside of the classroom. Indeed, these skills are acknowledged to be essential in the management of adult life, and in civic engagement, the workplace and higher education. Rather, it reflects the research that finds that adolescents’ primary functional interface with this skill set is academic.

Functional literacy: Functional literacy is defined as literacy employed to serve a purpose. As defined by Bean (1999), it includes the value, purpose, and agency that are
assigned to the task, which may be social, academic, cocurricular, or other. For the purpose of this study, adolescent use of communication technologies and digital reading were included under the umbrella of functional literacy.

**Literacy**: Literacy has an evolving definition that is largely context-dependent. Reading is a component of literacy—necessary to, but not sufficient for, its definition. For this study literacy was defined as "the ability to read and write critically in content and functional areas using both text and digital medium". Critical levels will be defined as "summary", "analysis" and "synthesis" (Horner, 2007).

**Reading achievement**: Based on the review of the literature, reading achievement is understood to refer to critical reading scores as measured by assessments including, but not limited to, State Assessments, Scholastic Achievement Tests (SATs), American College Tests (ACTs), Advanced Placement (AP) and International Baccalaureate (IB) exams. It is understood that these assessments continue to reside largely in the text environment.

**Summary Organization of the Study**

This dissertation is delivered in five chapters. In Chapter I, the Statement of the Problem provided an introduction to the subject and background information. It provided a problem statement, research questions, and definition of terms. Chapter II, Review of the Literature, presents relevant studies on text and digital reading and assessment, including recent comparability studies, as well as current research on adolescent literacy practices. Chapter III, Methodology, introduces the research design. It describes the population, instruments used, the method of analysis, and limitations/delimitations. Chapter IV presents the data and an analysis of the data in light of the research questions.
Chapter V summarizes the findings of the study, draws conclusions, and makes recommendations in the areas of policy, practice and further research.
CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The following review of the literature provides context for this study examining the success of high-achieving adolescents’ reading in a text environment versus their success in reading in a digital environment. While there is no conclusive evidence on how it is best to test students, or which students will perform best in each environment, there is voluminous data on adolescent literacy. This chapter is structured in accordance with the theoretical framework that views literacy in the categories of functional literacy and reading achievement.

The chapter begins by reviewing the literature on adolescent functional literacy. The literature is laid out to follow the evolving definition of functional literacy in order to better demonstrate the role that definition plays in measurement outcomes. Next is an overview of the literature on adolescent reading achievement. To bring some order and selectivity to this section it is organized around two questions: (1) On what data were the claims of A Nation at Risk based?, and (2) How has reading achievement fared in the era of school reform? These questions were selected for their ability to highlight the special difficulty of comparative analysis in reading and to illustrate the link between research findings and policy. Studies reviewed include, but are not limited to, Stedman (1987), Stedman and Kaestle (1985, 1986), Gadway and Wilson (1976), and Harris & Associates, Inc. (1970). Selected millennial studies broaden the context of comparison by putting the reading skills and challenges of today’s adolescents into an international context. Data is drawn from sources including, but not limited to, the

This chapter, then, reviews findings from other studies that explore digital reading. These studies fall predominately in the category of comparability studies—studies done to ascertain whether the digital version of a test is comparable to the paper/pencil version. Many of the recent comparability studies have been commissioned by states that are moving their high-stakes testing out of paper/pencil and into a digital environment. Early comparability work exists, but little of the work done prior to 1996 will be discussed here. Changing conditions, including increased access to technology, enhanced familiarity with technology, and improved technology, render the results of the early studies less relevant. Comparability data are drawn from studies including, but not limited to, the summative work of Peak (2005), the experimental works of Macedo-Rouet (2009) and Clariana and Wallace (2002), and work commissioned by states, including Texas (2008, 2006), Florida (2006), and Oregon (2002).

The chapter concludes with an investigation of literature related to 21st Century adolescents’ use of communication technologies. The purpose is to examine students’ use of communication technologies as a functional literacy, and to examine some of the research describing the impact of these functional literacies on reading achievement. Literature includes, but is not limited to, Bean (1999), Prensky (2001), and Bauerlein (2009).
Functional Literacy

Since 1985, much of the policy and practice related to adolescent literacy has been driven by the conclusions reached, and recommendations set forth, in the Commission for Educational Excellence’s 1983 report, A Nation at Risk (1983). A Nation at Risk, commissioned by then-President Reagan, investigated claims of declining student achievement throughout the decades of the 1960’s and 1970’s. This report made the following bold assertions in the area of functional literacy: “About 13 percent of all 17-year-olds in the United States can be considered functionally illiterate. Functional illiteracy among minority youth may run as high as 40 percent (p. 3)”. But measures of functional literacy varied widely, depending on how functional literacy was defined, the instrument used to measure it, and where the cut point was set. Three different approaches to defining and measuring functional literacy are described below, with an introduction to the corresponding data.

Stedman and Kaestle (1986) pointed to the initial definition and measurement of functional literacy as derived from the US Census. The 1840 Census defined functional literacy as “the ability to read and write a simple message as measured by self-reporting”. In 1870, 20% of the population self-reported as functionally illiterate. Subsequent census data showed steady improvement in functional literacy rates, up through the latter half of the 20th Century (census data is collected and analyzed every 10 years). In 1979, only 0.6% of adults claimed functional illiteracy. In the age group of 14-24, the percentage of persons self-reporting functional illiteracy was even lower: 0.19% (p.11). This assessment of functional literacy presented data that is significantly
different from that presented by A Nation at Risk. The biggest threat to the validity of the census data is its collection instrument—self-reporting.

Stedman and Kaestle (1986) credited the Civilian Conservation Corps (CCC) with advancing a different definition of the term **functional literacy**, one that initially resulted in similar findings. The CCC defined functional literacy by educational attainment, and in the 1930’s they set the completion of grade three as the line of demarcation between literacy and illiteracy. The Census picked this up and began collecting data on level of educational attainment in 1940. Using the CCC’s definition and a model that allowed for backwards forecasting, Stedman and Kaestle confirmed that, by the late 1970’s, outright illiteracy as a percentage of adolescents was very small—0.7%, according to their model. This percentage was similar to the census findings on adult functional literacy.

However, the location of the line of demarcation for functional literacy changes over time and varies by demographic setting, complicating the attempt to create trend lines and definitive comparisons. Technology and globalization, among other factors, have raised the bar on the functional literacy in the decades since the CCC put out their measure. Stedman and Kaestle (1986) pointed to researchers such as Hunter and Harmon (1979) and Carroll and Chall (1975), who moved the line of demarcation for functional literacy to the level of a high school diploma. Stedman and Kaestle asserted that, applying the changed definition to the same data, the outcome actually indicates declining functional literacy. “In 1930, about 88 percent of the population had a third-grade education or more; ...and in 1980, 68.7 percent had completed high school”
Hence, functional illiteracy for the population would have been just above 31%. This would be closer to the minority student figure claimed by A Nation at Risk.

A third approach to defining and quantifying functional literacy is found in functional literacy tests—tests that assess a person’s ability to complete literacy tasks associated with daily living. Stedman and Kaestle (1986) marked Dr. Guy Buswell’s 1937 study out of the University of Chicago as the first test of this sort. They described a subsequent gap in the use of such testing, followed by an uptick in the 1970’s. Results are shown to vary widely, depending on what skills are being tested, the instrument used, and the scale for distinguishing literacy from illiteracy. Stedman and Kaestle summarized five of those studies in the table below (p.28). The findings for functional illiteracy range from a low of 3% to a high of 53.6%. The populations included compromise the value of these studies to a longitudinal discussion on adolescent functional literacy, as four out of the five studies reviewed included adults in their population. However, it is the other methodological variations that truly complicate the ability to draw definitive conclusions about functional literacy from these tests.

As stated above, these studies differed in their selection of a population, with only the Mini-Assessment of Functional Literacy (MAFL) limiting itself to an adolescent population. Additionally, the studies varied in how they defined functional literacy. The MAFL tested a hierarchy of skills along the vertical axis of reading achievement, including “understanding word meanings, gleaning significant facts, comprehending main ideas and organization, drawing inferences, and reading critically” (Gadway & Wilson, 1976, p. 1). The Adult Functional Reading Test concentrated on
day-to-day-tasks in major life categories, including reading a train schedule, identifying hazardous liquids, and reading employment applications (Murphy, 1975). Finally, the criterion or cutoff for demonstrating functional literacy varied. The MAFL set a minimally adequate performance level at 75%, as determined by the Right-to-Read Committee that commissioned the tests (Gadway & Wilson, 1976). Anyone earning less than 75% was categorized as functionally illiterate. So, while Gadway and Wilson reported significant gains in 14 of the 20 subgroups that took the three versions of the test given in 1971, 1974 and 1975, the selected cutoff put 13% of adolescents into the category of functional illiteracy—the statistic highlighted in A Nation at Risk. Stedman and Kaestle (1986) pointed out that setting the criterion at 60% drops the percentage of functional illiteracy to 2.9% (p.53).

A measure of current adolescent functional literacy is provided by NAEP’s 2004 trend line data report. NAEP provides scale scores on reading achievement, but it also categorizes student scores by performance. The performance level of a student describes their functional reading skills and is delineated in 50-point increments. The lowest is 150 (carry out simple reading tasks), and the highest is 350 (learn from complicated material. For the 17-year-olds, only levels 250, 300, and 350 are used. Eighty percent students achieved a level of 250 (interrelate ideas and make generalizations), statistically unchanged from 1971 and down from a 1988 high of 86%. Thirty-eight percent of students achieved level 300 (understand complicated information)—statistically unchanged from 1971, although down from a 1993 high of 43%. Only 6% of students achieved at the 350 level, statistically unchanged from 1971.
Reading Achievement

Adolescent reading achievement has been measured since the latter part of the 19th Century. Since the middle of 20th Century, the literature has been reporting rising workplace and civic demands on reading, while reading achievement scores among adolescents have been stagnant or declining. Publications such as Flesch’s, Why Johnny Can’t Read (1955) and A Nation At Risk (1983) have fueled public ire and laid the blame squarely on the shoulders of public education. The data behind these reports is important. Closer inspection reveals the challenges of drawing valid and reliable conclusions about reading. To better understand the current status of reading achievement, this chapter reviews literature selected for its ability to assist in answering two questions:

1. What data supported the claims of a 20th Century decline in adolescent literacy?

2. Has adolescent literacy improved in the decades of school reform?

Reading Achievement Before A Nation at Risk

Evidence of falling reading achievement is historically drawn from one of two types of comparisons—then-and-now studies and trend line studies. Then-and-now studies are those in which a test was given to a group of students in a particular locale, and then, some years later, was compared to the results of the same or an equivalent test being given in the same locale to the current group of students. Studies of this type were largely used to support the claims of a steady rise over the first half of the 20th Century.
Charting nationally representative students’ results, on a particular test at intervals over time, creates trend line data. Trend lines of students’ reading achievement have been used to support the argument of a significant drop in achievement at some time during the 1960’s and/or 1970’s, the claims of generally stagnant national scores in the millennium, and the loss of ground internationally. SAT, NAEP, and Program for International Student Assessment (PISA) all create trend line data.

Farr (1974) undertook a longitudinal analysis of reading achievement, commissioned by Educational Testing Services, examining the factual underpinnings of the claims that many adolescents did not possess adequate literacy skills and that literacy skills were declining (p.10). As a part of this study, Farr reviewed 13 then-and-now studies that took place over the first half of the century. The criteria for inclusion were that the study measure reading at two points in time, and that the method of measurement be clearly described. The time frame for the tests ranged from 1916 to 1963. The studies included grades ranging from 1st through 11th, and had populations ranging from 51 students to 252,000 students.

Of the 13 studies, 12 showed some degree of gain. In many cases, however, these gains were not significant, or significance was not reported. Farr found this to be lukewarm support for significant change supported by three other summaries of then-and-now studies: Witty and Coomer (1951), National Education Association (NEA) (1952, 1951), and Geberich (1952). These studies reported summative findings of student reading achievement at the mid-century as being not worse, and only possibly better than, in earlier years, mitigating the theory of a golden age of literacy during the
first half of the century (a premise that must be accepted in the most assertions of a mid-century decline).

In 1986, Stedman and Kaestle undertook a longitudinal analysis of reading achievement, with the purpose of analyzing the data that substantiated claims made in A Nation at Risk. Both Farr (1974), and Stedman and Kaestle, cited methodological problems with the then-and-now studies that made it difficult to draw valid conclusions, or to generalize their finding. The first issue was created by methodologies that failed to create comparable populations. Many studies did not correct for demographic changes in the locale where the assessment was given. The widespread demographic changes in those decades made it very likely that this lack of correction impacted the validity of the findings. The instruments used to measure reading achievement created the second issue. The instruments varied widely, and often tested skills other than reading. The way that changing policies and laws, including but not limited to those regulating admission, attendance and promotion, confounded variables created the third issue. These changes caused more students, and different students, to be in school longer, and created grade levels of students that were younger in more recent decades than their counterparts in earlier decades. Linked to that was the final issue of the inconsistent application of the variables of age and grade. The only two studies cited by Farr that included high school students will be described in more detail for the purpose of illustrating the problems discussed above.

A 1948 study by Krugman and Wrightstone (as cited in Farr, 1974) compared the 1935-1941 New York City Nelson-Denny results for 9th graders and 11th graders to the 1944-1946 results. The 9th grade population had 20,467 students in the earlier test,
and 13,702 in the latter. The 11th grade population had 29,319 students in the earlier test, and 21,252 in the latter. The test results in each era were reported out against the national norms for the time. The former 9th graders were reported as one month above the national norm, while those in the latter test were reported as four months above. Eleventh graders in the earlier testing were reported as two months above the national norm, while those in the latter were one month above. Krugman and Wrightstone (as cited in Farr & Rogers, 1974) reported these findings as positive, but acknowledged they were not significant. Population shifts and age versus grade were not discussed.

In 1949, Tiegs (as cited in Farr, 1974) measured changes in the mean and median scores on the Stanford and Progressive Achievement tests before 1945 and after 1945, which he reported out in six separate tables. The population totaled approximately 230,000 students (grades 3 through 11), drawn from 60 communities in seven states. Total reading achievement for high school compared 2,570 10th graders from before 1945 to 2,613 after. It also compared 250 11th graders pre-1945 to 266 post-1945. The results are quantified in gains or losses of months, with 10th graders gaining .64 months and 11th graders showing no change. The study overall showed a gain in mean reading achievement of 1.8 months. Tiegs concluded that, while the gain was probably not significant, it at least signified that the performance of public school students was not declining. Farr asserted that the lack of clarification on the number of years that spanned the first and second testing rendered the interpretation of the findings so unclear as to make it not worth going further with the assessment of limitations. Two other studies (drawing their sample from grade 6) will be examined for the efforts made by the researchers to overcome the shortcomings of prior then-and-now studies.
Finch and Gillwater’s 1949 study (as cited in Farr, 1974) compared 144 sixth-grade students in 1931 to 198 sixth-grade students in 1948. Both groups took the Thorndike-McCall Reading Scale, Form 3. Finch and Gillwater corrected for age, noting that in 1931 the students were 1.56 months older than they were in 1948. They also corrected for socioeconomic status by matching students through their fathers’ occupations. There was a .78 rise in the mean score for the 1949 group. Standard deviation for the earlier group was 4.02, and for the latter group it was 6.32. Finch and Gillwater did not report any type of statistical analysis, but concluded that the difference served as reasonably good evidence of improvements in teaching methods. However, when Farr and Coomer’s team ran a t-test on the data, they found the difference to be not significant. An additional weakness of the study was that it did not address the possible lack of relevance of the test material to the latter group of students.

In 1980, Elligett and Tucco (as cited in Stedman, 1987) performed another study that corrected for issues previously identified. Their study was a districtwide assessment of approximately 7,300 sixth-grade students in Pinellas County Florida. The purpose was to determine whether sixth-grade grade students in 1978-79 were reading as well as sixth-grade students read in 1955. “Each student in the study took both the current edition and the 1950s edition of a major standardized test. Their performance on these two tests was then rated according to the national norms reported for these tests at the time of their publication” (p.698). This study overcame a weakness of prior then-and-now studies by using only one group of students—hence, eliminating compositional changes. Elligett and Tucco found losses of between 5 and 10 months, and concluded that “current sixth-graders are reading only as well as students who were
one-half to three-quarters of a grade level lower in the 1950s” (Stedman, 1987, p.699). Interestingly, below-average students were actually doing as well or better than their 1950’s counterparts. Above-average students, on the other hand, demonstrated losses of between three and twenty-seven months. Weaknesses of the study included failing to address the issues of student motivation and content relevance of the 1950’s test for the latter group of students.

In 1985, Stedman and Kaestle reviewed eight state or national then-and-now studies to help answer the following question: Did adolescent reading undergo a catastrophic decline during the decades of 1960’s and 1970’s? The studies included grades ranging from first grade to first year of college. Population size ranged from just over 1000 to 15,000, with n for some studies being indeterminate, as the number of schools included, or some other general descriptor, defined n, rather than a sample count. Four of these studies reported increases and four reported decreases.

The 1979 study by Farr, Fay and Nagley (as cited in Stedman and Kaestle 1985) is detailed here for its representative conclusion regarding adolescent reading achievement. The study compared test results for 10th grade students from Indiana in 1944 to results for a similar population in 1976. The instrument was the 1944-945 Iowa Silent Reading Test at each administration. The 1944 population was drawn from volunteer schools, and consisted of 11,424 students (25% of the state’s students at that time, nonrandomized and not stratified). The population in 1976 was 7,000 students from 31 high schools, representing 6% of the student population. Selected students were stratified and randomly selected. The process was repeated in 1986 with another
randomized, stratified population of 8,032 students from 40 schools, representing 10% of the 10th grade student population.

These authors concluded that, once they adjusted for age difference, the test scores in 1986 did not differ significantly from the scores earned four decades earlier (Stedman & Kaestle, (1986) p.29). Stedman and Kaestle (198) drew a similar summative conclusion. They wrote that the data, when corrected for age and other compositional factors, showed that, “school children of the same age and socioeconomic status have been performing at the similar levels throughout most of the twentieth century” (p.18).

An important exception to this conclusion is the period from approximately 1966 to 1979. Authors, including Copperman (1979), Stedman and Kaestle (1986), and Munday (1979), have agreed that an achievement decline occurred in those years, although they differed in their description of its timing, duration, and causes. This decline is important because of the conclusions that people drew from it and the subsequent policy decisions that were made.

Leon Copperman, author of the Literacy Hoax, and then-President of the Institute of Reading Development in California, documented a decline in reading achievement starting in the mid-1960’s, and going through the end of the 1970’s. In his 1979 article, The Achievement Decline of the 1970’s, Copperman pointed to SAT and ACT scores as evidence of a decline in the latter half of the 1960’s. He described SAT verbal scores as having peaked in 1963 (478), and then dropping to 460 in 1970 and 429 in 1978. He translated this as a drop of almost 4% of a Standard Deviation (SD)/year (p.736). Copperman described a decline in the verbal ACT starting in 1966, which,
while less consistent than the SAT declines, averaged 2% of a SD/year from the late 1960’s through the 1970’s. Copperman pointed out that students who take the SAT are more representative of the Northeastern American adolescents, while those that take the ACT are more representative of Central and Western American adolescents (p. 736).

Copperman provided the following data as evidence of a decline in the 1970’s:

Eighth-grade students’ scores declined 3% of a SD each year, for a 10-month loss between the fall of 1970 and the fall of 1978 on the Metropolitan Achievement test (MAT). Tenth-grade students lost 15 months during that same time period. The California Achievement test (CAT) test data between 1970 and 1977 showed a three- to six-month loss for the eighth graders and an 11- to 30-month loss for 11th graders (p.737). Editorial comment from Phi Delta Kappan and Stedman and Kaestle (1985) indicated that Paul Copperman attributed these declines exclusively to liberal educational practices, including but not limited to, straying from rigorous core curriculum, a lack of classroom management, and grade inflation.

While Stedman and Kaestle (1985) agreed that a mid-century decline in a reading achievement occurred, they disagreed on timing, which they believed supported their alternate theory on cause. They discredited a significant decline in the 1960’s, since that evidence relied heavily on changes in SAT scores. Stedman and Kaestle pointed out that Copperman’s argument regarding the drop in SAT scores was predicated on the validity of the rise in scores attributed to the preceding decade (9%, according to College Boards). Using Schrader’s 1968 analysis of test equating, they showed that the SAT equating (1957 and 1967) included private school students that had previously been excluded. Hence, they discredited the validity of the 9% rise. The
equating of the PSAT (1960 and 1966) and the Iowa Tests (1957 and 1962), both of which were consistent in the types of schools they included, showed stable trend lines at high school during that time, supporting Stedman and Kaestle’s assertion. Additionally, they cited population in general as a limitation of the SAT. Because students self-select to take it, they cannot be assumed to be representative of high school students in general, and findings cannot be assumed to be representative. Finally, compositional changes in the population taking the SAT were not accounted for in Copperman’s analysis. Stedman and Kaestle believed these changes accounted for approximately 66% of the score differential in SAT scores.

