8-9-2010

The Exploration of Demographics and Computer Adaptive Testing in Predicting Performance on State-Mandated Reading Assessments

Amy L. Maziarz
Indiana University of Pennsylvania

THE EXPLORATION OF DEMOGRAPHICS AND COMPUTER ADAPTIVE TESTING IN PREDICTING PERFORMANCE ON STATE-MANDATED READING ASSESSMENTS

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Education

Amy L. Maziarz

Indiana University of Pennsylvania

August 2010

Indiana University of Pennsylvania
The School of Graduate Studies and Research
Department of Educational and School Psychology

We hereby approve the dissertation of

Amy L. Maziarz

Candidate for the degree of Doctor of Education

6/2/2010 — Signature on File
William F. Barker, Ph.D.
Professor of Educational and School Psychology, Advisor

6/2/2010 — Signature on File
Wendy F. Dover, Ed.D.
Director of Exceptional Student Education
Rock Hill School District #3

6/2/2010 — Signature on File
Joseph F. Kovaleski, D.Ed.
Professor of Educational and School Psychology

6/2/2010 — Signature on File
Edward M. Levinson, Ed.D.
Professor of Educational and School Psychology

ACCEPTED

Timothy P. Mack, Ph.D.
Dean
The School of Graduate Studies and Research

Title: The Exploration of Demographics and Computer Adaptive Testing in Predicting Performance on State-Mandated Reading Assessments

Author: Amy L. Maziarz

Dissertation Chair: William F. Barker, Ph.D.

Dissertation Committee: Wendy F. Dover, Ed.D.
Joseph F. Kovaleski, D.Ed.
Edward M. Levinson, Ed.D.

No Child Left Behind (NCLB, 2001) included a broad spectrum of changes to the federal role in public education, including accountability provisions that mandated states to test all students. In an atmosphere of educational reform and federally mandated high-stakes testing, demands have increased for progress monitoring strategies that reliably predict outcomes on statewide assessments. This study investigated the predictive validity of demographic variables and the Measures of Academic Progress (MAP) Reading in relation to student performance on the South Carolina's Palmetto Assessment of State Standards (PASS) English language arts (ELA) test. Various demographic predictive factors of student performance were analyzed including sex, race/ethnicity, socioeconomic status, special education, and grade. The specific MAP predictive factors included the MAP Reading RIT score as well as the three MAP Goal Performance areas (i.e., Understanding and Using Literary Texts, Understanding and Using Informational Texts, and Building Vocabulary).

Archival test data and demographic information were obtained from five elementary and three middle schools located in the target school district. The sample was comprised of 3,861 students in grades 3-8. The data were analyzed using associational measures based on Cross-tabulation, Multi-factorial Analysis of Variance, Pearson correlation, and Multiple Linear Regression leading to the construction of a hypothetical path model. The main conclusions of the statistical analysis were that: (1)There were no relationships of practical significance between the demographic variables and the PASS ELA scores; (2) There were significant correlations

between the various MAP scores; and (3) Although the correlations were statistically significant between the MAP scores and the PASS ELA scores, the very small effect sizes implied that the linear relationships have little practical importance. In conclusion, while evidence was provided to indicate that the overall model, including the three MAP scores, was statistically significant, the low effect size was indicative of a model that had limited mathematical ability to accurately predict the PASS ELA scores.

ACKNOWLEDGEMENTS

Though only my name appears on the cover of this dissertation, a great many people have contributed to its production. I owe my sincere gratitude to all those people who have made this dissertation possible.

To Dr. William Barker, Dr. Wendy Dover, Dr. Joseph Kovaleski, and Dr. Edward Levinson - Thank you all for your expertise, time, and feedback. I appreciate everything you have done for me to make this whole process possible. A special thank you to my advisor, Dr. William Barker, your supervision and support from the preliminary to the concluding level enabled me to develop a better understanding of statistical analysis.

To Jess, Andrea, and Sally - My doctoral studies would not have been the same without the academic support and social diversions provided by my trusty cohort. The four of us were like the points on a Compass Rose, but without you three, I would have never been able to navigate my way.

To Wendy Wells – Thank you for your editorial grace. You have a knack for providing corrective feedback in the most positive way.

To the Special Services Gang – Thank you all for everything. From the daily words of encouragement, to covering for me when my time and attention were focused on this, you truly are an amazing group of individuals to work with.

Most importantly, none of this would have been possible without the love, patience, and support of my family.

To Tom, Kelly, Tim, and Laurie - Thank you for all the support and understanding throughout the years. I am truly grateful that not only are we family, but that I also consider you friends. I know I have missed numerous important family events over the past several years, but you always offered understanding and for that, I am grateful. You have also helped me celebrate many milestones throughout this process and nobody celebrates better than our family! Finally, I would like to thank you for bring into this world, seven of the most amazing children I will ever know.