When looking at the achievement decline of the 1970’s, Stedman and Kaestle (1985) referenced several demographic factors that exerted significant influence on the test scores of that decade (p.205). Family sizes increased during the 1950’s (the Baby Boomers were born), which dropped test scores because first- and second-born children tend to score higher than subsequent children. The number of immigrants increased, which raised the percentage of students who were English Language Learners (ELL). “Asian and Hispanic students raised the percentage of minority students from one-sixth to one-fourth” (p. 205). The dropout rate decreased, leaving more low-achieving students in school. Lower-achieving students began aiming at technical and community college, thus taking College Readiness exams, such as the SAT and ACT. Finally, policies such as automatic promotion and younger school-admission ages changed the composition of test takers, relative to prior decades. When all of these factors are combined, Stedman and Kaestle believed they had accounted for between 30 and 50% of the decline in standardized achievement test scores of that decade. They pointed out
that researchers such as Copperman attributed 100% of the cause to instructional practices.

As an historical note, the Sandia Report, published 1993 by Carson, Huelskamp, and Woodall (as cited by Stedman, 1994), attributed 100% of the decline to compositional changes. It stated that SAT scores rose in all subcategories during that decade, and that the average fell despite these rises. They described this as a statistical phenomenon known as Simpson’s paradox, where the proportion of the students at the top fell, causing a decline in the average, despite improvement in subcategories. Stedman (1994) rebutted the Sandia Report, pointing out that students are categorized into quintiles and that the percentage of students taking the test from the bottom two-fifths of their class rose between only 2 and 5% in the years between 1976 and 1992 (not a uniform rise). Stedman (1994) pointed out that the technique of reporting an SAT decline as an average percent is masking the verbal decline. He agreed with the Sandia Report’s assessment of NAEP math and reading as virtually unchanged, but pointed to the drop in science as further evidence of an actual decline.

**Reading Achievement in the Era of School Reform**

Most major tests showed reading achievement to be relatively unchanged throughout the latter decades of the 20th Century. For instance, the 2000 SAT Score Report described a rise in the verbal scores in the early 1980’s, and then a subsequent decline, with verbal scores ending the century 3 points above the 1980 mean score of 501, although not fully recovered to the 1972 level of 529. NAEP data for reading remained similarly stable.
NAEP is an unusually reliable provider of reliable longitudinal data, as it is collected by the federal government and is consistent in methodology of collection and analysis. Reading data is collected using two separate instruments, long-term trend assessments and main assessments. In 2004, NAEP released a longitudinal study that examined reading scores for 38,000 students, collected from 1971 to 2004. The 2004 report provided nationally representative reading performance data as measured by the long-term assessment starting in 1971, and by the main assessment starting in 1991. Reading data is scored on a 500-point scale and is presented in three different ways: average scale scores, percentile scores, and performance levels. Performance levels, set in increments of 50, help link reading achievement with functional literacy. At the top performance level of 350, students are described as being able to learn from specialized reading material. At the lowest level for high school, 150, students are described as being able to complete simple, discrete reading tasks.

Between 1971 and 2004, 17-year-olds, on average, remained statistically unchanged (despite a bubble in the late 1980's). The 2004 NAEP report also provided some interesting subgroup analyses. It revealed that 17-year-old boys continued to lag behind girls in reading (although girls have closed the gap in math). The 12-point spread back in 1971 had narrowed to eight points in 1988, but by 2004 had rebounded to a 14-point differential. Black and Hispanic students continue to perform below the national average, but have closed the gap considerably since 1971. Seventeen-year-old Hispanic students closed the gap by 11 points, and African American students closed it by 24 points (p. 15).
An additional subgroup category change worth noting is students with scale scores falling at or below the 59th percentile between 1992 and 2004. Average scale scores at the 50th percentile fell from 293 to 287. Average scale scores at the 25th percentile fell from 263 to 258. The fall at the 10th percentile began several years earlier. Then, average scale scores fell from 241 in 1988 to 227 in 2004. Each of those drops reached the level of statistical significance. Simultaneously, the data showed no expansion of students in the top performance category. Only 5-7% of adolescents performed at the top level.

The recently released 2009 reading results were similar. In reading assessments of 11 states (Arkansas, Massachusetts, Connecticut, New Hampshire, Florida, New Jersey, Idaho, South Dakota, Illinois, West Virginia, and Iowa), 52,000 students in grade 12 participated. The average score was 2 points higher in 2009 than in 2005, but 4 points lower than the score for the first reading assessment in 1992. The Nation’s Report Card for grade 12 reported that, “In comparison to 2005, scores in 2009 were higher for students at the 10th and 50th percentiles, and not significantly different from the scores for students at the 25th, 75th, and 90th percentiles (Figure 2).” In comparison to 1992, scores were lower in 2009 for students at the 10th, 25th and 50th percentiles, and not significantly different at the 75th and 90th percentiles (p.9). Although boys closed the gap with girls by 3 points, this was not significantly different from either 1992 or 2005 (p.12). NAEP began permitting accommodations in 1998. Figure 1 below shows the trend lines over two decades, differentiating between before accommodations and after accommodations.
The Program for International Student Assessment (PISA) provides an international comparison of adolescent reading achievement through a system of international testing that takes place every three years. Administered by the Organization for Economic Cooperation and Development (OECD), it was first given in 2000. The PISA reading assessment combines functional literacy with reading achievement. Fifteen-year-old students are assessed by assigning them critical reading in a series of real-world texts drawn from government forms, brochures, newspaper
articles, instruction manuals, etc. Answers are both multiple choice and constructed response. A supplement to the 2009 Condition of Education reported that PISA reading literacy data is available on the United States for 2000 and 2003 but not for 2006, due to a printing error (p. 13).

The same report illustrated a trend of static scores on the early-21st Century PISA, that were similar to those shown through the NAEP testing. The PISA data showed no statistically significant change for the reading scores of 17-year-olds in the US between 2000, when they were at the OECD average, and 2003. However, other countries are moving their adolescents’ reading achievement scores up. In 2000, six countries outperformed the United States in reading; in 2003 nine countries outperformed the US. Also, in 2003, the top 10% of US students had an average score of 600, below the OECD top 10% average of 617—a statistically significant difference. The report does provide a footnote stating that the static scores between 2000 and 2003, and being at the OECD average, may both be due, in part, to the large standard error assigned to the US data.

The US, largely driven by the school reform movement, has placed a lot of emphasis on test preparation. This investment of time, which occurs both in and out of school, has caused speculation by Bauerlein (2009) and others that improved test-taking skills is actually causing a bump in scores. Hence, scores that appear static may actually represent a decline in reading skill. Bauerlein posited that performance on content knowledge tests, including the NAEP, PISA and SAT II subject exams may be better arbiters of the reading achievement of adolescents. Bauerlein pointed to the percentile data as evidence of poor content literacy. On the 2006 NAEP subject test in history, for
example, 53% of 12th grade students tested scored below basic, and only 1% scored advanced. On the 2005 NAEP subject test in science, 46% did not reach basic, and only 2% reached advanced. Scale scores, however, present a more positive picture. In the 2006 NAEP subject test in history, 12th grade scale scores showed a significant improvement over both the 1996 and 2001 scores (pp. 9-10). The 2006 mean of 290 was significantly improved over the 1996 mean score of 286 and the 2001 mean score of 287. It should be noted, however, that accommodations were not permitted between 1996 and 2001, but were permitted between 2001 and 2006.

**Digital Literacy**

**Comparability Testing**

Gallagher, Bridgeman and Calahan (2002) described achievement testing as being driven to the digital environment by the pragmatism of cost and schedule, along with the benefits of more rapid turnaround time on results. State Departments of Education (DOEs) are taking the lead in administering large-scale trials of digital reading achievement testing. Any state moving some, or all, of its high-stakes testing to the digital environment is under a federal mandate to demonstrate the equity of the mean scores of students taking the online version with the mean scores of students taking the paper/pencil version. This is demonstrated by testing known as *comparability testing*.

In a review of comparability testing, Peak (2005) described inconsistencies in the results garnered from testing prior to 1993. Confounding factors that were identified included test takers’ unfamiliarity with computers, their inability to annotate, and mode effects such as scrolling and poor screen resolution. Peak acknowledged that many of the confounding factors have been mitigated in the intervening years. Increased use of
computers in classrooms (with a concomitant increase in student comfort and confidence with technology) and a reduction in access disparities have made it increasingly possible to use computers for assessment, including high-stakes assessment.

It has continued to be difficult, however, to produce consistency between students’ reading achievement in the text mode and their reading achievement in the digital mode. Choi and Tinkler (2002) performed a feasibility study for Oregon, assessing the viability of the widespread use of online testing. The goal of their research was to evaluate the score comparability of items administered to third- and tenth-grade students in mathematics and reading, across modes, and then to discuss methods to get the scores in each mode on a single scale. They included 14 high schools and approximately 800 students in the reading assessment. They took the 2001 spring test, which was already in two equal parts and converted each part to a Computer Based Test (CBT). They used a common person model and item level analysis of mode effect.

The correlation coefficient for 10th grade reading was 0.932, with a mean item difficulty differential of 0.10460. The correlation coefficient for 10th grade math was 0.931, but a mean item difficulty differential of only 0.01767 (p. 9). They broke the items out in terms of what level of specificity the question related to—word, phrase, sentence, or discourse. Tenth-grade students demonstrated the highest degree of mode interference with questions at the phrase level. Students were surveyed for preference and computer familiarity. A majority of students indicated a preference for the computer. Students who indicated a low degree of familiarity with computers had greater mode effect, meaning that they scored lower on the computer-based test.
In the spring of 2005, Way, Davis and Fitzpatrick (2006) administered comparability testing for Texas, in grades 8 and 11. They called their methodology "matched samples comparability analyses (MSCA)" (p.4). The population was students who either needed to retest, to pass their test for the first time (new to the state), or rising juniors. All students who took the test online were volunteers. Texas used an equating computation to correlate their paper/pencil results to the CBT results. The 11th grade reading test was technically an English Language Arts (ELA) test, and hence not strictly a reading achievement test. Nevertheless, students, on average, demonstrated more difficulty with the CBT than they did with the paper pencil test. As expected, students whose scores were being compared to their grade 10 scores did better. Students who were retaking an exit exam performed at approximately the same level. Overall, however, students taking the ELA demonstrated a mode effect that favored pencil and paper. The following excerpt from the Way, et al.'s Summary of Comparability points out the challenges of both administering the testing and of achieving comparability,

"In grade 8 reading, the mode differences were quite pronounced and warranted the use of the alternate score conversion table for reporting online results. In grade 11 mathematics and ELA, the differences were less pronounced and the ELA results were also complicated by the contributions of constructed response and extended essay items to the total scores. Nevertheless, the alternate score conversions were used for reporting scores with these tests, in part because of the magnitudes of raw score differences but also because of the high stakes associated with these tests" (p.13).
Pommerich performed a review of the comparability literature in 2004, which included her own primary research in comparability from 1998 and 2000. Her findings, like those of Choi and Tinkler, showed increased mode effect when all of the content could not be presented on a single screen and either scrolling or paging was required. Pommerich's population was grades 11 and 12, with an n of approximately 1800 students in the first study and approximately 3200 in the second. The tests were identical for content. Half of the students took the test on paper and half took it on the computer. In both studies, the reading portion consisted of four passages and was assessed by 10 associated multiple-choice questions. In both studies, the scores were significantly lower when the test was taken on the computer. The differential, however, was lower in 2000, possibly because of adaptations made to the computer test.

None of the aforementioned studies broke out their results by race or gender, although they represent potentially significant subcategories for differential impact. The 2002 work by Gallagher, et al., however, reviewed the results of several national tests, looking at how mode effect was influenced by race or gender. The four tests included were: SAT I, GRE, GMAT, and Praxis Exams for beginning teachers. Only the SAT I results are reported here, as that is the only test designed for, and administered to, the adolescent population.

Gallagher et al. used data derived from a 1996 study in which students took a paper/pencil version of the SATs, and then were offered the opportunity to take it on the computer. Students were incentivized through the option of adding their CBT scores to their score report. The 1,401 students who reported making a "good" or "strong" effort on the computer-based version were the students whose scores were included in the
study. Analysis was done on mean difference, which they referred to as "the impact". They set up the analysis to correct for differences in sample sizes, and then examined mean differences between the two versions, by race and by gender.

Gallagher et al. noted that there were small, but statistically significant, differences within the racial and gender groups. African Americans, and to a somewhat lesser degree, Hispanics, demonstrated reduced impact from the computer version of the assessment over the White reference group. Females demonstrated increased impact from the CBT, with White females demonstrating the greatest impact. It must be restated that although the sample size was impressive (1,732 students), the fact that the data was generated six years prior to the 2002 publication date of this study affects the relevance of the findings. Those years produced significant changes in both technology and access to technology, rendering the results less representative.

While the larger comparability studies have focused on producing equivalent instruments, there have been some smaller scale studies which have focused on other factors, in line with a recommendation made by Pommerich who drew this conclusion from her work: “Because computer technology is continually changing, testing programs should conduct their own comparability studies using their own tests and technology, as comparability results might not generalize beyond a given test and computer interface. Likewise, it is important for testing programs to conduct their own comparability studies, as results do not always turn out as might be expected” (p. 4).

In 2008, Latham summarized his research out of the University of Florida, in which he attempted to understand how a student’s high school preparation and sense of self-efficacy related to their performance on an assessment of digital literacy skills.
Latham conducted three studies over the course of three years, each of which pointed to variability in high school preparation of students in the area of digital literacy and to inconsistent adolescent mastery of information literacy skills (of which digital reading is an essential underpinning). The second study, for instance, was comprised of 52 first-semester college freshmen. Forty-five percent of these students scored below proficiency on an assessment of information literacy. Additionally, their pre- and post-assessment of their performance significantly overrated their own abilities in the area of information literacy. Finally, on a pretest questionnaire regarding where they had learned their skills, 74% reported being self-taught and 41% reported learning from peers (students were allowed to select multiple sources). In contrast, only 45% reported having received formal instruction in a media center, and only 26% said they had received instruction in a non-library class. Since Latham’s sample size was small and his students were a part of a program whose competitive admission policy kept them from being representative, these results are not inconsistent with other small-study findings.

Clariana and Wallace (2002) explored the role of four factors in students’ performance on a CBT versus their performance on a PPT. The study was a posttest-only assessment of 105 college freshmen who had just completed a Computer Fundamentals course. The course had four sections of students. Two sections were randomly selected and assigned to take the computer-based assessment. The other two sections took the identical test in a paper-based version. The computer-based groups significantly outscored the paper-based group. “Gender, competitiveness, and computer familiarity were NOT related to this performance difference, though content familiarity
was. Higher-attaining students benefited most from computer-based assessment relative to higher-attaining students under paper-based testing” (p.593).

Macedo-Rouet (2009) used data from 122 first-year college students enrolled in a math course in 2003. The data compared student success in using notes and taking a quiz on paper versus online. These students demonstrated greater challenge in utilizing the web-based notes and quiz. Seventy-four percent stated that they would rather use paper than online resources for the course (381). As a result of Macedo-Rouet’s study, the following observation was made: “The use of online course work in higher education is on the increase. But we have little empirical evidence of how our students perform in online reading and response, nor what their affinities are regarding digital versus text reading and response” (p. ). Limitations for this study were both size and the fact that the content being tested was math. The subject of math may present a greater cognitive load for some students, confounding the mode effect.

**Adolescent Functional Literacy**

Evidence of the changed functional literacies of adolescents in the millennium comes from the National Endowment for the Arts. Their surveys on the state of reading (2007, 2004), found that high school students are doing less traditional reading. From 1984 to 2004, the number of 17-year-olds who said they almost never read rose 10 percentage points. Interface with the computer and TV rose. In a 2007 NJDOE survey on student health, the average high school student was found to be spending about three hours a day in front of a screen. Multitasking is rampant. The 2007 report on literacy by the National Endowment of the Arts reported that 53% of the 13-17 year olds who do read asserted that they usually engaged with some other media while reading.
The correlation between the functional literacies of adolescents and their reading achievement is not clear, and there is heated debate in the literature about how schools should be treating these functional literacies. Marc Prensky (2001), who coined the phrase "digital natives", typifies one position. He asserted that students are fundamentally changed by their interface with technology, and that their academic problems stem largely from the outdated methods and content of millennial schools. He has been a vocal advocate for schools' changing both instructional methodologies and assessments to better address the literacy lives of our adolescent learners.

In his 2001 articles, *Digital Natives* and *Digital Immigrants I & II*, Prensky made his arguments in catchy language, speaking of legacy content, cognitive difference and neuroplasticity. The research supporting the descriptions and recommendations, however, made a less compelling case. Prensky did not engage in original research. Part I does not contain a bibliography or employ citations. Part II does not have a bibliography, but does have a notes section. Three sources—Lightspan Partnership, Click Health, and Scientific Learning—are educational software companies providing evidence of the efficacy of the products they provide. Additionally, Prensky provided data in quotes such as the following:

The numbers are overwhelming: over 10,000 hours playing videogames, over 200,000 emails and instant messages sent and received; over 10,000 hours talking on digital cell phones; over 20,000 hours watching TV (a high percentage fast speed MTV), over 500,000 commercials seen—all before the kids leave college. And, maybe, at the very most, 5,000 hours of book reading. These are today's "Digital Native" students (p.1).
The note for this data clarifies: "These numbers are intended purely as "order of magnitude...and obviously vary widely for individuals" (p. 8)..

Mark Bauerlein (2009) made the other side of the case in his book, *The Dumbest Generation*. He described adolescents as being academically disadvantaged by the time they are spending engaged with communication technologies. He argued that, because the goal of adolescents' functional literacies is social, their investment of time not only does not produce knowledge; it produces a lack of respect for knowledge. Bauerlein used content achievement tests, including the NAEP history exam, to support his premise that adolescent literacies are causing students to lose ground academically. He reported that, in 2001, 57% of high school seniors fell below basic on the NAEP history exam and only 1% achieved advanced proficiency. These percentages were identical to those obtained on the 1994 NAEP history exam. Although the 2006 history exam, administered to 29,000 seniors, saw those below proficient drop 4 percentage points, those at advanced proficiency remained at only 1%.

Bauerlein also pointed to a 2006 study by Educational Testing Service (ETS). The study assessed the digital research skills of 6300 students. It tested 15 tasks related to locating, assessing, evaluating, and creating information in the digital world. The results found that less than 50% of the students were proficient on almost all tasks, and that only 12% of students were able to construct a persuasive side and stick to the argument. He pointed to these as tasks that should be allowing digital natives to shine. The executive summary provided the following pieces of qualitative data. Students ranked convenience, connectedness and course management ahead of learning as benefits of technology. Additionally, the most technologically sophisticated students
design allowed this study to serve as action research in that direction, helping to identify challenges such testing presents. Additionally, the size of the sample could not be anticipated. Even with 114 participants, splitting them into two groups created the risk that some factorial analysis would be eliminated due to an inadequate sample size. Hence, the risks posed by the common person model were deemed to be the least detrimental to the goals of the study.

Finally, an analysis of the PSAT scores of these students was done using the same contextual factors. The PSAT is a norm-referenced test of critical reading administered to all students at the Career Academies in grades 10 and 11. The PSAT was taken within one month of the administration of the standards-based tests. The analysis of PSAT results for these same students was intended to explore their text reading in the context of norm-referenced (rather than standards-based) testing, which is the basis of most research findings.

Permission to conduct this research within the Monmouth County Vocational School District was provided by Mr. Timothy McCorkell, Superintendent of Schools, and by the Board of Education. The Principal of each of the four schools also gave written permission to conduct research at their buildings and with their students. Institutional Review Board (IRB) approval was required, as human subjects were being involved.

Building principals solicited lead teachers according to best fit for time and place of testing. It was made clear that teachers had the right to refuse to participate, and any teacher who agreed to participate signed an Informed Consent Form. Consenting teachers solicited the participation of 10th grade students at their schools. Because they
are minors, recruitment of students was a two-step process. Parents gave permission to invite student participation via an Informed Consent Form. An Informed Assent Form, signed by both the student and a parent, confirmed willing participation. Only students that returned both forms signed were participants.

A trial of the study was conducted in the spring of 2010, to identify unanticipated complications. During the spring trial, 58 of the 74 second-semester 10th grade students participated. These students came from the high school where I am an administrator, which likely accounted for the high return rate. The trial resulted in some modifications to methodology that will be described where relevant. The data from the trial will not be presented for analysis.

The study was conducted in the fall of 2010 with a comparable population. The study, however, did not include students for whom I had direct supervisory responsibility. For both the trial and the study, each student took the two tests in one sitting. The four schools engaged in the study administered the tests within three weeks of each other. The students all took the PSAT on October 13, 2010.

Population

The population for this study was 10th grade students from four of the five Career Academies administrated by the Monmouth County Vocational School District (MCVSD). The participating schools were: Academy of Allied Health (AAHS), Communications High School (CHS), High Technology High School (HTHS), and Marine Academy of Science and Technology (MAST). The fifth academy, Biotechnology High School (BTHS) was excluded to avoid the appearance of coercion and to avoid having results that were impacted by my supervisory relationship. The
MCVSD Career Academies are college preparatory programs with a career theme. The Academies share common curriculum in core courses, including English and Computer Applications, equalizing students’ reading preparation and formal high school exposure to technology.

Gender distribution across the academies is 54% female and 46% male. Ethnic distribution is 2% African American, 3% Hispanic, 23% Asian, and 72% White. The sample was disaggregated by the subcategory of gender. Ethnicity and English as a Second Language (ESL) were not included as subcategories for this study, as the numbers were deemed too small for statistical analysis. The exception to this would be the ethnic subcategory Asian. However, mean critical reading scores for Asian students have not traditionally differed significantly from the mean for White students, and hence were not disaggregated.

All student participants live in Monmouth County, located in the Central Eastern section of New Jersey and comprised of 52 different municipalities. Although Monmouth County boasts one of the highest per capita income levels in the country, the county is also home to areas of extreme poverty. There are five towns in the county that were formerly designated as Abbott Districts, including Asbury Park, one of the state’s most challenged school districts. Career Academy students are drawn from these extremes.

A student applies for a spot in a Career Academy in his or her 8th grade year. The highest scoring, eligible applicant from each sending district is guaranteed an offer of admission. Career Academy students were selected for this study because they are a homogenous population with above-average achievement. The use of this population
builds on the 2002 findings of Clariana and Wallace. Their posttest-only study randomized 150 college freshmen to take a content-based test on either the computer or on paper. The tests were identical and all items were multiple-choice. Clariana and Wallace found that the performance of high-achieving students benefited more from computer assessment than the performance of low-achieving students. The reading scores of the top performers in the US have dropped over the last 10 years. This study focused exclusively on academically talented students, in order to further explore the functional literacies of this population and the potential benefits that digital reading and assessment may have for them.

**Instruments**

The PPT and the CBT were adapted from the publicly available sample versions of the Washington State Assessment of Student Learning from 2007 and 2008. Permission to use these tests for educational purposes was granted, since no monetary benefit was being derived from their use. The PPT was adapted from the 2008, Grade 10, Sample Assessment in Reading (see Appendix A). The CBT was adapted from the 2007, Grade 10, Sample Assessment in Reading (see Appendix B). Instrument validity was tested using single factor, multifactor and comparative analysis. The tests were determined to be valid, with the validity being described as lying primarily in the content tested. The report states that reliability was estimated based on internal consistency measures. Full analysis of the validity and reliability of these instruments was reported in the associated technical reports published by Washington State (associated statistics can be found in Appendix 1).
A student who is a proficient programmer constructed the CBT by transferring the text to the computer. Precedent for converting a PPT into a CBT was demonstrated by Choi and Tinkler (2002). The CBT was guided by conclusions garnered from comparability studies regarding the mitigation of mode effects (Peak, 2005). The need to scroll, poor screen legibility, and loss of the ability to annotate, have all been previously identified as factors that exacerbate mode effect (making digital reading more difficult) for students. While scrolling was required of students in this study, students were working on equipment that was familiar to them and the sections were numbered to allow for visual landmarking, a feature that can be lost when scrolling is required. Screen legibility has been addressed through the progress of technology, and does not pose an issue at any of the sites due to the frequency of equipment upgrades. Students were not able to annotate on the screen, but this was not deemed to be a significant problem as the passages were only between five and eight paragraphs long, and students could scroll back at any time.

The number, type and scoring of items on the combined, adapted test attempted to mirror the level and cognitive load of the original test. The original test used four reading passages and presented readings in two categories: literary and informational. Three types of questions were used to assess students' understanding of their reading: multiple choice, short answer (SA), and extended answer (EA). The four readings were assessed through a total of 23 items on the sample test. The item analysis showed three categories: comprehension, analysis, and thinking critically. Multiple-choice questions were worth 1 point, SA worth 2, and EA worth 4.
Each of the adapted tests had two reading passages that were informational in nature and linked in content. Each test assessed the students' reading through 10 multiple-choice questions, one SA and one EA question. The value of each answer was multiplied by 1.5 from the original values, in order to preserve the ratio of weights while acknowledging the reduced number of items from which a score was being derived. So, the 10 multiple-choice questions per test were each worth 1.5 points, the one SA question was worth up to 3 points, and the one EA was worth up to 6 points. Each participant could earn a maximum composite score of 19 points.

On both versions, students could answer questions in any order, and change answers until they submitted their test. Multiple-choice questions were either right or wrong. The EE, which asked students to summarize a reading with three supporting details, was scored on a rubric that awarded up to 6 points. The SAE asked the students to support a statement with a detail from each reading. This answer was scored on a rubric that awarded up to 3 points. The rubrics for the essays were derived from the teacher guides provided by Washington State. They were modified slightly to accommodate the expanded point values.

The other instrument that was used was the 2010 PSAT Critical Reading. Critical Reading is one of three components on the PSAT. The other two are Writing and Math. The Critical Reading component is made up of two 25-minute sections containing a total of 48 questions. There are two types of questions: sentence completion and passage-based reading questions. Sentence completion questions test students' knowledge of the meanings of words and whether students understand how parts of a sentence should fit together logically. Passage-based reading questions test
whether students understand what they have read, whether they can extrapolate the meaning of a word from its context, and higher-level thinking skills such as analysis and synthesis.

Data Collection

Student participants were each assigned a unique, random student ID number (SID). This SID number was assigned by the programmer and provided to the student at the start of testing through the label affixed to the Scantron form (for use with the PPT). The student then used the same number as the login for the CBT.

The tests were given at one sitting. Although teachers received verbal and written directions for administering the tests that included directions on the order of administration, there was some deviation from the directions. The directions asked for each school to split students into two groups, with half of the students at each school taking the PPT first and half taking the CBT first. For whatever reason (confusion, access to technology, ease of supervision), this instruction was inconsistently implemented, resulting in an uneven split in the order of testing (a variable in the analysis).

Directions were read at the start of each test, and students were encouraged to read along (the same directions were available on the student test). Students were given 35 minutes per test, and teachers were instructed to post the start time and finish time on the board. If a student finished early he or she was encouraged to look over the test but could not go on to the next test.

All PPT answers were recorded on the Scantron sheet. The Scantron and test booklet were collected at the end of testing and returned to me. All CBT answers were
entered digitally. CBT answers were submitted by retyping a statement (a measure instituted to prevent unintended submission), and then hitting submit. The additional contextual information for each student was self-reported as a coversheet to the CBT. Students were asked to identify gender, and to select one of two levels for the contextual factors of preferred mode of reading, amount of time spent leisure reading, amount of leisure time spent on communication technologies, order of testing, and prior instruction and assessment in digital reading.

The scoring procedure for the short-answer essay and the extended essay on each test were derived from the Washington State teacher materials associated with the tests. Two English teachers from a nonparticipating school and I served as graders. The three examiners selected anchor answers for each of the four essays (answers that exemplified each of the rubric options). Then, each examiner was assigned two-thirds of the students from each school, so every student received two scores for each essay. The final score for each student essay was the mean of the two examiner scores.

Research Questions

The research questions, first presented in Chapter I, are detailed here for the reader to view in light of their associated hypotheses.

1. Are reading scores different, on average, on a standards-based assessment of text literacy than on a standards-based assessment of digital literacy, for high-achieving 10th grade students?

2. Are reading scores different, on average, on a standards-based text literacy than on a standards-based assessment of digital literacy, for high-achieving 10th grade students, when controlling for those contextual factors that the literature suggests
may mediate the effect of mode, including: a) school attended, b) gender, c) reading preference (text or digital), d) amount of time spent leisure reading, e) amount of time spent on communication technologies, f) order of testing, and g) prior instruction and assessment in digital reading?

3. Within one literacy mode (text or digital), are reading scores significantly different, on average, for high-achieving, 10th grade students of different gender and reading habits, when controlling for the selected contextual factors detailed above?

4. Within the arena of text literacy, are mean reading scores on a norm-referenced test (Preliminary Scholastic Achievement Test [PSAT]) related to mean reading scores on a standards-referenced test?

5. Within the arena of text literacy, do the set of selected contextual factors (v. supra) influence reading scores on a norm-referenced test (PSAT) in the same way as scores on a standards-referenced test, for high-achieving 10th grade students?

**Individual Hypotheses**

The null hypotheses for the research questions presented above are as follows:

1. High-achieving 10th grade students will perform the same, on average, on two standards-based assessments of critical reading, independent of the mode (text or digital).

2. Reading scores will be the same, on average, on a standards-based assessment of text literacy as on a standards-based assessment of digital literacy, for high-achieving 10th grade students, when disaggregated by
the following contextual factors: a) school attended, b) gender, c) reading preference (text or digital), d) amount of time spent leisure reading, e) amount of time spent on communication technologies, f) order of testing, and g) prior instruction and assessment in digital reading?

3. Within one literacy mode (text or digital), reading scores are the same, on average, for high-achieving, 10th grade students, when disaggregated by the selected contextual factors detailed above.

4. Within the arena of text literacy, reading scores on a norm-referenced test (the PSAT) are, on average, not related to reading scores on a standards-referenced test, for high-achieving 10th grade students.

5. Within the arena of text literacy, a set of selected contextual factors (detailed above) influence reading scores on a norm-referenced test (PSAT) in the same way that they influence scores on a standards-referenced test, for high-achieving 10th grade students.

Data Analysis

The two primary statistical treatments used in this study are the t test and the one-way ANOVA. These tests are designed to compare means, and both fall into the category of statistical treatments known as *General Linear Models*. As such, they share three critical assumptions about the data. For the results of these tests to be valid, the data must comply with these assumptions.

First is an assumption of normalcy. Nonconforming data must be transformed using the appropriate formula for skewed variables, and transformed data must be used in all subsequent analysis. For a distribution that is negatively skewed, as this one was,
values must first be reversed, or reflected, and the distribution becomes positively skewed. The transformations are then computed on the values in the positively skewed distribution. “Reflection is computed by subtracting all of the values for a variable from one plus the absolute value of maximum value for the variable. This results in a positively skewed distribution with all values larger than zero. When an analysis uses a transformation involving reflection, we must remember that this will reverse the direction of all of the relationships in which the variable is involved. Our interpretation of relationships must be adjusted accordingly”


The second assumption is of equality of variance. The Levene test is used to compare the squared standard deviations of the dependent variables. The assumption is met when the p values are not significant, indicating no significant difference between the standard deviations. If the data violated the equality of variance assumptions, the alternative Brown-Forsythe statistic (which does not presume on an equality of variance) was provided.

Finally, there is an underlying assumption of power. Statistical power is derived from three factors: sample size, level of significance and effect size. Power that is too low increases the risk of Type 1 and Type 2 errors (wrongly rejecting or accepting the null hypothesis). The sample size of 114 participants lends strong statistical power to the data, as does the use of the standard alpha of .05. Finally, when significance is demonstrated by a p value of .05 or less, effect size will be calculated. The d family of effect will be used, as this focuses on the magnitude of difference. When sample sizes are the same, the pooled standard deviation is used as the divisor. The formula for
calculating the d family of effect is derived from Leech, Barrett, and Morgan (2005).
Effect size will be reported according to Cohen’s four levels: small (.2), medium (.5),
large (.8), or very large (1 or more).

The two-tailed t test of paired samples demonstrates significant difference
through t values that have a p of .05 or less. One-way ANOVA, also tests mean, but
allows for disaggregating the data by factor, and was selected for its contribution to
computational ease in the Student Version of SPSS being used. A significant
relationship in an ANOVA is indicated by an f value with a p of .05 or less. The Levene
statistic was included for each ANOVA. If the Levene statistic does not meet the
assumption for equality of variance, the alternative Brown-Forsythe F-statistic and the
Welch F-statistic will be provided. Results will be displayed in table form. For the
factor, School Attended, which has levels, the post-hoc table will be provided when the
ANOVA reports significant difference.

**Hypothesis 1**

Hypothesis 1 tests whether high-achieving 10th grade students perform the same
on two standards-based assessments of critical reading, on average, independent of the
mode of delivery (one is paper/pencil, one is on the computer). The instruments are
described as the paper and pencil test (PPT) and the computer-based test (CBT). These
tests were administered as part of the study, and all participants were volunteers. Each
participant took both tests. The data was treated as needed to meet the assumption of
normalcy.

The independent variable for this hypothesis was mode, and the dependent
variable was reading comprehension. A two-tailed t test for paired samples with a level
of significance of .05 was used to compare students’ critical reading on the PPT, on
average, to their critical reading on the CBT. Requisite descriptives and statistics were
analyzed and presented in table form. If significance was demonstrated by a t value
with an associated p of .05 or less, effect size was reported.

A crosstab analysis was used to further explore this research question. The
goal of the analysis was to differently explore the relationship between student
performance in the context of text literacy and the context of digital literacy. Crosstab
analysis allowed a view of students’ movement through quartiles, in order to see if
students in any particular PPT quartile appeared to have added positive or negative
mode effect when testing in the digital context.

In setting up the crosstab, the PPT served as the independent variable, and the
CBT served as the dependent variable. Quartile frequencies were run on both the PPT
scores and on the CBT scores. A range of scores for each quartile was established for
each test. Two new variables were computed, and student scores were assigned to a
quartile for each test. In each case, the levels for the new variables were coded 1 for the
lowest quartile and 4 for the highest.

**Hypothesis 2**

Hypothesis 2 tests whether the reading scores of high-achieving 10th grade
students are the same, on average, independent of mode, when disaggregated by the
following contextual factors: a) school attended, b) gender, c) reading preference (text
or digital), d) amount of time spent leisure reading, e) amount of time spent on
communication technologies, f) order of testing, and g) prior instruction and assessment
in digital reading. Mode served again as the independent variable, and reading
achievement as the dependent variables (PPT and CBT scores). Factor information was self-reported by each participant as a part of the CBT. Each factor was coded as a dichotomous variable, with the exception of school attended, which had four levels. In the instance that any factor had a level with less than 15 participants, the analysis was automatically eliminated, due to the negative effect of a small sample size on statistical power. Post hoc analysis was run with the variable of "school attended", to demonstrate the source of the between subject significance. Tukey was used unless the assumption of equality of variance was violated, in which case the Games Howell was used.

In order to establish whether any difference in students' performance on text literacy compared to digital literacy was related to any of the factors above, a variable called mean difference was created. This variable was then disaggregated by each of the above factors. The descriptives and statistics associated with the one-way ANOVA were provided for each analysis and included homogeneity of variance. If the assumption of equal variance was not met, the alternative F-statistics would be provided. A significant relationship was demonstrated by an F-statistic, with an associated p value of .05 or less.

**Hypothesis 3**

Hypothesis 3 tests whether, within one literacy mode (text or digital), reading scores are the same, on average, for high-achieving, 10th grade students when disaggregated by the selected contextual factors detailed above. The contextual factors served as the independent variables, and reading achievement within mode served as the dependent variables. One-way ANOVA was the statistical treatment. The descriptives associated with the one-way ANOVA were provided for each analysis and included
homogeneity of variance. If the assumption of equal variance was not met by the Levene statistic, the alternative F-statistics were provided. In any case in which the F-statistic for the between subjects was significant (p value of .05 or less), the associated statistics were presented in table form and effect size was calculated.

**Hypothesis 4**

Hypothesis 4 tests whether, within the arena of text literacy, reading scores on a norm-referenced test (PSAT) are the same, on average, as on a standards-referenced test, for high-achieving 10th grade students. PSAT testing is a part of the routine of district-sponsored testing that takes place across the Career Academies. These 10th grade students took the PSAT on October 13, 2010, within three weeks of when they took the standards-based test generated by this study. The score from the critical reading section is used in this analysis. It should be noted, however, that students took the math and writing sections at the same sitting. The data was made available by the principal of each participating school. The scale score for the critical reading section of the PSAT is 20-80. For Hypothesis 4, test construction serves as the independent variable and reading achievement as the dependent variable.

**Hypothesis 5**

Hypothesis 5 tests if, within the arena of text literacy, a set of selected contextual factors (detailed above) influence reading scores on a norm-referenced test (PSAT) in the same way that they influence scores on a standards-referenced test, for high-achieving 10th grade students. For a description of the PSAT, see above. The selected contextual factors are those described in Hypothesis 2, with the exception of order of testing, as it is not applicable. Factor served as the main effect, while text
reading was the dependent variable with two levels: norm-referenced (PSAT) and standards-based. One-way ANOVA, with its associated descriptives, was the statistical treatment. All statistics are provided and when an F-value had an associated p of .05 or less, effect size is reported.

**Limitations**

Reading achievement research has faced significant methodological challenges, making it difficult to draw conclusions that can be generalized, or to draw valid conclusions about changes in reading achievement over time. This study sought to extend the research on text literacy compared to digital literacy through a demonstration of action research at the district level. Nevertheless, limitations arose that must be acknowledged.

1. This study used a high-achieving population, homogenous for prior demonstration of above-average mastery of math and language arts. This group, however, was suburban and without significant SES or ethnic subgroup variability. This sets a limitation on the ability to generalize the findings to all high achieving populations.

2. Participation in this study was voluntary, and the PPT and the CBT did not result in a grade for students. Given what we know about high-achieving students, it is likely that both the self-selection and the low-stakes outcome influenced student motivation and effort, and hence, outcomes.

3. This study acknowledges that the personal levels of digital competency and affinity possessed by the individuals in this study create a
limitation on the validity of the results. While there has been much leveling of the digital playing field in general and locally via a 9th grade Computer Applications course, it is inevitable that students have varying levels of competency in, and affinity for, using the computer. This may have influenced outcomes in ways that are difficult to quantify. The decision not to include a variable for digital competency was supported by Latham’s (2008) finding that students who were less competent in information literacy tended to overrate their competency.

4. This study acknowledges that, while every effort was made to create a CBT and PPT that were parallel in construct and equivalent in cognitive load, students may not have experienced these tests as equivalent. An affinity for the topic on one test versus the topic on the other test may have influenced the scores.

5. Finally, this study acknowledges that testing the digital reading skills of this high-achieving population for the first time posed certain challenges. While it was the informed professional opinion of three English teachers familiar with this population that these tests would pose a moderate challenge to their reading skills, it is possible that it posed less of a challenge than anticipated. According to the literature, this would have mitigated mode effect.

**Delimitations**

This study sets several delimitations.
1. First, while the terms digital literacy and text literacy were used in association with this study, they were applied in the study only as measures of students' ability to read and respond to content in the digital environment and the text environment. The term digital literacy is acknowledged to include the ability to access, retrieve and evaluate information in an online environment—skills that were not assessed in this study.

2. The second delimiter refers to the discussion of the technology associated with reading. Some of the research that has contributed to our understanding of adolescent reading in the digital environment has come from work with ESL students and struggling readers. This is important work, but was included only from the perspective that it has informed us about digital reading in general.

3. Finally, this study sets a delimitation on the purpose associated with gathering information on the communication practices of the participants. The information that the students provided was used solely to differently parse the data. It was used to advise whether a relationship exists between time spent on communication technology and reading success across modes, for the purpose of better addressing the reading needs, abilities, and preferences of adolescents. It did not seek to address whether students should engage in these practices, or whether parents should limit students' time on technology.
CHAPTER IV
ANALYSIS OF THE DATA

Introduction

This chapter reports the results of the analysis of the data obtained from a study of high-achieving 10th grade students. The study analyzed the students' standards-based text reading, relative to their standards-based digital reading. The analysis included disaggregating across and within mode by a selected set of contextual factors. It further examined their text reading through an analysis of their performance on a norm-referenced assessment of critical reading, compared to their performance on a standards-based assessment.

This study was conducted at four of the five Career Academies of the Monmouth County Vocational School District in central New Jersey. The data used was generated by this study, with the exception of the PSAT data, which was collected by district procedure. While student names were initially associated with the ID number for ease of distribution of text materials and associated login, the data was scrubbed of any names prior to analysis. The study design, the population, and the instruments used, were as described in the previous chapter. This chapter provides a detailed description of the sample, followed by a restatement of each hypothesis, and a report of the analysis of the results organized around each of the five hypotheses. The chapter will close with a summary of the findings.
Description of the Sample

The sample was solicited from the population per IRB protocol. Letters of Solicitation were given to all grade 10 students at each of the four schools. All students who participated signaled parental approval and personal willingness through the return of a signed Informed Assent Form. When the tests results were compiled, five students had failed to submit tests that were complete in both modes. In all cases, students had failed to complete either the PPT short answer essay or the PPT extended answer essay. These students were eliminated from the sample. Table 1 below shows final participation by school.

Table 1

<table>
<thead>
<tr>
<th>School</th>
<th>Total Student Body</th>
<th>Total 10th Graders</th>
<th>Participated</th>
<th>Submitted Complete data</th>
<th>% of 10th Graders Participating</th>
<th>School as % of Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>284</td>
<td>73</td>
<td>19</td>
<td>19</td>
<td>26.0</td>
<td>17.0</td>
</tr>
<tr>
<td>1</td>
<td>292</td>
<td>73</td>
<td>21</td>
<td>17</td>
<td>29.0</td>
<td>15.0</td>
</tr>
<tr>
<td>2</td>
<td>268</td>
<td>69</td>
<td>31</td>
<td>31</td>
<td>50.0</td>
<td>27.0</td>
</tr>
<tr>
<td>3</td>
<td>291</td>
<td>71</td>
<td>48</td>
<td>47</td>
<td>68.0</td>
<td>41.0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1135</td>
<td>286</td>
<td>119</td>
<td>114</td>
<td>42.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Gender was the only subcategory used to disaggregate data in this study, due to insufficient n’s in other categories. Since the literature reports a persistent gender gap in
reading achievement (with girls continuing to score significantly higher than boys),
gender data from each of the four schools is explored in some depth and reported in
Table 2 below.

Table 2

<table>
<thead>
<tr>
<th>School</th>
<th>Females as % of Total Student</th>
<th>Females as % of 10th Graders</th>
<th>Number of Participants</th>
<th>Number of Female Participants</th>
<th>Females as Percent of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>66.0</td>
<td>58.0</td>
<td>19</td>
<td>12</td>
<td>65.0</td>
</tr>
<tr>
<td>1</td>
<td>62.0</td>
<td>69.0</td>
<td>17</td>
<td>12</td>
<td>71.0</td>
</tr>
<tr>
<td>2</td>
<td>38.0</td>
<td>44.0</td>
<td>31</td>
<td>15</td>
<td>48.0</td>
</tr>
<tr>
<td>3</td>
<td>51.0</td>
<td>47.0</td>
<td>47</td>
<td>19</td>
<td>41.0</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>58</td>
<td></td>
<td>58</td>
<td>51.0</td>
</tr>
</tbody>
</table>

The gender split by school appears to conform to stereotypes associated with the related
career themes. The pre-engineering program is heavily male, while the
communications- and medical arts-themed schools have more females. The marine
science-themed school is the most gender balanced. The demographics by school did
not predict the gender balance of the participants. Nevertheless, the final gender split
was very even, at 49% male and 51% female.
Basic Descriptive Statistics

One hundred nineteen students were tested. Valid results were obtained for 114 participants, and all analysis was based on that number. Student data were entered into SPS Statistics Student Version 17.0. Students were identified by their randomly assigned SID number as well as by school. Frequency tables of final participation by school and by gender were exhibited above.

All variables were dichotomous, with the exception of school attended, which had four levels. Each variable was numerically coded. Table 3 below summarizes the variables, their value, and percentages. They are included before a restatement of the hypotheses to assist the reader in further conceptualizing the study.
Table 3

Distribution of Sample by Independent Variables, N=114 (in Percent)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>17.0</td>
</tr>
<tr>
<td>School Attended</td>
<td>2</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>27.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>41.0</td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>51.0</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>49.0</td>
</tr>
<tr>
<td>Reading Preference</td>
<td>Computer</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td>Text</td>
<td>75.0</td>
</tr>
<tr>
<td>Amt. Leisure Reading</td>
<td>2 hr/wk or less</td>
<td>46.5</td>
</tr>
<tr>
<td></td>
<td>More than 2 hr/wk</td>
<td>53.5</td>
</tr>
<tr>
<td>Time Spent on Communication Technologies</td>
<td>3 hr/day or less</td>
<td>32.0</td>
</tr>
<tr>
<td></td>
<td>More than 3 hr/day</td>
<td>68.0</td>
</tr>
<tr>
<td>Order of Testing</td>
<td>Computer First</td>
<td>39.5</td>
</tr>
<tr>
<td></td>
<td>Paper/Pencil First</td>
<td>60.5</td>
</tr>
<tr>
<td>Prior Instruction/Assessment in Digital Reading</td>
<td>Yes</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>98.0</td>
</tr>
</tbody>
</table>

Table 3 shows that there was adequate student representation for statistical analysis of all factors, with the exception of prior instruction/assessment in digital reading. Because only two participants claimed to have had prior
instruction/assessment, this factor was eliminated from analysis. As noted earlier, overall gender was a remarkably even split, despite the fact that individual schools showed gender imbalance and representation by school was not symmetrical. Interestingly, although students were twice as likely to describe themselves as spending more time with technology, they were three times more likely to prefer text reading over digital. Students were almost evenly split on whether they described themselves as doing more or less leisure reading. Almost twice as many students took the paper-based test first.

The distributions of the standards-based text literacy (PPT) scores and the standards-based digital literacy scores were tested for normalcy. The data for both the PPT and CBT showed negative skewness: -.531 and -.942 respectively (Table 4). The parametric statistics used in this study, two-tailed t test and ANOVA, are quite robust and may show little effect, even from skews that are +/- 1 (Leach, Barrett, and Morgan). However, the study pursued transformation to protect the integrity of the findings. Transformation with reflection was implemented, using the transformation formula for negatively skewed variables within SPSS. Table 4 below shows the descriptives before transformation, and Table 5 shows them after transformation. All subsequent data analysis used the transformed data.
Table 4

The Distribution of Text Literacy and Digital Literacy Scores before Transformation

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Skew Statistic</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT</td>
<td>114</td>
<td>11.0</td>
<td>18.75</td>
<td>15.46</td>
<td>1.75</td>
<td>-.53</td>
<td>.22</td>
</tr>
<tr>
<td>CBT</td>
<td>114</td>
<td>9.75</td>
<td>18.55</td>
<td>15.78</td>
<td>1.45</td>
<td>-.94</td>
<td>.22</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5

The Distribution of Text Literacy and Digital Literacy Scores After Transformation

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Skew Statistic</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/P</td>
<td>114</td>
<td>1.0</td>
<td>2.96</td>
<td>1.91</td>
<td>.35</td>
<td>-.04</td>
<td>.23</td>
</tr>
<tr>
<td>CBT</td>
<td>114</td>
<td>1.0</td>
<td>3.13</td>
<td>2.24</td>
<td>.37</td>
<td>-.29</td>
<td>.23</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Both the PPT and the CBT showed a negative skew, with CBT having a more pronounced skew—close to one. The transformed data shown above meets the assumption of normalcy, with only residual negative skew remaining.

This study does not presume on a correlation of student performance across modes, and indeed presumes that the different skills and affordances associated with each mode will disrupt correlation. If the intended use for the scores had required high positive correlation, the scores for these students would have needed to be equated in the subsequent analysis. This would be the case, for instance, when offering a
standardized test assumed to measure the same underlying construct in both text and digital modes. It must be noted, however, that the lack of correlation raises the possibility that there were unintended inequalities in the two instruments. For instance, students may have found the content in the reading passages more interesting in the CBT than in the PPT passages, making it appear easier and contributing to what appears to be mode effect.

**Findings for Hypothesis 1**

$H_0$: Hypothesis 1 states that high-achieving 10$^{th}$ grade students will perform the same, on average, on two tests of critical reading, independent of the mode.

**Preliminary Findings**

Table 6 below reports the basic distribution of student scores on the two tests. A glance at the table suggests that CBT scores were higher than PPT scores, although the CBT distribution was more skewed. Table 7 reports the results of a comparison of means between the PPT and the CBT using a two-tailed t test for paired samples.

Table 6

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Skew Statistic</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT</td>
<td>114</td>
<td>1.0</td>
<td>2.96</td>
<td>1.92</td>
<td>.35</td>
<td>-.04</td>
<td>.23</td>
</tr>
<tr>
<td>CBT</td>
<td>114</td>
<td>1.0</td>
<td>3.13</td>
<td>2.22</td>
<td>.37</td>
<td>-.29</td>
<td>.23</td>
</tr>
</tbody>
</table>
Students had a mean score of 1.91 on the PPT and a mean score of 2.22 on the CBT. The mean difference was -.31, with a standard deviation of .42. The mean score difference was significant at the .000 level with a t value of -7.77 and df (113), suggesting that the null hypothesis should be rejected; i.e., that scores on the two modes were significantly different. The effect size was small, with d equal to .21. To amplify these results, crosstabs were used to compare how respondents’ quartile position on the distribution on one instrument differed from their quartile position on the distribution of the other. Table 8 below reports the associated distribution.
Table 8

_A Comparison of Respondents Quartile Positions on PPT and CBT Distributions_,

_N=114_

<table>
<thead>
<tr>
<th>PPT Quartiles</th>
<th>CBT Quartiles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Count</td>
<td></td>
<td>12</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>% PPT quartile</td>
<td></td>
<td>37.5</td>
<td>37.5</td>
<td>18.8</td>
<td>6.3</td>
<td>100.0</td>
</tr>
<tr>
<td>2 Count</td>
<td></td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>% PPT quartile</td>
<td></td>
<td>20.0</td>
<td>32.0</td>
<td>28.0</td>
<td>20.0</td>
<td>100.0</td>
</tr>
<tr>
<td>3 Count</td>
<td></td>
<td>11</td>
<td>6</td>
<td>9</td>
<td>12</td>
<td>38</td>
</tr>
<tr>
<td>% PPT quartile</td>
<td></td>
<td>28.9</td>
<td>15.8</td>
<td>23.7</td>
<td>31.6</td>
<td>100.0</td>
</tr>
<tr>
<td>4 Count</td>
<td></td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>% PPT quartile</td>
<td></td>
<td>10.5</td>
<td>15.8</td>
<td>31.6</td>
<td>42.1</td>
<td>19.0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>30</td>
<td>29</td>
<td>28</td>
<td>27</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26.3</td>
<td>25.4</td>
<td>24.6</td>
<td>23.7</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Chi square testing found the quartile differences to be significant at the .048 level, with a value of 17.05 df (9). Of the students who participated, 10.5% of them scored in the lowest quartile on the PPT, and also scored in the lowest quartile on the CBT. Another 10.5% of students scored in the lowest quartile on the PPT, and also scored in the second lowest quartile on the CBT, indicating a pattern of poor performance that was independent of mode. Beyond those students, however, there appeared to be much less consistency of association. Ten percent of the students who scored in the 3rd quartile of the PPT scored in the first quartile of the CBT; another 10% scored in the same 3rd quartile of the PPT, but scored in the 4th quartile of the CBT.
Summary Findings

The null hypothesis is rejected for Hypothesis 1. Students did not perform the same, on average, on the two reading assessments. Students performed significantly better on a standards-based assessment of digital literacy than on a standards-based assessment of text literacy. The crosstab analysis showed significant differences between students' quartile scores on text literacy and their quartile score on digital literacy. Twenty-one percent of students scored in the bottom quartile on the PPT, and also in the bottom half of the CBT, indicating a weakness in reading that manifested in both contexts. Beyond that, performance in the text context appeared less well associated with performance in the digital context, lending credibility to viewing text and digital literacy as separate constructs.

Findings for Hypothesis 2

H_{02}: Hypothesis 2 proposes that performance will be the same, on average, on a standards-based assessment of text literacy and a standards-based assessment of digital literacy, for high-achieving 10^{th} grade students, when disaggregated by the following contextual factors: a) school attended, b) gender, c) reading preference (text or digital), d) amount of time spent leisure reading, e) amount of time spent on communication technologies, and f) order of testing (prior instruction and assessment in digital reading was eliminated due to an inadequate sample size).

Preliminary Findings

Table 9 below reports the mean difference in PPT-CBT scores by school attended. School attended met the equality of variance assumption at the .121 level.
Table 9

<table>
<thead>
<tr>
<th>School code</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19</td>
<td>-.38</td>
<td>.47</td>
<td>-.60</td>
<td>-.15</td>
<td>-.34</td>
<td>.57</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>-.24</td>
<td>.36</td>
<td>-.42</td>
<td>-.06</td>
<td>-.96</td>
<td>.38</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>-.33</td>
<td>.33</td>
<td>-.45</td>
<td>-.21</td>
<td>-1.10</td>
<td>.49</td>
</tr>
<tr>
<td>3</td>
<td>47</td>
<td>-.29</td>
<td>.48</td>
<td>-.42</td>
<td>-.15</td>
<td>-1.24</td>
<td>.78</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>-.31</td>
<td>.42</td>
<td>-.38</td>
<td>-.23</td>
<td>-1.34</td>
<td>.78</td>
</tr>
</tbody>
</table>

The mean mode difference ranged from -.24 at School 1, to -.38 at School 0, with an average difference of -.31. Students at each of the four schools performed better on the CBT than on the PPT. Table 10 below reports the associated statistics.

Table 10

<table>
<thead>
<tr>
<th>Mean PPT-CBT Difference by School Attended, ANOVA, N=114</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
</tr>
<tr>
<td>Mean difference Between Groups</td>
</tr>
<tr>
<td>Within Groups</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Mean difference by School Attended was not significant at the .74 level.
Gender. Gender was split 58 females (52%) to 56 males (43%). Gender met the equality of variance assumption at the .829 level. Table 11 reports mean difference in PPT-CBT scores by gender.

Table 11

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>58</td>
<td>-.35</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-1.34</td>
</tr>
<tr>
<td>Male</td>
<td>56</td>
<td>-.25</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-1.24</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>-.31</td>
<td>.42</td>
</tr>
</tbody>
</table>

Females had a greater mean difference between their PPT scores and their CBT scores than males had, perhaps indicating greater mode effect for female students. Table 12 below reports the associated statistics.

Table 12

<table>
<thead>
<tr>
<th>Mean PPT-CBT Score Differences by Gender, ANOVA, N=114</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
</tr>
<tr>
<td>Mean difference Between Groups</td>
</tr>
<tr>
<td>Within Groups</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>
Mean score difference did not attain statistical significance when disaggregated by gender.

**Reading preference.** Eighty-six students (75%) claimed to prefer to read text, while the remaining 28 (25%) claimed to prefer to read digitally. Reading preference met the equality of variance assumption at the .670 level. Table 13 below reports the mean differences in PPT-CBT scores by reading preference.

Table 13

*Mean PPT-CBT Difference Scores by Preferred Mode for Reading, N=114*

<table>
<thead>
<tr>
<th>Test Factor</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer</td>
<td>28</td>
<td>-.26</td>
<td>.44</td>
<td>-1.24</td>
<td>.65</td>
</tr>
<tr>
<td>Print</td>
<td>86</td>
<td>-.32</td>
<td>.41</td>
<td>-1.34</td>
<td>.78</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>-.30</td>
<td>.42</td>
<td>-1.34</td>
<td>.78</td>
</tr>
</tbody>
</table>

Students described themselves as preferring text reading to digital reading at a 3:1 ratio, although students, on average, scored higher on the CBT than on the PPT. The associated statistics are reported in Table 14 below.
Table 14

*Mean PPT-CBT Difference Scores by Preferred Mode for Reading, ANOVA, N=114*

<table>
<thead>
<tr>
<th>Mean difference</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>.072</td>
<td>1</td>
<td>.07</td>
<td>.41</td>
<td>.52</td>
</tr>
<tr>
<td>Within Groups</td>
<td>19.51</td>
<td>112</td>
<td>.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19.58</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean score difference disaggregated by preferred reading mode was not significant.

**Time spent leisure reading.** Amount of time spent leisure reading was self-reported by students at one of two levels. Students either described themselves as spending two hours or less per week leisure reading or more than two hours per week leisure reading. Fifty-three students (46.5%) claimed to spend two hours or less per week leisure reading. Sixty-one students (53.5%) claimed to do more than two hours of leisure reading per week. The amount of time spent leisure reading met the equality of variance assumption at the .765 level. Table 15 below reports mean difference in PPT-CBT scores by time spent leisure reading.
Table 15

*Mean PPT-CBT Difference Scores by Preferred Time Spent Leisure Reading, N=114*

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤2 hr/wk</td>
<td>53</td>
<td>-.30</td>
<td>.41</td>
<td>-1.24</td>
<td>.78</td>
</tr>
<tr>
<td>&gt;2 hr/wk</td>
<td>61</td>
<td>-.30</td>
<td>.42</td>
<td>-1.34</td>
<td>.65</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>-.30</td>
<td>.42</td>
<td>-1.34</td>
<td>.78</td>
</tr>
</tbody>
</table>

Students' had virtually no mean score difference when controlling for the amount of time they spend leisure reading. The mean score difference was .30, whether they read more than two hours/week or less than two hours/week. Table 16, below, reports the ANOVA for the mean score difference when controlling for time spent leisure reading.

Table 16

*Mean PPT-CBT Score Difference by Amount of Time Spent Leisure Reading, ANOVA, N=114*

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean difference</td>
<td>.00</td>
<td>1</td>
<td>.00</td>
<td>.00</td>
<td>.959</td>
</tr>
<tr>
<td>Between Groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Groups</td>
<td>19.58</td>
<td>112</td>
<td>.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19.58</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Students' mean reading score difference was not significant when disaggregated by the amount of time a student spent leisure reading.

**Time spent on communication technology (CT).** Students self-reported the amount of time per day they spent on communication technologies (CT). CT was described as blogging, surfing the Internet, texting, phoning, visiting Facebook, etc. Students were encouraged to count time in which they were multitasking with CT as time on CT. Students had to describe their CT usage as three hours or less per day, or more than three hours per day. Thirty-six students (32%) described themselves as spending three hours a day or less on CT. Seventy-eight students (82%) reported spending more than three hours per day on CT. CT met the assumption of equality of variance at the .232 level. Table 17, below, reports mean difference in PPT-CBT scores by time spent on communication technology.

Table 17

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤3 hr/day</td>
<td>36</td>
<td>-.22</td>
<td>.36</td>
<td>-.92</td>
<td>.63</td>
</tr>
<tr>
<td>&gt;3 hr/day</td>
<td>78</td>
<td>-.34</td>
<td>.44</td>
<td>-1.34</td>
<td>.78</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>-.31</td>
<td>.42</td>
<td>-1.34</td>
<td>.78</td>
</tr>
</tbody>
</table>
Students selected the higher time investment as a descriptor at a 2:1 ratio. Table 18, below, reports the ANOVA for the mean difference between reading scores on the PPT and the CBT, controlling for time spent on communication technologies.

Table 18

<table>
<thead>
<tr>
<th>Mean PPT-CBT Score Difference by Amount of Time Spent on Communication Technologies, ANOVA, N=114</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean difference</td>
</tr>
<tr>
<td>Between Groups</td>
</tr>
<tr>
<td>Within Groups</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Mean score difference disaggregated by the amount of time a student spent on communication technologies was not significant.

**Order of testing.** The students reported which test they took first. Forty-five students (40%) reported taking the CBT first. The remaining 69 students (55%) reported taking the PPT first. As noted in Chapter III, inconsistent adherence to the directions related to the order of testing, resulted in an uneven split of participants for the variable of order of testing. The irregularity associated with this variable will not be addressed again. Order of testing met the assumption of equality of variance at the .137 level. Table 19, below, reports mean difference in PPT-CBT scores by order of testing.
Table 19

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Minimum</td>
</tr>
<tr>
<td>CBT</td>
<td>45</td>
<td>-.37</td>
<td>.38</td>
</tr>
<tr>
<td>PPT</td>
<td>69</td>
<td>-.26</td>
<td>.44</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>-.31</td>
<td>.42</td>
</tr>
</tbody>
</table>

Regardless of which test they took first, the mean difference for both groups indicated that students performed better on the digital test. Students who took the CBT scored .37 higher on the CBT. Students who took the PPT first scored .26 higher on the CBT. Table 20, below, reports the ANOVA associated with the mean difference in scores when controlling for order of testing.

Table 20

<table>
<thead>
<tr>
<th>Mean PPT-CBT Score Difference by Order of Testing, ANOVA, N=114</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Mean difference</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The mean difference in reading scores between the PPT and CBT, disaggregated by the order of testing, was not significant.
Summary Findings

Performance across mode showed no significant difference when controlling for the selected contextual factors. None of the factors that were controlled for were associated with a significant mean difference in reading scores between text reading and digital reading. Hence, the null hypothesis for H₀₂ is accepted, as student scores across mode were, on average, the same when controlling for each of six selected contextual factors. The seventh factor, prior exposure to digital reading and assessment, was eliminated due to inadequate n.

Findings for Hypothesis III

H₀₃: Within each literacy mode (text or digital), the critical reading for high-achieving 10th grade students is the same, on average, when controlling for the selected contextual factors detailed above.

Preliminary Findings

For Hypothesis 3, the data was examined within mode. One-way ANOVA was used to parse scores within each type of literacy by the same contextual factors described above. The descriptives and associated statistics are presented for each contextual factor.

School attended. School attended met the assumption for the Homogeneity of Variance for both the PPT and the CBT at .199 and .105, respectively. Table 21, below, reports critical reading within mode, controlling for school attended.
Table 21

*The Effect of School Attended on PPT and CBT Scores, N=114*

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Minimum</td>
</tr>
<tr>
<td><strong>PPT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>1.90</td>
<td>.30</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>1.99</td>
<td>.30</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>2.05</td>
<td>.30</td>
</tr>
<tr>
<td>3</td>
<td>47</td>
<td>1.82</td>
<td>.40</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>1.91</td>
<td>.35</td>
</tr>
<tr>
<td><strong>CBT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>2.28</td>
<td>.48</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>2.23</td>
<td>.27</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>2.37</td>
<td>.31</td>
</tr>
<tr>
<td>3</td>
<td>47</td>
<td>2.10</td>
<td>.36</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>2.22</td>
<td>.37</td>
</tr>
</tbody>
</table>

Students from each of the four schools performed better, on average, on the CBT. Rank order on the two tests varied slightly. On the PPT, the order of performance by school from lowest to highest was 3, 0, 1, 2. On the CBT, the order of performance was 3, 1, 0, 2. In both cases, School 2 was the highest performing school, on average, and School 3 was the lowest performing. Table 22 reports the associated post hoc analysis.
Table 22

The Effect of School Attended on PPT and CBT Scores, Tukey HSD, N=114

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>School Code</th>
<th>School Code</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval Lower Bound</th>
<th>95% Confidence Interval Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT</td>
<td>0</td>
<td>1</td>
<td>-0.09</td>
<td>0.12</td>
<td>0.85</td>
<td>-0.39</td>
<td>0.21</td>
</tr>
</tbody>
</table>

had mixed feelings about technology in their courses. In the face of conflicting evidence, it is left to individual schools and districts to determine the essential skills and knowledge their graduates will need and how best to deliver and assess them.
and was significant at the .03 level with df (3, 110). The between groups for the CBT had an F Statistic of 3.77, and was significant at the .01 level with df (3, 110). There was a medium effect size: d was equal to .64 for the PPT and equal to .79 for the CBT.

Historically, these four schools have shown differences in mean scores on standardized tests (SAT, PSAT, and High Speed Packet Access [HSPA]), despite a common process for competitive admission. This difference is likely due to middle school attended, level of parental education, and general academic drive associated with the student body.

**Gender.** Fifty-eight females (51%) and 56 males (49%) all took both tests. The assumption of homogeneity of variance was met by the PPT at the .162 level and by the CBT at the .077 level. Table 23, below, reports the mean differences by gender for PPT and CBT scores in turn. The findings show that males and females both had higher mean scores on the CBT. Females, however, had significantly higher mean scores in both modes than males did. Table 24, below, reports the results of a one-way ANOVA of the effect of gender on PPT and CBT scores.
Table 23

The Effect of Gender on PPT and CBT Scores, N=114

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>58</td>
<td>1.98</td>
<td>.317</td>
<td>1.32</td>
<td>2.96</td>
</tr>
<tr>
<td>Male</td>
<td>56</td>
<td>1.85</td>
<td>.377</td>
<td>1.00</td>
<td>2.74</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>1.92</td>
<td>.353</td>
<td>1.00</td>
<td>2.96</td>
</tr>
<tr>
<td>CBT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>58</td>
<td>2.33</td>
<td>.324</td>
<td>1.29</td>
<td>3.13</td>
</tr>
<tr>
<td>Male</td>
<td>56</td>
<td>2.11</td>
<td>.383</td>
<td>1.00</td>
<td>2.89</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>1.22</td>
<td>.371</td>
<td>1.00</td>
<td>3.13</td>
</tr>
</tbody>
</table>

Table 24

The Effect of Gender on PPT and CBT Scores, ANOVA, N=114

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT</td>
<td>.48</td>
<td>1</td>
<td>.48</td>
<td>4.00</td>
<td>.048</td>
</tr>
<tr>
<td>Between Groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Groups</td>
<td>13.56</td>
<td>112</td>
<td>.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14.05</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT</td>
<td>1.512</td>
<td>1</td>
<td>1.51</td>
<td>12.06</td>
<td>.00</td>
</tr>
<tr>
<td>Between Groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Groups</td>
<td>12.05</td>
<td>112</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15.56</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The main effect of gender was significant in both types of literacy. Girls significantly outperformed boys, on average, on both the PPT and the CBT. The between-groups variance on the PPT was statistically significant at the .048 level, with an F statistic of 4.00 and df (1,112). On the CBT, the between-groups variance was
significant at the .001 level, with an F statistic of 12.06 and df (1,112). Effect size was medium again, with \( d \) equal to .37 for the PPT and .61 for the CBT.

**Preferred mode of reading.** The survey question asked the students to identify whether they preferred to read on the computer or preferred to read text. Of the 114 students sampled, 28 (25%) claimed they preferred to read on the computer. The remaining 86 (75%) claimed they preferred to read text. The assumption of the equality of variance was violated by the PPT at the .01 level. Using the alternate Brown-Forsythe statistic, the CBT met the assumption at the .288 level and the PPT met the assumption at the .50 level. A one-way ANOVA tested the effect of preferred mode of reading on scores on each literacy test. Table 25, below, reports the mean differences by reading preference for PPT and CBT scores in turn. Table 26, below, reports the associated ANOVA of the PPT and CBT scores by in turn, by reading preference.
Table 25

_The Effect of Preferred Mode of Reading on PPT and CBT Scores, N=114_

<table>
<thead>
<tr>
<th>Test Factor</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/P Computer</td>
<td>28</td>
<td>1.84</td>
<td>.46</td>
<td>1.00</td>
<td>2.96</td>
</tr>
<tr>
<td>Print</td>
<td>86</td>
<td>1.94</td>
<td>.31</td>
<td>1.13</td>
<td>2.74</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>1.92</td>
<td>.35</td>
<td>1.0</td>
<td>2.96</td>
</tr>
<tr>
<td>CBT</td>
<td>28</td>
<td>2.11</td>
<td>.32</td>
<td>1.63</td>
<td>2.70</td>
</tr>
<tr>
<td>Print</td>
<td>86</td>
<td>2.26</td>
<td>.38</td>
<td>1.00</td>
<td>3.13</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>2.22</td>
<td>.37</td>
<td>1.00</td>
<td>3.13</td>
</tr>
</tbody>
</table>

Table 26

_The Effect of Preferred Mode of Reading on PPT and CBT Scores, ANOVA, N=114_

<table>
<thead>
<tr>
<th>Test</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT</td>
<td>Between Groups</td>
<td>.21</td>
<td>1</td>
<td>.21</td>
<td>1.70</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>13.84</td>
<td>112</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>14.05</td>
<td>113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT</td>
<td>Between Groups</td>
<td>.46</td>
<td>1</td>
<td>.46</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>15.10</td>
<td>112</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>15.56</td>
<td>113</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The students who claimed they preferred to read print earned higher scores, on average, in both modes. Students who claimed to prefer to read on the computer earned
the lowest average scores on the test with paper and pencil. The differences between
groups clustered by reading preference were not significant for either test.

**Time spent leisure reading.** The survey question asked the student to identify
how much time they spent leisure reading per week at one of two levels. Of the 114
students sampled, 53 (25%) claimed they read two hours/week or less. The remaining
61 (75%) claimed they read more than two hours/week. The assumption of equality of
variance was met by the PPT at the .951 level, and by the CBT at the .248 level. A one-
way ANOVA examined the mean difference in scores within each literacy, parsed by
amount of leisure reading. Table 27, below, reports the mean differences by amount of
time spent leisure reading for both the PPT scores and the CBT scores in turn.
Table 27

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 hr/wk</td>
<td>53</td>
<td>1.19</td>
<td>.34</td>
<td>1.00</td>
<td>2.46</td>
</tr>
<tr>
<td>&gt;2 hr/wk</td>
<td>61</td>
<td>2.01</td>
<td>.34</td>
<td>1.13</td>
<td>2.96</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>1.92</td>
<td>.353</td>
<td>1.00</td>
<td>2.96</td>
</tr>
<tr>
<td>CBT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 hr/wk</td>
<td>53</td>
<td>2.13</td>
<td>.33</td>
<td>1.00</td>
<td>2.89</td>
</tr>
<tr>
<td>&gt;2 hr/wk</td>
<td>61</td>
<td>2.31</td>
<td>.38</td>
<td>1.29</td>
<td>3.13</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>2.22</td>
<td>.37</td>
<td>1.00</td>
<td>3.13</td>
</tr>
</tbody>
</table>

Students who described themselves as doing more leisure reading had higher mean scores in both types of literacy. On the PPT, their mean was 2.01, compared to 1.19 for students who did less leisure reading. On the CBT, their mean score was 2.31 compared to 2.13 for the other group. Table 28, below, reports the ANOVA for the reading scores on PPT and CBT scores in turn, parsed by leisure reading.
Table 28

*The Effect of Amount of Time Spent Leisure Reading on PPT and CBT Scores, ANOVA, N=114*

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.98</td>
<td>1</td>
<td>.93</td>
<td>8.42</td>
<td>.004</td>
</tr>
<tr>
<td>Within Groups</td>
<td>13.07</td>
<td>112</td>
<td>.117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14.05</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>.97</td>
<td>1</td>
<td>.97</td>
<td>7.43</td>
<td>.007</td>
</tr>
<tr>
<td>Within Groups</td>
<td>14.59</td>
<td>112</td>
<td>.130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15.56</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Students who described themselves as engaging in more leisure reading performed significantly better within both types of literacy. The mean difference on the PPT was significant at the .004 level, with an F statistic of 8.42 and df (1,112). The mean difference on the CBT was significant at the .007 level, with an F statistic of 7.43 and df (1,112). There was, again, a medium effect size, with d equal to .57 on the PPT and .50 on CBT.

**Time spent on communication technologies.** The question asked students to quantify the amount of time they spend on communication technologies. Students either said they spent three hours per day or less, or they spent more than three hours per day on communication technologies. Of the 114 participants in this study, 36 (32%)
described themselves as spending three hours or less on communication technologies. The remaining 78 students (68%) claimed they spent more than three hours/day on communication technologies. One-way ANOVA was used to compare means within mode, parsed by the amount of time spent on communication technologies. The assumption of homogeneity of variance was met on the PPT at the .105 level, and on the CBT at the .532 level. Table 29, below, reports the mean differences by time spent on communication technology for the PPT scores and the CBT scores in turn. Table 30, below, reports the associated ANOVA for the PPT and the CBT, also in turn.

Table 29

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/P ≤3 hr/day</td>
<td>53</td>
<td>1.96</td>
<td>.29</td>
<td>1.41</td>
<td>2.48</td>
</tr>
<tr>
<td>&gt;3hr/day</td>
<td>61</td>
<td>1.90</td>
<td>.38</td>
<td>1.00</td>
<td>2.96</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>1.92</td>
<td>.35</td>
<td>1.00</td>
<td>2.96</td>
</tr>
<tr>
<td>CBT ≤3</td>
<td>53</td>
<td>2.18</td>
<td>.38</td>
<td>1.29</td>
<td>2.91</td>
</tr>
<tr>
<td>&gt;3hr/day</td>
<td>61</td>
<td>2.25</td>
<td>.37</td>
<td>1.00</td>
<td>3.13</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>2.22</td>
<td>.37</td>
<td>1.00</td>
<td>3.13</td>
</tr>
</tbody>
</table>
Table 30

*The Effect of Amount of Time Spent on Communication Technologies on PPT and CBT Scores, ANOVA, N=114*

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.08</td>
<td>1</td>
<td>.08</td>
<td>.64</td>
<td>.425</td>
</tr>
<tr>
<td>Within Groups</td>
<td>13.84</td>
<td>112</td>
<td>.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14.05</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.11</td>
<td>1</td>
<td>.11</td>
<td>.83</td>
<td>.365</td>
</tr>
<tr>
<td>Within Groups</td>
<td>15.44</td>
<td>112</td>
<td>.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15.56</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On the PPT, students who spent less time on communication technologies had a higher mean than students who spent more time: 1.96 compared to 1.90. On the CBT, the findings were reversed: students who spent less time on communication technologies had a lower mean than students who described themselves as spending more time: 2.18 compared to 2.25. The mean differences in scores, however, were not significant on either test.

**Order of testing.** The question asked the students which test they took first—the PPT or the CBT. Of the 114 students who took the test, 69 (61%) said that they took the PPT first, and the other 45 (39%) said that they took the CBT first. Mean score difference was then disaggregated by order of testing, for each type of literacy, using one-way ANOVA. The assumption of equality of variance was met on the PPT at the
.873 level, and on the CBT at the .810 level. Table 31, below, reports the mean differences by order of testing for both the PPT scores and the CBT scores in turn.

Table 31

<table>
<thead>
<tr>
<th>Test</th>
<th>Order</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT</td>
<td>CBT first</td>
<td>45</td>
<td>1.98</td>
<td>.34</td>
<td>1.22</td>
<td>2.74</td>
</tr>
<tr>
<td>PPT</td>
<td>first</td>
<td>69</td>
<td>1.88</td>
<td>.36</td>
<td>1.00</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>114</td>
<td>1.92</td>
<td>.35</td>
<td>1.00</td>
<td>2.96</td>
</tr>
<tr>
<td>CBT</td>
<td>CBT first</td>
<td>45</td>
<td>2.35</td>
<td>.37</td>
<td>1.29</td>
<td>2.13</td>
</tr>
<tr>
<td>CBT</td>
<td>first</td>
<td>69</td>
<td>2.15</td>
<td>.35</td>
<td>1.00</td>
<td>3.91</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>114</td>
<td>2.22</td>
<td>.37</td>
<td>1.00</td>
<td>3.13</td>
</tr>
</tbody>
</table>

The students who took the CBT first had higher means on both the PPT and the CBT. On the PPT, students who took the CBT first had a mean score of 1.98, compared to the 1.88 of students who took the PPT first. On the CBT, students who took the CBT first had a mean score of 2.35, compared to the 2.15 of the students who took the PPT first. Table 32, below, reports the results of the one-way ANOVA of the effect of order of testing on the PPT scores and on the CBT scores, in turn.
Table 32

The Effect of the Order of Testing on PPT and CBT Scores, ANOVA, N=114

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.27</td>
<td>1</td>
<td>.27</td>
<td>2.15</td>
<td>.145</td>
</tr>
<tr>
<td>Within Groups</td>
<td>13.78</td>
<td>112</td>
<td>.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14.05</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>1.22</td>
<td>1</td>
<td>1.22</td>
<td>9.52</td>
<td>.003</td>
</tr>
<tr>
<td>Within Groups</td>
<td>14.34</td>
<td>112</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15.56</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Students who took the CBT first had significantly higher scores, on average, on the CBT, than the students who took it as their second test. The between-groups on the CBT was significant at the .003 level, with an F statistic of 9.52 and df (1,112). The effect size for order of testing was small, with d=.29. Order of testing was not associated with a significant mean difference in the scores on the PPT.

Summary Findings

Certain selected contextual factors were associated with significant differences within one or both modes of literacy. Specifically, school attended, gender, amount of time spent leisure reading and order of testing. In summary:

- Schools 2 and 3 performed significantly differently from each other, on average, on both the PPT and the CBT.
- Girls performed significantly better than boys, on average, on both the PPT and the CBT.
- Students who claimed to do more leisure reading performed significantly better in both modes, on average, than students who described themselves as doing less leisure reading.
- Order of testing was significant within the digital mode. Students who took the CBT first performed significantly better on it than students who took it second.
- Reading preference and time on communication technology were not significant effects.

Hypothesis 3 supposed that scores would be the same, on average, within each mode, independent of the main effect. Therefore, the null hypothesis is rejected for Hypothesis 3.

**Findings for Hypothesis 4**

**H04:** Within the arena of text literacy, reading scores on a norm-referenced test (the PSAT) are, on average, not related to reading scores on a standards-referenced test, for high-achieving 10th grade students.

**Preliminary Findings**

The standards-based assessment of text literacy was the instrument previously described, created for this study and based on a practice assessment developed by Washington State. It was graded, as previously described, by a team assembled for the purpose of this study. The PSAT is designed and scored by College Boards. Grade 10
Career Academy students take it diagnostically and for practice for the SAT. Students took the tests within three weeks of each other in the fall of 2010. Both tests required the students to read text and answer with paper and pencil.

The tests were compared using the Spearman rho Correlation for Rank Order. This test is used frequently in test-retest situations, or in situations in which multiple forms of a test are used to ensure that the instruments are measuring the same set of skills and knowledge. The data meets the assumptions, in that it is ordinal and monotonic. The Spearman Correlation was used, as opposed to the Pearson, due to the residual skew in the data. A correlation coefficient is reported, and its relative strength described. The significance is reported out as a p value. Table 33, below, reports the mean reading scores on the PPT and PSAT in turn. The PPT before transformation was on a scale of 0-19; after transformation it was on a scale of 0-3.5. The PSAT is on a scale of 0-80. Table 34, below, reports the correlation between the reading scores on the PPT and reading scores on the PSAT.

Table 33

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT</td>
<td>1.92</td>
<td>.35</td>
<td>114</td>
</tr>
<tr>
<td>PSAT CR</td>
<td>60.91</td>
<td>7.73</td>
<td>114</td>
</tr>
</tbody>
</table>
Table 34

A Comparison of Text Literacy Constructs: Correlation, N=114

<table>
<thead>
<tr>
<th>Spearman's rho</th>
<th>PPT</th>
<th>PSAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT</td>
<td>Correlation Coefficient</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>114</td>
</tr>
<tr>
<td>PSAT</td>
<td>Correlation Coefficient</td>
<td>.411**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>114</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed)

With r .411, the correlation is significant at the .01 level, indicating a relationship between students' text reading, as measured by the standards-based test, and by the norm-referenced test. The .411 level indicates a moderately strong relationship.

Summary Findings

For Hypothesis 4, the null hypothesis is rejected. The scores did show a relationship between student performance on the norm-referenced assessment of critical reading and the standards-based assessment. The correlation between the scores on the two text assessments was higher than the correlation between the standards-based text scores and the standards-based digital scores. This may lend validity to the idea of viewing digital and text literacy as distinct constructs.

Hypothesis 5

H_{05}: Within the arena of text literacy, a set of selected contextual factors (detailed above, excluding order of testing) will influence reading scores on a norm-
reference test (PSAT) in the same way as they influence scores on a standards-referenced test (PPT), for high-achieving 10th grade students.

**Preliminary Findings**

All Grade 10 students at each of the four schools took the PSAT as part of the district regiment of testing. The assumption of normality was met with a skewness of -.010. The descriptives and statistics associated with the PPT were reported in the tables associated with Hypothesis 3, and will not be presented again. The associated findings, however, will be referred to in the analysis and the summary. In summary, for Hypothesis 3, one-way ANOVA found significant mean score differences on the PPT when disaggregating by the following factors: school attended, gender, and time spent leisure reading. The generalized descriptives for mean score on the PSAT can be found in Table 32, above.

**School attended.** The Levene Statistic was significant at the .046 level, which violated the assumption for the equality of variance. The alternative Brown-Forsythe was significant at the .000 level. Table 35, below, reports the mean score difference on the PSAT Critical Reading when controlling for school attended.
Table 35

The Effect of School Attended on PSAT Critical Reading Scores, \(N=114\)

<table>
<thead>
<tr>
<th>School code</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19</td>
<td>58.21</td>
<td>4.53</td>
<td>56.03</td>
<td>60.39</td>
<td>49.00</td>
<td>66.00</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>59.41</td>
<td>7.73</td>
<td>55.44</td>
<td>63.39</td>
<td>43.00</td>
<td>66.00</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>65.84</td>
<td>8.33</td>
<td>62.78</td>
<td>68.90</td>
<td>47.00</td>
<td>76.00</td>
</tr>
<tr>
<td>3</td>
<td>47</td>
<td>59.30</td>
<td>7.05</td>
<td>57.23</td>
<td>61.38</td>
<td>44.00</td>
<td>72.00</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>60.91</td>
<td>7.72</td>
<td>59.48</td>
<td>62.35</td>
<td>43.00</td>
<td>76.00</td>
</tr>
</tbody>
</table>

School 2 had the highest mean, at 65.84. School 3 had the lowest mean, at 59.30. This was consistent with the findings from the PPT in which School 2 had the highest mean and School 3 had the lowest mean. Table 36, below, reports the mean difference in PSAT scores from a One-Way ANOVA, when controlling for school attended.
Table 36

*The Effect of School Attended on PSAT Critical Reading Scores, Games Howell,*

$N=114$

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>School Code</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>PSAT</td>
<td>0</td>
<td>-1.20</td>
<td>2.40</td>
<td>.943</td>
<td>-7.09</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-7.63*</td>
<td>2.10</td>
<td>.001</td>
<td>-12.48</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-1.09</td>
<td>1.50</td>
<td>.879</td>
<td>-.4.97</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.20</td>
<td>2.40</td>
<td>.943</td>
<td>4.69</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-6.43</td>
<td>2.17</td>
<td>.052</td>
<td>-12.89</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.11</td>
<td>1.67</td>
<td>1.000</td>
<td>5.98</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>7.63*</td>
<td>2.10</td>
<td>.001</td>
<td>12.48</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6.43</td>
<td>2.17</td>
<td>.052</td>
<td>12.89</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6.54*</td>
<td>1.67</td>
<td>.004</td>
<td>11.35</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.09</td>
<td>1.96</td>
<td>.879</td>
<td>4.97</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-.11</td>
<td>2.04</td>
<td>1.000</td>
<td>5.75</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-6.54*</td>
<td>1.67</td>
<td>.004</td>
<td>-1.74</td>
</tr>
</tbody>
</table>

*The mean difference is significant at the 0.05 level.

Students from School 2 performed significantly better, on average, than the students from two of the other schools: School 0 and School 3. The difference between school 0 and School 2 was significant at the .001 level. The difference between School 2 and School 3 was significant at the .004 level. The effect size was medium, with $d$ equal to .69. This was consistent with the findings on the PPT, the standards-based test of text literacy, on which the students from School 2 performed significantly better, on average, than the students from School 3.
**Gender.** There were 58 females and 56 males. The assumption of equality of variance was met with a Levene statistic that was not significant at the .234 level. Table 37, below, reports the mean difference in PSAT Critical Reading (CR) scores. Table 38 reports the ANOVA for the mean difference in PSAT CR scores, when controlling for gender.

Table 37

*The Effect of Gender on PSAT Critical Reading Scores, N=114*

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/P Female</td>
<td>58</td>
<td>61.93</td>
<td>6.92</td>
<td>45.00</td>
<td>76.00</td>
</tr>
<tr>
<td>Male</td>
<td>56</td>
<td>59.86</td>
<td>8.42</td>
<td>43.00</td>
<td>76.00</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>60.91</td>
<td>7.73</td>
<td>43.00</td>
<td>76.00</td>
</tr>
</tbody>
</table>

Table 38

*The Effect of Gender on PSAT Critical Reading Scores, ANOVA, N=114*

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>278.67</td>
<td>1</td>
<td>278.67</td>
<td>3.113</td>
<td>.08</td>
</tr>
<tr>
<td>Within Groups</td>
<td>10024.56</td>
<td>112</td>
<td>89.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10303.24</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Girls had a higher mean score than boys on the PSAT, but the between-groups variance was not significant. The main effect of gender influenced student performance
on a norm-referenced assessment of text literacy differently than it influenced performance on a standards-based assessment. The effect of gender on standards-based reading scores was significant, with females outperforming males. The effect of gender on the norm-referenced reading scores was not significant. Males and females scored the same, on average, on a norm-referenced test of reading.

**Reading preference.** Twenty-eight students described themselves as preferring to read digitally; the other 86 described themselves as preferring to read text. The data met the assumption of equality of variance. The Levene statistic was not significant at the .223 level. Table 39, below, reports the mean difference in PSAT scores when controlling for a student's preferred mode of reading. Table 40, below, reports the ANOVA for mean score difference disaggregated by preferred reading mode.

<table>
<thead>
<tr>
<th>Preferred Mode</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSAT Digital</td>
<td>28</td>
<td>58.61</td>
<td>8.76</td>
</tr>
<tr>
<td>Print</td>
<td>86</td>
<td>61.66</td>
<td>7.26</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>60.91</td>
<td>7.73</td>
</tr>
</tbody>
</table>
Table 40

*The Effect of Preferred Mode for Reading on PSAT Critical Reading Scores, ANOVA, N=114*

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSAT Between Groups</td>
<td>197.22</td>
<td>1</td>
<td>197.22</td>
<td>3.37</td>
<td>.069</td>
</tr>
<tr>
<td>Within Groups</td>
<td>6547.90</td>
<td>112</td>
<td>58.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6745.12</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The mean score for students who preferred to read digitally was 58.6071. The mean score for students preferring to read text was 61.663. This mean difference was not significant at the .069 level. Reading preference was not a significant factor for mean difference in reading scores on the standards-based assessment of text reading, either.

**Amount of time spent leisure reading.** Fifty-three students described themselves as engaging in two hours or less per week of leisure reading. The other 61 said that they read more than two hours per week. The assumption of equality of variance was met by a Levene statistic that was not significant at the .824 level. Table 41, below, reports the mean difference on the PSAT, when controlling for amount of time students spent leisure reading. Table 42, below, reports the associated ANOVA.
Table 41

The Effect of Amount of Time Spent Leisure Reading on PSAT Critical Reading Scores,

N=114

<table>
<thead>
<tr>
<th>Test</th>
<th>Amt. LR</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSAT</td>
<td>≤2 hrs/wk</td>
<td>53</td>
<td>59.10</td>
<td>7.89</td>
<td>43.00</td>
<td>76.00</td>
</tr>
<tr>
<td></td>
<td>&gt;2 hrs/wk</td>
<td>61</td>
<td>62.49</td>
<td>7.28</td>
<td>47.00</td>
<td>76.00</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>114</td>
<td>60.91</td>
<td>7.73</td>
<td>43.00</td>
<td>76.00</td>
</tr>
</tbody>
</table>

Table 42

The Effect of Amount of Time Spent Leisure Reading on PSAT Critical Reading Scores,
ANOVA, N=114

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSAT</td>
<td>Between Groups</td>
<td>327.35</td>
<td>1</td>
<td>327.35</td>
<td>5.71</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>6417.77</td>
<td>112</td>
<td>57.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6745.12</td>
<td>113</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Students who read more in their leisure time had a higher mean score than students who read less: 62.49 compared to 59.10. Students who read more also had a higher minimum score than students who read less: 47 compared to 43. The between-group variance for this factor was significant, with an F statistic of 5.713 and df (1,112) at the .019 level. The effect size associated with the amount of time spent leisure reading was smaller than is typical, with d equal to .14. This was consistent with the
analysis of the standards-based text literacy in which students who read more also significantly outperformed students that read less.

**Time on communication technologies (CT).** Thirty-six students described themselves as spending three hours per day or less on CT, while the other 78 said that they spent more than three hours per day. The Levene statistic for this data was not significant at .803, meeting the assumption of equality of variance.

Table 43, below, reports the mean difference in PSAT reading scores, when controlling for time students spent on communication technologies. Table 44, below, reports the ANOVA for the mean difference in PSAT scores, when controlling for the amount of time students spent with communication technologies.

Table 43

*The Effect of Amount of Time Spent on Communication Technologies on PSAT Critical Reading Scores, N=114*

<table>
<thead>
<tr>
<th>Test</th>
<th>Amt. LR</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSAT</td>
<td>≤3 hrs/day</td>
<td>36</td>
<td>61.75</td>
<td>7.59</td>
<td>48.00</td>
<td>76.00</td>
</tr>
<tr>
<td></td>
<td>&gt; 3 hrs/day</td>
<td>78</td>
<td>60.53</td>
<td>7.81</td>
<td>43.00</td>
<td>76.00</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>114</td>
<td>60.91</td>
<td>7.73</td>
<td>43.00</td>
<td>76.00</td>
</tr>
</tbody>
</table>
Table 44

The Effect of Amount of Time Spent on Communication Technologies on PSAT Critical Reading Scores, ANOVA, N=114

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSAT Between Groups</td>
<td>36.92</td>
<td>1</td>
<td>36.92</td>
<td>.62</td>
<td>.43</td>
</tr>
<tr>
<td>Within Groups</td>
<td>6708.20</td>
<td>112</td>
<td>59.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6745.12</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Students who spent less time on communication technology had a higher mean score on the PSAT reading than students who spent more time: 61.75 compared to 60.53. Students who spent less time on communications technologies also had a higher minimum score than students who spent more time: 48 compared to 43. The between-groups variance was not significant at the .434 level. The amount of time spent on communications technologies was also not a significant factor in mean difference on standards-based reading scores.

Summary Findings

Hypothesis 5 assumed that a set of selected contextual factors would influence performance on a norm-referenced assessment of text literacy in the same way as they influenced a standards-based assessment of text literacy. The factors of Preferred Mode of Reading and Amount of Time Spent on CT were not significant for either of the assessments of text literacy (PPT or PSAT). The factors of School Attended and Time Spent Leisure Reading each showed a significant relationship to the both types of text
literacy. For norm-referenced text literacy, however, the effect size for School Attended was much larger than the effect size for Time Spent Leisure Reading.

School Attended showed that students from School 2 significantly outperformed students from at least one other school on each test. Time Spent Leisure Reading showed that students who did more leisure reading scored significantly better, on average, on both the standards-based and the norm-referenced assessments of text literacy. Surprisingly, gender was the factor that showed disparity. On the standards-based assessment, the girls significantly outperformed the boys (as they did on the assessment of digital literacy). However, on the norm-referenced assessment, there was no significant difference in mean score when controlling for gender. Hence, the null hypothesis was rejected for Hypothesis 5; the factors did not affect the two versions of text literacy equally.

**Summary of the Data Analysis**

Table 45, below, reports a summary of the effects of each contextual factor, by mode, for the convenience of the reader. Within the text mode, standards-based (SB) and norm-referenced (NR) testing are reported separately.
Table 45

**Summary: Contextual Factor Effects by Mode, N=114**

<table>
<thead>
<tr>
<th>Mode</th>
<th>School Attended</th>
<th>Gender</th>
<th>Preferred Mode of Reading</th>
<th>Time Spent On Leisure Reading</th>
<th>Time Spent On Communication Technologies</th>
<th>Order of Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB Text: PPT</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>NR Text: PSAT</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Digital: CBT</td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

*=Significance

The study examined students’ text literacy compared to their digital literacy.

The study examined data related to students’ performance on a standards-based assessment of text literacy and a similar assessment of their digital literacy. The study further disaggregated the data by a gender and a set of contextual factors. A two-tailed paired-samples t test and one-way ANOVA were the primary statistical tools. The design of this study did not presume on a rank order correlation and, indeed, presupposed that, based on the unique skills and affordances associated with each type of literacy, student scores would diverge when compared across mode. This data was derived from two parallel tests (one text, one digital) that were designed for this study. Text literacy was further explored using PSAT data on the same students who had been collected per district policy. The study tested the null hypothesis associated with five research questions related to student performance across and within the two modes of literacy.
The analysis revealed that this group of high-achieving 10th grade students performed significantly better on a digital reading assessment than they did on an assessment of text literacy, despite the fact that 75% of participants claimed to prefer to read in text. Gender and the selected contextual factors did not prove to have a significant influence on differences across modes, when analyzed through one-way ANOVA.

When literacy modes were examined separately, however, several factors revealed a significant influence. First, there was a significant difference in scores in each mode when controlling for the factor of school attended. A significant difference existed between School 2 and School 3, at a minimum, in both digital literacy and text literacy. Gender was also associated with a significant mean difference in each mode, with girls outperforming boys in both text literacy and digital literacy. Finally, the between-subjects difference associated with the amount of time a student spent leisure reading was significant within both modes. Students who did more leisure reading significantly outperformed students who did less leisure reading, in both text literacy and digital literacy. Order of testing was associated with a significant mean difference in the reading scores only within the digital mode of literacy. Students did better on the CBT, regardless of which test they took first. However, students who took the CBT first did significantly better on it than students who took it second.

The analysis of text literacy was then extended to a comparison of standards-based text literacy and norm-referenced text literacy. ANOVA testing of the contextual factors was extended to the PSAT as an example of norm-referenced text reading. The
associated null hypothesis presumed that the contextual factors would produce the same findings in both types of text literacy. Four of the five factors tested produced the same findings. Preferred Mode of Reading and Time on CT were not significant for either. School Attended and Amount of Time Spent Leisure Reading were significant for both. Gender was where the two versions of text literacy parted ways. Girls significantly outperformed the boys on the standards-based assessment (and on the digital assessment), but gender was not associated with a significant mean difference on the PSAT. Chapter V will discuss the implications of these findings for practice and policy, and suggest possible applications for further research.
CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Introduction

Since the 1983 report, A Nation at Risk, launched the school reform movement, the reading achievement of American adolescents has been tested extensively. Success on high-stakes tests has been the driving force shaping the curriculum and instructional strategies of literacy education in high schools across America during this era. But, despite dedicating three decades to the school reforms demanded by this report, the reading achievement scores of American adolescents are stagnant at best. In addition, technology has changed the literacy landscape and complicated its analysis.

Some of the ways that technology has changed literacy for adolescents are well understood in the field of education. For instance, it is understood that technology has raised the level of literacy demanded of a high school graduate. The globalization of the job market and the concurrent automation and/or outsourcing of many lower-skilled jobs have raised the level of literacy required for the jobs that remain. The educational system has subsequently been required to graduate students who can demonstrate higher levels of literacy than previous graduates.

It is also understood that technology has pushed more postsecondary and civic literacy into the digital world. Cost effectiveness and ease of access have pushed college classes, databases, manuals and registrations into the online environment. The absence of direct human assistance for these activities raises the level of literacy needed by all citizens. Finally, it is understood that technology has captured the attention of adolescents, leading them to spend significant amounts of time in the digital world. What
is not well understood is how proficient students are with academic tasks in the digital world, or what the relationship is between the time adolescents spend in the digital world and their academic success and motivation.

Digital reading, as a facet of 21st Century literacy, was the focus of this study. This study compared the constructs of text literacy and digital literacy for a group of high-achieving 10th grade students. Text literacy was assessed on a standards-based assessment of critical reading, as well as on a norm-referenced assessment (the PSAT). Digital literacy was assessed on a parallel standards-based assessment of critical reading. Chapter V presents an overview of the study, a summary and discussion of the principle findings, and recommendations for associated literacy practice and policy as well as recommendations for further research.

**Overview of the Study**

This study assumed that literacy consists of multiple constructs, consisting of related but distinct skills. It furthers assumes that success in each literacy is highly dependent on the context and the motivation of the individual adolescent. The study examined students' critical reading in two modes: text and digital. The study compared students' mean scores on a paper/pencil reading test to their mean reading scores on a computer-based reading test. The study then disaggregated scores within mode by the subcategory of gender and the contextual factors of school attended, reading preference (computer or text), time spent leisure reading, time spent on communication technologies, order of testing, and prior instruction and assessment in digital reading.

The goal of this study was to add to the field of research exploring digital literacy and factors that influence student success in that mode. In particular, this study extended
the research in a direction recommended by both Pommerich (2004) and Russell et al. (2003). Each suggested that the inconsistencies in the findings to date pointed to the need for districts to undertake testing of their own students, with an eye to finding the optimal mode for each student. Disaggregating the data by contextual factors continues the effort to unearth insights on the relationship between the current functional literacies of adolescents and their reading achievement. Recommendations were based on the data analysis, as well as on the procedural and process lessons learned.

**Research Design**

This study was a cross-sectional analysis of data collected, using a common person design. Valid data was collected on 114 10th grade students from four Vocational School District Career Academies in Monmouth County, New Jersey. Because admission to these academies is competitive, the sample population can be described as homogenous for high achievement. Each participant took both a standards-based assessment of text literacy (PPT) and an equivalent assessment of digital literacy (CBT). In addition, the district provided PSAT scores for these students.

Content for each of the standards-based tests was modified from a Washington (State) Assessment of Student Learning sample test (WASL) (2007, 2008), which were publicly available online from Washington State. Each test consisted of two linked reading passages that were informational in nature. Each test had 10 multiple-choice questions, one short-answer question and one extended-essay response. The combined number of items mimicked the cognitive load of a single sample test, and the procedure for grading the tests was adapted from the 2007 Washington State Technical Report.
A senior student on internship built the web-based version of the CBT and imputed the adapted 2008 content as a part of his capstone senior project. The CBT was designed to mitigate mode effect, according to lessons learned from prior research. Paragraphs were numbered, providing landmarks for retrieving information, a feature that is lost when scrolling is involved. The test also enabled students to go back and revisit prior questions, mimicking text literacy. And finally, the CBT was given on equipment that the students were familiar with and which had good screen resolution.

Participants, both students and teachers, were recruited using IRB-approved Letters of Solicitation, and Informed Consent and Informed Assent Forms. Each of the four schools administered the tests to students' sequentially, although schools varied in which test they administered first. Sixty-two percent of the students took the CBT first, while the other 38% took the PPT first. A unique, randomly generated ID number was assigned to each student for preservation of anonymity.

**Individual Hypotheses**

The following are the null hypotheses that were tested:

1. High-achieving 10th grade students will perform the same, on average, on two standards-based assessments of critical reading, independent of the mode (text or digital).

2. Reading scores will be the same, on average, on a standards-based assessment of text literacy as on a standards-based assessment of digital literacy, for high-achieving 10th grade students, when disaggregated by the following contextual factors: a) school attended, b) gender c) reading preference, (text or digital), d) amount of time spent leisure reading, e)
amount of time spent on communication technologies, f) order of testing, and g) prior instruction and assessment in digital reading.

3. Within one literacy mode (text or digital), reading scores are the same, on average, for high-achieving 10th grade students, when disaggregated by the selected contextual factors detailed above.

4. Within the arena of text literacy, reading scores on a norm-referenced test (the PSAT) are, on average, not related to reading scores on a standards-referenced test, for high-achieving 10th grade students.

5. Within the arena of text literacy, a set of selected contextual factors (detailed above) influence reading scores on a norm-referenced test (PSAT) in the same way that they influence scores on a standards-referenced test, for high-achieving 10th grade students.

Data Analysis Procedures

SPSS Student Version 17.0 was used in the analysis of the data collected for this study. Hypothesis 1 focused on the mean difference between student achievement on the PPT and student achievement on the CBT. Primary modes of analysis were a two-tailed t test of paired samples and one-way ANOVA. The assumption of normalcy was tested, and the data was transformed to correct for negative skewness. Prior to ANOVA testing, all data was subjected to the Levene test for equality of variance. If the assumption of equality was violated, the more robust Brown-Forsythe and Welch tests were performed. Further, if the assumption of equality was not met where post hoc analysis was needed, the F statistic provided by the Games-Howell was used in lieu of the Tukey.
Summary of the Findings

Hypothesis 1

Hypothesis 1 tested mode effect by comparing the students' reading achievement on the PPT (a standards-based assessment of text reading) to their reading achievement on the comparable CBT (a standards-based assessment of digital literacy). Mean difference between PPT and CBT scores was analyzed using a two-tailed t test of paired samples. The null hypothesis was rejected, as there was a significant difference in the mean scores.

Students performed significantly differently when reading in the digital mode than when they were reading in the text mode. This mean difference in scores lends support to the contention of this investigation—that digital literacy and text literacy are related but distinct constructs. A somewhat unexpected result was that students scored better, on average, in the digital mode, than in the text mode. Peak (2005) found that studies comparing reading across modes had, in general, found that the digital mode had a negative effect on reading performance. The longer the passages, the more pronounced the mode effect. The need to scroll, difficulty landmarking, and the inability to annotate have all been cited as contributing factors.

However, inconsistencies in the findings prompted Pommerich (2004) and Russell et al. (2003) to call on schools to test their own students, rather than allowing the state’s findings, or the literature, to guide their decision-making. Additionally, Clariana & Wallace (2002), found enhanced positive mode effect for students with high levels of content mastery. It was a small-scale study, which tested students in an academic arena other than reading. However, it is plausible that, since the population for this study was
homogeneous for reading achievement, they had a similar positive mode effect in the area of reading achievement due to content competency.

There are several other explanations that may have been responsible for, or related to, the apparent mode effect. First, the cognitive load associated with the CBT may have been under the threshold needed by this group to manifest a negative mode effect. Perhaps, with a longer test or denser reading passages, the findings would have been reversed. Second, as mentioned in the section on limitations, it is possible that the content of the CBT was more interesting to the students, and hence more engaging. Although both tests had parallel multiple-choice and short-answer essay questions based on similar linked passages (both were historical in nature), the possibility of unintended inequalities in the instruments cannot be ruled out, due to the correlation below .7. For instance, since the PPT discussed the history of silk and the CBT discussed the history of baseball, the students may have found the CBT content more enjoyable, making it seem easier. This would be an unintended inequity masquerading as mode differences. Finally, it is possible that the net benefit provided by writing on the computer was greater in magnitude than the net deficit of reading on the computer. These findings, however, do point to mode benefits being afforded to high-achieving 10th graders when the digital mode is used to deliver and assess their standards-based reading.

A crosstab provided additional analysis for this hypothesis. The chi square for this analysis was significant. The quartile data indicated that students at the ends of the achievement spectrum were most likely to have equivalent quartile scores in each mode. Thirty-eight percent of the students who scored in the lowest quartile on the PPT scored in the lowest quartile on the CBT. Forty-two percent of students who scored in the
highest quartile on the PPT also scored in the highest quartile of the CBT. There was more variability among the students scoring in the 2nd and 3rd quartiles (only 24% of students who scored in the 3rd quartile on the PPT also scored in the 3rd quartile on the CBT). This variability supports the argument that, because the constructs differ, student success in each mode will vary by individual. It lends credence to the idea of accommodating individual proclivities to optimize individual performance.

**Hypothesis 2**

Hypothesis 2 looked at mean PPT-CBT difference when controlling for a set of contextual factors. The selected contextual factors were: a) school attended, b) gender, c) reading preference (text or digital), d) amount of time spent leisure reading, e) amount of time spent on communication technologies, f) order of testing, and g) prior instruction and assessment in digital reading. The factors were selected, either because prior research had shown them to have an influence on reading achievement, or because the literature was divided on their effect.

The null hypothesis for Hypothesis 2 was accepted, insofar as no contextual factor was associated with a between-groups variance that was significant. The mean difference was significant, but none of the factors selected for subsequent analysis was associated with a significant difference. The lack of significance associated with any of the six factors tested again points back to the two modes being fundamentally different constructs. Given the factor significance that surfaced within each mode, the lack of significance associated with this hypothesis points to the shortcomings of using mean difference across mode as an analytical method for learning about students' mode-related
strengths and weakness. It is really a methodology that best serves the alignment of instruments and high-stakes scores.

Hypothesis 3

Hypothesis 3 tested the effect of a set of selected contextual factors (described above) on PPT and CBT scores. The null hypothesis for Hypothesis 3 proposed that there would be no significant differences in either mode when controlling for the selected contextual factors. The null hypothesis was rejected, as several of the selected factors were associated with a between-groups variance that was significant. Factors that were not associated with significant difference were reading preference and time spent on communication technology. As mentioned before, prior instruction and assessment in digital reading were eliminated, due to an inadequate sample of students claiming that they had been instructed and assessed in digital reading.

School attended was associated with a between-groups variance that was significant. There was a significant difference between the scores from School 2 and School 3, in both modes. In both modes, School 2 scored higher than School 3. This is consistent with other results that the district has published over the past five years, although significance is rarely tested or reported. Mean score for School 2 was also higher than for School 0 and School 1, although there was no significance associated with these differences. This finding is likely support for the literature quantifying the role of parents; level of education in student achievement. School 2 draws its student body largely from towns in the county that are associated with higher mean incomes and higher degrees of professionalism, more so than School 3.
Gender was also associated with a significant between-subjects variance. Girls performed significantly better than boys, on average, on both the PPT and the CBT. This is consistent with the findings from the literature that girls outperform boys, on average, in text reading achievement at the high school level. The literature, however, reports inconsistent findings for the mode effect when controlling for gender. While some comparability studies, including that of Gallagher, Bridgeman and Calahan (2002), found that mode had a greater negative impact on females, other studies have failed to replicate this finding. So, while it might have been anticipated that the girls would outscore the boys on the PPT, their performance on the CBT was more difficult to forecast. The findings from this study, in which girls showed no negative mode effect, may again support the role of content mastery in mitigating mode effect.

PPT and CBT scores were also associated with a significant difference when they were disaggregated by the amount of time a student spent leisure reading. This is consistent with the 2005 report by the National Endowment for the Arts, which found that reading achievement scores improved by as much as 16 points when a student engaged in more leisure reading.

**Hypothesis 4**

Hypothesis 4 compared text literacy in two different contexts: a standards-based assessment (PPT) and a norm-referenced assessment (PSAT). The associated null hypothesis stated that students' scores on the two assessments would, on average, not be related. A General Linear Model could not be used due to the different score scales, so performance on the two tests was compared using a Spearman Correlation of Rank Order.
The null hypothesis was rejected, as the correlation of .411 was significant at the .000 level, indicating that the scores were related.

While these two instruments both tested critical reading in the text mode, two factors differentiated them from the outset. First, the PSAT critical reading test is administered alongside the math and the writing components, in a single sitting. This being the case, cognitive load would have been a powerful differentiator. Additionally, the students take the PSAT very seriously as practice for the SAT, and to put them in line for Merit Scholarship. On the other hand, participation in the PPT was voluntary, and the students had nothing riding on the outcome. Motivation should have been a second, powerful differentiating factor. Nevertheless, the standards-based assessment of text literacy was more closely associated with the norm-referenced text literacy than it was with the standards-based assessment of digital literacy. This, again, lends support to the premise that text literacy and digital literacy should be viewed as distinct constructs, which need distinct instruction and assessment.

Hypothesis 5

Hypothesis 5 tested critical reading within norm-referenced text reading, to see if it was differently influenced by selected contextual factors than standards-based text reading. The null hypothesis posited that the contextual factors would influence the two tests in the same way. Order of testing was eliminated from the list of contextual factors, as it was not relevant to the PSAT. The null hypothesis was rejected. The remaining selected factors influenced scores on the two tests differently.

The factors of reading preference and amount of time spent on communication technologies were not associated with a significant difference for either type of text
literacy. School attended and leisure reading were associated with significant differences on both the standards-based assessment and the norm-referenced assessment. School 2 performed significantly better than one other school (School 3) on the standards-based assessments. On the norm-referenced assessment, School 2 performed significantly better than two of the other schools (School 0 and School 3). This is consistent with the literature that finds parent education and expectations to be associated with student success. School 2 has a high percentage of Asian students, and students whose parents work in engineering, telecom, and other related professions. Relative to the other academies, School 2 has a very high percentage of students that accelerate their progression through the math sequence. Their total PSAT scores were, on average, significantly higher than the other four academies, as were their critical reading scores viewed alone. The fact that they outscored two schools on the PSAT (versus one on the study test) might be related to their motivation associated with that test.

The factor of leisure reading was also associated with a significant between-group variance. Students who did more leisure reading scored significantly better on both tests of text literacy, than students who did less leisure reading. This is again consistent with the literature. In the NAE study (2005), students who spent increased time leisure reading experienced a 16-point boost on a norm-referenced test for reading.

Surprisingly, gender was the factor that influenced the standards-based assessment differently from the norm-referenced assessment. Girls performed significantly better than the boys on both the CBT and the PPT. However, on the PSAT, gender was not associated with a significant between-groups variance. This finding
contradicts the literature, which has described a gender gap in reading scores for at least 15 years.

Again, motivation may have been at least a part of the story here. There has been some research around boys and competition, with boys performing better when there is something at stake. Because of the high stakes nature of the PSAT, boys may have put forth a better effort than they did on the standards-based test for which they were volunteering.

**Discussion and Implications for Practice**

The literature described in Chapter II showed that studies comparing reading achievement across modes have produced conflicting results. These studies have been conducted largely by states, as they attempted to move their high-stakes testing into the online environment. When the scores have differed significantly across modes, the solution has been to use a mathematical solution called *equating*, so as not to disadvantage any student based on mode. This study took the position that digital literacy and text literacy are different constructs. The study further explored factors for their influence within each of these literacies.

Instruction and assessment in digital literacy has been pushed to the back burner in public education by the pressure to prepare students for high-stakes reading tests. High-stakes reading tests in public education continue to reside primarily in the paper and pencil realm. Digital literacy has been further marginalized by the standards through which it has been delivered and assessed over the past 15 years. The standards for digital literacy have typically been integrated standards, not associated with a core subject. This has led to inconsistent delivery and infrequent assessment.
This study of 10th graders in the Monmouth County Vocational District had some findings that corresponded with the prevailing literature, and some that diverged. The significant differences associated with school attended and amount of time spent leisure reading were very much in line with well-established research findings. Because school attended is frequently associated with socioeconomic status and level of parent education, it tends to have a large effect on student achievement, as it did here. The literature has also shown that students who do more reading in their leisure time tend to score higher on achievement tests. This study reaffirmed those findings. Additionally, gender has repeatedly been associated with significant differences in students’ reading performance. Girls, on average, outperform boys, as they did in this study. The fact that gender was not associated with a significant difference on the norm-referenced PSAT was a surprise. A possible reason may be found in the motivation of high-achieving adolescent males in high-stakes testing. Finally, the overall finding that these high achieving students performed better, on average, on the CBT, was not entirely expected. Adolescents have frequently performed less well in the digital mode when reading longer passages and scrolling. These students did not show the typical mode effect, whether because they were more motivated by the CBT, found responding on the computer easier than using paper and pencil, or another confounding factor.

Several design features of this study complicated the interpretation of this result. Use of the common-person design ensured that the samples for each mode were the same. However, it forced the use of two different instruments, introducing the possibility that students did not experience the reading passages as equivalent. Students may have found variation in the challenge provided or in their level of interest.
1. Participants in this study were volunteers. Self-selection introduced the possibility that this group already felt more comfortable with computer-based testing, or was in some way otherwise predisposed to greater success in the digital environment.

2. The fact that students did not have traditional educational outcomes (i.e., grades) associated with the results of these tests introduced the possibility of uneven effort in one mode or the other, influencing outcomes.

3. This study did not make an effort to account for, or to distinguish between, digital reading and computer-based response. A positive mode effect provided by typing responses on the computer may have masked a negative mode effect during reading.

Confounding factors notwithstanding, the evidence supported the notion that digital literacy and text literacy are different constructs. It also substantiated that moving more testing to the computer for students who have already demonstrated content competency has the potential to improve student outcomes.

Additionally, more testing in the blended mode (reading text and responding digitally) should be considered. Seventy-five percent of the students identified themselves as preferring to read text materials, and yet students performed better, on average, on the CBT for critical reading. Given the amount of time students are spending on communication technologies (83% claimed to spend more than three hours per day), it can be assumed that they are facile with a keyboard. Blended-mode testing might be the optimal combination of positive mode affordances for many students and should be explored.
The significance attributed to the variable of School Attended is a reminder of the limitations of broad generalizations applied to large groups of students. Even within this self-selected group of 114 students, deemed to be homogeneous for high achievement, significant differences in performance existed that were associated with which of the four schools the student attended. Generalities can save time, but they can also mask significant data points related to students’ success.

Gender, in this study, produced inconsistent findings. Girls outperformed boys in both modes of the standards-based testing. On the norm-referenced PSAT, however, gender was not associated with a significant difference in scores. As mentioned earlier, the most likely explanation for this inconsistency was the difference in the outcome impact of the two tests. The standards-based tests had no grades associated with them, and students volunteered to take them. The PSAT had a high-stakes outcome and publicized results. Males may have tried harder on the PSAT, motivated by the outcome. This finding goes back to the role of motivation in student achievement and differentiated motivation based on subgroup. High schools may need to do more qualitative investigation, exploring students’ own descriptions of what motivates them regarding achievement in general and reading achievement, specifically.

Increased leisure reading was associated with higher scores, on average, within all types of literacy. This finding reinforces the value of high schools’ setting school-based goals around engaging students in more leisure reading. It is unlikely that one strategy will work for all students. It is more likely that schools will need to design diverse strategies, differentiated by factors such as gender, socioeconomic status (SES), and geography (urban, rural, etc). Additionally, it is not clear how mode will fit into inspiring
high school students to do more leisure reading. With 68 of the students describing themselves as spending three or more hours per day on communication technologies, it would be easy to assume that a digital reader might inspire more leisure reading. Yet, 75 of the same students said that they preferred to read text over digital. Mode may be best used as a selectively applied strategy to increase motivation, as well as a selectively applied option for high-stakes testing.

While there has been a lot of discussion in the literature regarding the interaction between time spent on communications technologies and reading achievement, at no point in this study was there significance associated with the between-groups variance for time spent on communications technologies. It is not clear what this finding means in the broader context of students' functional literacies and their interaction with reading achievement. This finding opens the door to the possibility that the negative impact associated with communication technologies use is mitigated by something that these high-achieving students are doing, such as increased leisure reading. A more detailed exploration would be needed to understand how these students spend their time. A larger sample would also help so that multifactorial analysis could be undertaken.

It is noteworthy, however, that 98% of these students stated that they had never had digital reading instruction and assessment, and yet they still performed better, on average, on the CBT. It may be that students' exposure to computers has finally moved their academic performance in reading. It may also be that the level of reading competency of this group mitigated any mode effect. However, given the speed with which the world is moving online, instruction and assessment of digital reading should
not be haphazard or optional. All students should leave high school having demonstrated competency in digital literacy at a level aligned with their postsecondary plans.

Order of testing is known to impact outcomes in tests of many kinds, going all the way up to the exams for doctoral candidates. The level of significance associated with the order of testing is unclear in the context of this study. For this group of students, order of testing was significant only within the CBT. Students who took it first did significantly better. It may have been that the novelty of testing on the computer was at its peak for those students who had not been impacted by the cognitive load imposed by the PPT. It is also possible that the negative mode effects of the CBT, such as visual fatigue, were enhanced for students who had previously engaged with material at a high level of concentration. Qualitative investigation, such as interviewing students pre- and posttest, would provide valuable insight into the significance of these findings.

**Recommendations for Policy and Further Research**

The recommendations coming out of this study are presented in two sections. The first section has two policy recommendations extrapolated from the findings of this study, nested in the prevailing literature. The second section has five recommendations for further research, presented in the hope that adapting, expanding, and improving upon the basic methodology of this study will yield findings that improve the reading achievement of adolescents in the coming years.

**Policy Recommendations**

1. States should amend their Core Content Standards so that the standards for digital literacy (with digital reading as a component) are incorporated into the standards for Language Arts. Ninety-eight percent of these high-achieving NJ students stated that they
had not received instruction and assessment in digital reading. In NJ, digital literacy has been delivered through integrated technology standards since 1996. The data collected in this study supports findings that describe inconsistent instruction and assessment of the skills that make up digital literacy. Digital literacy needs to be the responsibility of the teachers most able to teach and assess reading at the high school level. Ensuring that students are facile with academic content that is delivered digitally is a critical task for 21st Century high schools.

2. High-stakes reading assessment should be standards-based. As highlighted in the discussion related to Hypotheses 1, 2 and 5, student performance was influence by mode and, within text literacy, performance varied between standards-based and norm-referenced testing. With standards being the driving force behind curriculum, they should also be the driving force behind assessment. Moving away from norm-referenced testing would eliminate the need for uniform testing conditions for all students. Students’ preferred mode, or combination of modes, could be more easily accommodated, maximizing success for every student.

**Recommendations for Further Research**

1. This study used a common-person design with two versions of a critical reading test. One version was digital, and one version was paper/pencil. It is recommended that this study be improved on by using an experimental design. If the common-person design is maintained, both versions of the test need to be offered in both modes. Then, students would be randomly assigned to distinct versions in each mode. Otherwise, the same version of the test should be used in each mode, and students should be randomly
assigned to mode. This will require a larger population, so that the factorial analysis will still have an adequate n.

2. This study used a population that was homogenous for high achievement. Although analysis revealed significant differences in achievement when disaggregated by school attended, this population was not intended to be described as a diverse sample. Gifted and talented students are a distinct population, with extensive research dedicated to their unique motivations and challenges. This investigation intentionally targeted that population, independent of populations of other abilities. However, broadening the population could extend this study. The sample could continue to include students known to have above-average reading skills, but also include students who are average- to below-average readers. This juxtaposition would allow relative mode affect to be measured among different populations.

3. This study attempted to use prior instruction and assessment in digital reading as a factor by which to disaggregate results. However, the number of students that had experienced prior instruction was too small to do a meaningful analysis. The literature cited prior instruction as a differentiating contextual factor. Using prior instruction as a treatment and measuring its effect would be an interesting direction to extend this study. The population would have to be larger so that the multifactorial analysis would not produce n’s that were too small.

4. This study used quantitative data exclusively. Students self-reported on a variety of contextual factors that were coded at two levels. Adding a qualitative component could extend this study. Having students describe which test they liked better, and why, would clarify some issues that this study can only speculate on. For instance, did students do
better on the CBT because they preferred typing their responses, or were they just unaffected by the digital mode? Giving the students a chance to elaborate on their relationship with reading and responding on the computer could yield some valuable information regarding student motivation, a critical piece of what is being sought.

5. When this study disaggregated the norm-referenced test scores by gender, it produced an unexpected result. There was no significant difference in scores, on average, based on gender. This contradicted most major research findings, and it contradicted the findings on the standards-based tests. Further research could be done that explores more deeply the habits and motivation of high-achieving boys. By taking a larger sample, one that included mixed ability, and including a qualitative component, a less speculative and more reliable conclusion could be drawn about this finding.

6. Finally, this study has made several references to the notion of customizing mode according to student preference, in order to maximize student performance. This study could be extended by allowing students to select their mode for testing, to see whether student performance could be enhanced by comparing a group of students allowed to select their mode of testing to a group of students assigned to a mode of testing. Students who are allowed to select should be able to mix and match test and digital reading and response.

Conclusion

Three converging factors have put the spotlight squarely on high school reading. First, expectations for postsecondary literacy are rising. Literacy expectations are being driven upwards by globalization, technology, and the transition from an industrial to an information economy. Second, test scores in critical reading for secondary students in
the U.S. have been essentially flatlined for the past 30 years. Despite 30 years of school reform, critical reading scores have proven very resistant to improvement. Finally, technology has permanently changed literacy, and education is having a hard time adapting. Adolescents' functional literacies have moved almost exclusively into the digital world, while academic reading instruction and achievement languish in the paper/pencil world.

Education and its associated research have attempted to treat digital literacy and text literacy as variations of a single construct. When the data has failed to support their equivalency, the solution has been to doctor the scores so they are equal (a process known as equating). This study supports the contention that, for the purpose of instruction and assessment, digital literacy and text literacy may be better viewed as distinct but related constructs. Viewing them as distinct entities will promote instruction and assessment in each literacy. Ensuring that all students are proficient readers within both literacies should be an essential goal for 21st Century education.
References

Achieve, Inc. (2004). *Ready or not: Creating a high school diploma that counts.*


Washington, DC. National Commission on Excellence in Education. (ERIC Document Reproduction Service No. ED226006)


## Appendix A

Statistics Related to Instrument Reliability and Validity

Table 1

<table>
<thead>
<tr>
<th>Grade</th>
<th>Model</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>( \chi^2/df )</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS Reading</td>
<td>Single-factor</td>
<td>22757.56</td>
<td>629</td>
<td>36.18</td>
<td>0.95</td>
<td>0.056-0.057</td>
</tr>
<tr>
<td></td>
<td>Multi-factor</td>
<td>22136.03</td>
<td>614</td>
<td>36.05</td>
<td>0.95</td>
<td>0.056-0.057</td>
</tr>
<tr>
<td>Comparison</td>
<td></td>
<td>621.53</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2

Reliability for the CBT: Test & Content Strand Reliability Estimates

<table>
<thead>
<tr>
<th>Strand</th>
<th>Alpha Coefficient</th>
<th>Raw Score Standard Error of Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>.86</td>
<td>3.03</td>
</tr>
<tr>
<td>LC</td>
<td>.46</td>
<td>.96</td>
</tr>
<tr>
<td>LA</td>
<td>.60</td>
<td>1.45</td>
</tr>
<tr>
<td>LT</td>
<td>.40</td>
<td>1.16</td>
</tr>
<tr>
<td>IC</td>
<td>.55</td>
<td>1.07</td>
</tr>
<tr>
<td>IA</td>
<td>.50</td>
<td>1.34</td>
</tr>
<tr>
<td>IT</td>
<td>.47</td>
<td>1.24</td>
</tr>
<tr>
<td>Writing*</td>
<td>.78</td>
<td>1.50</td>
</tr>
<tr>
<td>COS</td>
<td>.72</td>
<td>1.22</td>
</tr>
<tr>
<td>CONV</td>
<td>.74</td>
<td>.71</td>
</tr>
</tbody>
</table>

*Writing was not assessed in this study but was retained in this table for statistical consistency. 2007 HS WASL Technical Report, p. 42.
Table 3

Validity for the PPT: Model Goodness-of-Fit Statistics

<table>
<thead>
<tr>
<th>Grade</th>
<th>Model</th>
<th>$X^2$</th>
<th>df</th>
<th>$X^2$/df</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS Reading</td>
<td>Single-factor</td>
<td>19466.44</td>
<td>629</td>
<td>30.95</td>
<td>0.95</td>
<td>0.050-0.051</td>
</tr>
<tr>
<td></td>
<td>Multi-factor</td>
<td>18097.83</td>
<td>614</td>
<td>29.48</td>
<td>0.95</td>
<td>0.049-0.050</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>1368.61</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2008 HS WASL Technical Report, p. 37
Table 4

Reliability for the PPT: Test & Content Strand Reliability Estimates

<table>
<thead>
<tr>
<th>Strand</th>
<th>Alpha Coefficient</th>
<th>Raw Score Standard Error of Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>.87</td>
<td>2.83</td>
</tr>
<tr>
<td>LC</td>
<td>.45</td>
<td>1.08</td>
</tr>
<tr>
<td>LA</td>
<td>.49</td>
<td>1.19</td>
</tr>
<tr>
<td>LT</td>
<td>.41</td>
<td>1.36</td>
</tr>
<tr>
<td>IC</td>
<td>.66</td>
<td>1.14</td>
</tr>
<tr>
<td>IA</td>
<td>.53</td>
<td>0.81</td>
</tr>
<tr>
<td>IT</td>
<td>.50</td>
<td>1.50</td>
</tr>
<tr>
<td>Writing*</td>
<td>.76</td>
<td>1.48</td>
</tr>
<tr>
<td>COS</td>
<td>.67</td>
<td>1.25</td>
</tr>
<tr>
<td>CONV</td>
<td>.72</td>
<td>0.67</td>
</tr>
</tbody>
</table>

*Writing was not assessed in this study but was retained in this table for statistical consistency. 2008 HS WASL Technical Report, p. 42.
Appendix B

Instrument: Paper/Pencil Based Test
PAPER AND PENCIL
ASSESSMENT OF READING

NOTE: THE FORMATTING OF THIS ASSESSMENT HAS BEEN MODIFIED FOR PRINT PURPOSES

Adapted From:
WASL - Washington Assessment of Student Learning
A Component of the Washington State Assessment System

Copyright © December 2008

All rights reserved. Educational institutions within the State of Washington have permission to reproduce this document. All other individuals wishing to reproduce this document must contact OSPI.
ACKNOWLEDGMENTS


Directions to the Student

Today you will take two reading assessments based on a tenth grade reading assessment from Washington State. You will take one part using pencil and paper and one part on the computer. Each test consists of two linked passages. Your answers will be used to find out how well you read on the computer compared to how well you read paper text. You will now begin your First section, the Paper/Pencil Test.

You will read a story and a related selection and answer some questions. You may look back at the story or selection when you are answering the questions. There are two different types of questions. There are multiple-choice questions that require you to choose the best answer and there are short-answer questions for which you will write one or more paragraphs. You will put both types of answers right on the Scantron.

Sample questions have been included. These sample questions do not relate to the selections you are about to read. They have been included to show you the two types of questions you will find in the booklet and how to mark or write your answers.

There are several important things to remember:

1. Complete each reading. You may look back at the reading selection as often as you want.
2. The paragraphs in the reading passages are numbered to help you remember where information is found.
3. Read each question carefully. Then choose or write the answer that you think is best.
4. When you are supposed to write your answers, write them neatly and clearly on the lines provided. Cross out or erase any part of your work you do not want to include as part of your answer. For short-answer questions be sure to write complete answers and write neatly so your answers can be read.
5. Use only a No. 2 pencil, not a pen, to write or mark your answers directly in the space provided in your booklet. If you do not have a No. 2 pencil, ask your teacher to give you one.
6. You have 35 minutes to complete this section. If you do not know the answer to a question, go to the next question. You can come back to that question later.
7. If you finish early, you may check over your work in this Reading session only.
8. When you reach the end you may go back and check your work and then close your book.
Sample Questions

To help you understand how to answer the test questions, look at the sample questions below. These questions do not refer to the selections you are about to read. They are included to show you what the questions in the test are like and how to mark your answers on the computer.

Multiple-Choice Sample Question

For this type of question you will select the answer and fill it in on the Scantron.

0. According to the bar graph, which of these planes flies the fastest?
  A. The Boeing 747
  B. The Concorde
  C. The DC-10
  • D. The SR-71

For this sample question, the correct answer was D. Therefore, the circle D on the Scantron would be filled in completely.

Short-Answer Essay (1)

For this question you will type an answer consisting of approximately 1 well-constructed paragraph. It should be built on information from your reading and clearly answer the question. Write directly on the Scantron.

Extended-Answer Essay (1)

For this question you will write an answer consisting of approximately 3 well-constructed paragraphs. It should contain specific information from your reading and clearly answer the question. Write directly on the Scantron.

STOP: PROCEED TO TEST WHEN INSTRUCTED BY THE TEACHER
Silk: The Caterpillar Thread
by Laurel Kendall

1 Thousands of tiny jaws crunch mulberry leaves, hour after hour and day after day, all day and all night, with a pause now and again to shed an outgrown skin. This is the lifelong banquet of the Bombyx mori, the tiny caterpillar we call “silkworm.” In twenty-five days or more, the feast is done, and the silkworm spins its thick cocoon, a continuous strand of liquid silk that hardens when it touches the air.

2 The cocoons must be washed in very hot water to remove the sticky coating from the silken strand. Nimble fingers, usually women’s fingers, gather her thin threads from several cocoons and reel them into a single long and lustrous thread, strong enough for sewing or weaving.
3. Silk is perhaps the world’s most wondrous fabric. We say “silky” or “silken” to describe lovely skin and hair and often use the expression “as soft as silk.” But silk is also strong, stronger than a steel wire the same size as a thin silken thread. Japanese samurai used silk cords to bind together the pieces of their armor. Surgeons use silk thread to stitch wounds and incisions. Pilots trust their lives to the sturdy silk of their parachutes. And silk is beautiful. Sometimes we say “silken” when we mean luxurious. Silk robes embroidered with dragons of silk thread were clothing fit for a Chinese emperor. Today, even a simple dress or shirt made of silk is a very special piece of clothing.

4. Nearly four thousand years ago, Chinese farmers living along the Yellow River in north-central China unraveled the caterpillar’s threads and began to spin and weave silk. From at least the second century B.C., Chinese silk was traded over great distances to the world outside China. Silk reached the Roman Empire nearly two thousand years ago, and the luxury-loving Romans paid great sums of gold to clothe themselves in silk. For many centuries, Chinese silk was worn in Persia, in Indian kingdoms, in the Middle East, and around the Mediterranean Sea.

- Glass
- Apples
- Ivory
- Carrots
- Pomegranates
- Paper
5. Brave traders traveled the Silk Road, a network of difficult, often dangerous roads over mountains and across deserts. Some silk traveled a distance of nearly forty-five hundred miles from western China, over the high Pamir Mountains, and through what is now Iran. Silk for the European and North African market was carried to ports in the eastern Mediter

6. No single merchant caravan traveled the entire length of the Silk Road. Silk and other valuable goods changed hands many times as they were traded and retraded along the way. If merchants survived attacks by bandits, hunger, thirst, and exhaustion, they would grow rich. States along the Silk Road encouraged trade by protecting caravans, and through trade, their cities prospered.

7. As a link between East and West, the Silk Road influenced world history. Buddhist monks followed the Silk Road and brought Buddhism to China. Many splendid Buddhist shrines still remain in central Asia. Other travelers began to venture further along the trade routes and to write about the peoples they encountered. The most famous Western traveler was Marco Polo, who left Venice as a boy in 1271 and followed the Silk Road to China. He returned twenty-four years later and wrote about his adventures in the land of Kublai Khan. The wonders he described were so fantastic that many refused to believe him. Others, such as Christopher Columbus, wanted nothing more than to follow in Marco Polo's footsteps. Columbus, seeking a route to China by sea, bumped into the New World before he ever reached the Orient.
8. Even though silk cloth traveled thousands of miles, the marvel of the silk-producing caterpillar was, for a long time, China’s secret. Romans thought that the shimmering cloth came from “the hair of a sea shrimp.” Eventually, the secret reached other lands—Japan and Korea in the third century, central Asia in the fourth, Byzantium, in the eastern Mediterranean, in the sixth. According to one legend, monks arrived in Byzantium with the silkworm hidden in their hollow walking sticks. Even though silk could be manufactured outside China, Chinese silk remained a very special commodity. Today, although the secret is out, China is still the world’s foremost producer of silk.

Please answer the following questions:

1. What is the purpose of the text box Brought to You by the Silk Road?
   A. To identify some items traded on the Silk Road
   B. To describe the trade routes on the Silk Road
   C. To explain how silk was traded for jade
   D. To list some products made from silk

2. Why did so many traders risk the dangers of the Silk Road?
   A. To become Buddhists
   B. To buy silk clothing
   C. To become rich
   D. To follow in Marco Polo’s footsteps

3. Silk fabric is made from:
   A. Caterpillars
   B. Threads spun by caterpillars
   C. Hair
   D. Skin

4. What is the main idea of the passage?
   1. The Silk Road was dangerous.
   2. The Silk Road brought Buddhism to China.
   3. Silk is a valuable commodity that was originally produced in China.
   4. Silk is a wonderous fabric.
I. Extended Essay.

Write a summary of the selection *Silk Road*. Be sure to include a topic sentence and three important ideas from the selection in your summary.

**PLEASE PUT YOUR ANSWER ON THE SCANTRON!!!**
Directions: Read this story, which goes with the selection you just read. Then answer the questions.

Journey on the Silk Road
by Luann Hankom

1 My name is Fa Zang. I am 12 years old, and the year is A.D. 742. I am excited! I am joining my father on my first caravan to a far-off city called Dunhuang. My father has obtained porcelain, rhubarb, herbal medicine, and silk cloth to trade. What treasures will we find on our journey?

2 We begin our journey in Chang’an, China, where we live. Chang’an is a bustling city with two million people. Our caravan includes private merchants such as my father, Chinese government officials, and of course, camels. Camels may be slow, but they are sturdy animals that can carry our heavy loads.

3 Our prized trading item is silk, which comes from silkworms. People in foreign lands use our silk cloth for fancy clothes.
We leave Chang’an and travel through the Wei River Valley along the Imperial Highway. The landscape is green and yellow—bright green fields and mulberry trees. The ground is yellow with loess, a fine dust that blows in the wind. If the wind is harsh, I will put a mask over my face, so the dust doesn’t get inside my mouth or eyes.

At night, my feet are sore from walking. Our caravan stops at a shelter, so we don’t have to sleep out in the open. Other traders are at the shelter, too. They have dates, pistachio nuts, peaches, and pears. Someone tosses me a pear. Its sweet, slippery juice drips down my chin while I eat it.

The days and nights continue. We stop at farms for food along the way and meet caravans coming and going. We continue northwest through forests and hills and cross the Huang River, sloshing through the water. We travel the foothills of the Nan Shan Mountains until we reach Dunhuang. I am tired and sore from the journey that has lasted many weeks. I am intrigued as my father starts exchanging goods with caravans from the West.

There are rare items such as green and white jade, fine-colored glass, and exotic perfumes. My father trades his silk for white jade and Persian metalwork. He trades the rhubarb for pistachio nuts and walnuts. He exchanges the herbal medicine for musical instruments. The government officials trade silk for horses. The officials are pleased—the horses will be for the emperor’s army.
8 The men from the West describe unusual, foreign places on their journeys: Tyre and Byzantium. I have not heard of these cities before. They speak of the difficult journeys through the Taklamakan Desert and the Pamirs. Such adventures!

9 I can barely fall asleep, for thoughts of these exotic places and peoples fill my brain. I will travel to these cities someday! My father and I will travel back to Chang'an, so I must get my rest. It has been an exciting journey. I dream of the travels yet to come.

Please answer the following questions:

5. According to the story, which word best describes Fa Zang?
   A. Respected
   B. Confused
   C. Worried
   D. Amazed

6. What is the main idea of the story?
   A. Fa Zang learns how to ride a camel.
   B. Fa Zang receives gifts while traveling.
   C. Fa Zang takes her first caravan trip with her father.
   D. Fa Zang travels through the scenic Wei River Valley.
7. Based on the information in the story, what inference can the reader make about Fa Zang’s attitude toward her journey?

   A. She embraces the new experience.
   B. She distrusts the merchants from the West.
   C. She is unaware of the significance of the trip.
   D. She is invigorated by the long walk to Dunhuang.

8. Fa Zang sees her journey on the Silk Road as an adventure, but in actuality it is:

   A. A way for her father to get pistachio nuts and walnuts.
   B. A way for her family to survive.
   C. A way
   D. for her to see the world.
   E. A way for her father to meet new people.

9. What is the main similarity between travelers on the Silk Road in both the selection and the story?

   A. Travelers endured hardships along the Silk Road.
   B. Travelers slept in shelters along the Silk Road.
   C. Travelers rode camels along the Silk Road.
   D. Travelers were safe along the Silk Road.
10. According to both the selection and the story, which sentence best explains why silk cloth is popular around the world?

   A. Silk cloth is a sturdy material.
   B. Silk cloth is easy for workers to produce.
   C. Silk cloth is a common material accessible to most people.
   D. Silk cloth is considered a luxurious fabric used for clothes.

II. Short Answer Essay:

Both the selection and the story show that silk is a valuable product to trade. Write a paragraph that supports this idea using at least one detail from “The Caterpillar Thread” and one detail from “Journey on the Silk Road”.

PLEASE PUT YOUR ANSWER ON THE SCANTRON!!!!

STOP
Appendix C

Instrument: Computer-Based Test
ACKNOWLEDGMENTS


Reading Assessment

Directions to the Student

Today you will take two reading assessments. You will take one part using pencil and paper and one part using the computer. Each test consists of two passages. Your answers used to find out how well you understand what you read.

You will read stories and selections and answer some questions. You may look back at the story or selection when you are answering the questions. There are three different types of questions. There are 6 multiple-choice questions, a summarizing question linked to one essay, and a question that requires information from both readings. All answers are typed directly into the computer.

Sample questions have been included to show you the two types of questions you will find in the booklet and how to mark or write your answers.

There are several important things to remember:

1. Read each selection. You may look back at the reading selection as often as you want.
2. The paragraphs in the reading passages are numbered. A question about a particular paragraph may refer to the paragraph number.
3. Read each question carefully. Then choose or type the answer that you think is best.
4. For multiple choice questions, click the circle next to the answer that you think is best.
5. For short-answer questions, you may have more space than you need. You do not need to fill the whole space. Be sure to write complete answers. You can delete and retype anything you are not happy with.
6. You have 45 minutes to complete this section. If you do not know the answer to a question, go to the next question. You can come back to that question later.
7. If you finish early, you may check over your work in this Reading session only.
8. When you reach the word STOP on this test, you are at the end. If this is your second test you are done. If this is your first test, you must wait to go onto the second section until the teacher instructs you.
Sample Questions

To help you understand how to answer the test questions, look at the sample questions below. These questions do not refer to the selections you are about to read. They are included to show you what the questions in the test are like and how to mark or write your answers in your test booklet.

Multiple-Choice Sample Question

For this type of question you will select the answer and click the circle next to it.

According to the bar graph, which of these planes flies the fastest?

- A. The Boeing 747
- B. The Concorde
- C. The DC-10
- D. The SR-71

For this sample question, the correct answer was D. Therefore, the circle next to D was clicked and showed up as selected.

Short-Answer Sample Question

For this type of question you will type a short answer consisting of a few phrases or sentences into the textbox provided. You should include information from your reading in your answer.

What are two similarities between Matt and LeShaun? Include information from both the selection and the story in your answer.

Both Matt and LeShaun like to read mystery novels. Also, they are both helpful because they both rescued the bird when it flew into the abandoned shed.
In the Beginning
by Janet Wyman Coleman with Elizabeth V. Warren

Bats and Balls
1 Before there was a game called baseball, Americans had discovered the fun of swinging a stick at a ball. In the early 1800s, children held tree limbs above their shoulders and swatted at walnuts wrapped in rags. Adults swung at balls with the same enthusiasm. Broomsticks made great bats, as did large pieces of wood called “wagon tongues,” named after the part of a wagon that jutted out and held the horses’ reins. If players had the skill and time, they carved and sanded pieces of ash or hickory into long, graceful bats. Sometimes the bats were painted with a faux (false) grain to imitate the look of expensive woods and then used as trophies of good games. Balls were also made by hand, of rags, pieces of old mattress fabric, or horsehide.

2 The simple equipment made it possible to play “ball” almost anywhere. Soldiers enjoyed a game at Valley Forge during the Revolutionary War, and the Indian leader Geronimo fielded a team of Apaches against the U.S. Army at Fort Sill, Oklahoma, in the late 1800s. The Apaches won.
Everyone Wants to Play
3 In the 1840s and 1850s, thousands of ambitious young men left their families in Europe and immigrated to New York. They took jobs as policemen, firemen, and shipbuilders, and discovered baseball. Many of the immigrants were proud to be Americans and they wanted to play the American game. Teams evolved out of the different professions. Shipbuilders pitched to firemen. Undertakers caught fly balls hit by doctors. Schoolteachers tagged out bartenders on fields and lots around the city.

This pigskin ball looks like it was well used.

4 However, there was a problem with the balls. They couldn’t be thrown very far, because they were so light. Doc Adams of the Knickerbocker Club found a saddler who taught him how to sew horsehide and stuff it with rubber cuttings. At first, Adams made the balls himself “not only for our club but for other clubs when they were organized,” but soon workers in the leather trade were also producing and selling balls. By the late 1860s, demand was so great, baseballs had to be mass-produced in factories.

5 In 1857, the Knickerbockers and fifteen other clubs that played by the same rules created a league, the National Association of Base Ball Players. Doc Adams was the president. It was decided that baseball must continue to be an amateur game. Money would be its ruination, so the players should never be paid.
Historical Note:

By 1861, there were sixty-two teams in various states. In 1869, admission was charged and players were paid for the first time. As heroes of the game emerged in the 1880s, baseball cards were created and circulated among fans. In America, baseball remains a popular sport and continues to grow in popularity in other parts of the world.

Please answer the following questions:

1. What is the most important idea the author presents in the selection “In the Beginning”?

   - A. Interest in baseball increased during the 1800s.
   - B. Factories began producing baseballs in the 1860s.
   - C. In the late 1800s, U.S. soldiers played baseball with Apache Indians.
   - D. In the 1840s and 1850s, many immigrants played baseball in New York.

2. Based on the information in the selection “In the Beginning,” what conclusion can the reader draw about the impact of baseball in America?

   - A. Baseball inspired young men to join the military.
   - B. Baseball was a model for other amateur sports.
   - C. Baseball was unappealing to immigrants.
   - D. Baseball acted as a unifying force.
3. Why was it so simple for people to play baseball almost anywhere?
   
   o A. It was fun to swing a stick at a ball.
   o B. There were a lot of people around to play.
   o C. Anyone could find something to use to make a bat and a ball.
   o D. Soldiers played at Valley Forge.

4 Why did baseball players not get paid when teams were first formed?
   
   o A. The sport wasn’t popular enough to make money.
   o B. There were too many teams in too many states.
   o C. People refused to pay admission.
   o D. People believed that money would ruin the fun of the game.

**I. Extended Answer Essay:** Write a summary of the section *Everyone Wants to Play*. Be sure to include a topic sentence and **three** important ideas from the selection in your summary.

--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------

Go on to next Section
Directions: Read this selection, which goes with the selection you just read. Then answer the questions.

Baseball Smarts
by Bill Wise

Put yourself in Charlie Waitt’s shoes for a moment.

It’s 1875 and you play big-league baseball for the St. Louis Brown Stockings. You’re a 21-year-old rookie outfielder who’ll be filling in at first base for your team’s next game.

The thought of playing first base makes you wince. You’ve played there twice before, and it’s nothing like playing the outfield. At first base, you’ll be catching sharply hit ground balls, whistling line drives, and stinging throws, all from close range. The last time you played first base, your hands ached for days.

In 1875, baseball is played bare-handed!

Charlie knows that fielding a baseball without a glove is tricky. Players hold their hands in the shape of a box to keep the ball from hitting their palms. This works well for outfielders; by the time the ball has traveled that far, it usually has slowed down quite a bit.

Infielders and catchers, however, aren’t so lucky. They often get cuts, bruises, and even broken bones from the hard-hit balls.
Players think the banged-up fingers and hands are just a part of the game. But Charlie Waitt decides to do something about it. Waitt takes an ordinary leather glove and cuts off the fingers. He hopes that the leather will reduce the sting of the ball.

Other players and the fans watching the game don’t like Waitt’s idea. They think wearing a glove is a sign of weakness. But Waitt finds that the glove does make a difference, and he continues to wear it in other games. Eventually, other players begin to wear gloves, too. Albert Spalding, a star pitcher for the Boston Red Stockings, asked Waitt about his glove.

“Waitt confessed that he was ashamed to wear it, but he said he had it on to save his hand,” Spalding later wrote. “He also admitted that he had chosen a color as inconspicuous as possible because he didn’t care to attract attention.”

Waitt’s invention made a lasting impression on Spalding. Later, after becoming a first baseman, Spalding began wearing a glove. But he went one step further and added a thin layer of padding inside the glove for more protection.
Spalding was highly respected, and his use of the glove convinced others that it was all right to wear one. Some players even began to wear gloves on both hands. This new Spalding-style glove worked fine for everyone except the catchers. They needed more protection. Recognizing this need, former player Harry Decker designed a heavily padded mitt in 1890. It wasn’t nearly as big as today’s catcher’s mitts, but it was a big improvement over the thin gloves worn by the rest of the fielders.

By 1896, every big-league player was using a glove. There were far fewer injuries and errors as a result. Albert Spalding went on to found the Spalding Sporting Goods Company. But what about Charlie Waitt?
Charlie played in 113 professional games from 1875 to 1883, never spending more than one season with any team. He was what baseball folks call a journeyman ballplayer.

A newspaper article in 1882 declared that “a more honest and harder-working player than Charles Waitt would be hard to find.” It would also be hard to find a player who did more to change the way a baseball is fielded.

Please Answer the Following Questions:

5. Why did some players become convinced it was acceptable to use gloves?
   - o A. Fans urged their favorite players to use gloves.
   - o B. Players thought gloves made them appear tough.
   - o C. Albert Spalding used a glove and he was well-respected.
   - o D. Charlie Waitt designed a glove that was small and heavily padded.

6. What are the authors’ purposes for writing both selections?
   - o A. To explain the development of professional baseball teams
   - o B. To explain the popularity of baseball in the United States
   - o C. To explain the development of baseball in the 1800s
   - o D. To explain advances in the baseball glove
7. What is the main similarity between Doc Adams and Albert Spalding?
   o A. Both made baseball equipment.
   o B. Both used broomsticks for bats.
   o C. Both men started baseball leagues.
   o D. Both men were paid to play baseball.

8. Based on both selections, what inference can the reader make about Doc Adams and Charlie Waitt?
   o A. They were concerned about injuries baseball players suffered.
   o B. They were inventive people who found creative solutions to problems.
   o C. They were focused on making the game of baseball available to more people.
   o D. They were competitive people who wanted to change the rules of baseball.

9. Why did the idea of playing first base make Charlie Waitt wince?
   o A. He didn't want to get hurt.
   o B. He thought he wouldn't be able to catch the ball.
   o C. He had more fun in the outfield.
   o D. He didn't want to wear a glove.
10. What is the author's attitude about Charlie Waitt?
   - Condescension
   - Admiration
   - Confusion
   - Neutral

II. Short Answer Essay:

Both selections explain how baseball changed over time. Write a paragraph that supports this idea using at least one detail from "In the Beginning" and one detail from "Baseball Smarts".

STOP