To Ryan, Maddie, Kyle, Tyler, Abbie, Kaitlyn, and Emmie – Thank you for your smiling faces, constant laughter, and pure enjoyment of life. I regret not being able to spend more time with you these past several years, but the times that I did were deeply cherished and renewed my spirit!

To Shannon – The words "Thank You" do not even approach expressing my gratitude or indebtedness to you. By far, you got the brunt end of the stick in this whole process, as you had to live with me on a daily basis. The emotional support, encouragement, and understanding given by you while not always acknowledged by me were always appreciated. While the completion of my doctorate is a milestone in my life, it is not "The most significant thing I'll ever do". Thank you for being a constant reminder of what that is!

To my Parents – I wish to dedicate this dissertation to you. From the very beginning, you instilled in me that which was required to complete this journey -- high expectations, a strong work ethic, and perseverance. In addition, you graciously provided the needed support, both emotionally and financially, to make my goal a reality. Thank you for always believing in me even when I did not believe in myself. I love you both very much and am thankful every day for the life you have given me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The effectiveness of public education in the United States has long been scrutinized in our society. Publications, such as the *Coleman Report* (Coleman et al., 1966) and *A Nation at Risk* (U.S. Department of Education, 1983), have argued the United States will not remain competitive in a global market without substantial changes to American schools. As a result, the ability to measure school effectiveness has become a major issue in educational policy and legislation. The most recent attempt at such legislation occurred in January of 2002, when then President Bush signed the No Child Left Behind (NCLB, 2001) act into law. NCLB set forth a broad spectrum of changes to the federal role in public education, including accountability provisions that mandated states to test all students.

Scores from these tests are used for a variety of purposes, from identifying whether individual students are proficient with respect to designated academic standards, to helping determine whether schools are achieving adequate yearly progress (AYP). Under NCLB (2001), schools consistently not meeting AYP may be placed under sanctions requiring additional educational opportunities for students. If a continuing lack of progress is evident, proposed sanctions include allowing students to transfer to schools with better performance and, eventually, the closing of schools that fail to meet standards. In an atmosphere of educational reform and federally mandated high-stakes testing, demands have increased for progress monitoring strategies that reliably predict outcomes on statewide assessments (McGlinchey & Hixson, 2004).

Statement of the Problem

Each year, students in South Carolina participate in testing as part of South Carolina's accountability program. In previous years, statewide testing accountability requirements were met in the state through the administration of the Palmetto Achievement Challenge Test (PACT). Recently, efforts have been made to unify South Carolina's state and federal school accountability systems. In June 2008, the South Carolina legislature amended the Education

Accountability Act of 1998 (Act 282) to incorporate a new assessment in grades 3-8 known as the Palmetto Assessment of State Standards (PASS). According to the South Carolina Department of Education (SCDOE, n.d.a), beginning with the 2008-2009 school year, the PASS will be used for statewide accountability testing purposes and will include tests in writing, English language arts (reading and research), mathematics, science, and social studies. In addition, the revisions to the state's standardized testing system will result in a reduction in the amount of time students are tested as well as provide more detailed information regarding individual student performance to the teachers (SCDOE, 2007).

In addition to the PASS, the use of formative assessments is increasing in schools throughout South Carolina. To address the need for reliable and valid formative assessments available to schools, the SCDOE created a statewide adoption list of formative assessments. These assessments satisfy professional measurement standards and align with the South Carolina Academic Standards. One of the assessments on the final adoption list of formative assessments for the 2008-2009 school year was the Measures of Academic Progress (MAP) published by the Northwest Evaluation Association ([NWEA], SCDOE, n.d.b). The MAP is designed to deliver assessments matched to the capabilities of each individual student through computerized adaptive testing (CAT).

In order to use the MAP as a reliable progress-monitoring tool, an alignment of the scores from the state and the MAP tests is as important as the curriculum alignment. NWEA (2007a) conducts regular state alignment studies to examine the correspondence between the MAP and state standardized tests used to measure student achievement. Each alignment study identifies the specific scores from the MAP that correspond to the various proficiency levels for each subject (i.e., reading, mathematics, etc.) and for each student grade. Alignment studies also estimate the probability that a specific score on MAP would achieve a status of "proficient" or better on the state test. In addition to scores, Goal Performance areas are also reported and are aligned to the content of the individual state standards or benchmarks. NWEA has conducted four prior studies to investigate the alignment of cut scores between the spring administrations of the PACT and the MAP tests (Hauser, 2002; Cronin & Hauser, 2003; Cronin, 2004; Cronin, 2007). These studies have suggested that the MAP tests are reliable and valid predictors of student performance on the PACT. The present study sought to extend these findings to the newly implemented statewide assessment, the PASS.

Research Questions

This study investigated the predictive validity of demographic variables and the MAP in relation to student performance on the South Carolina's PASS English language arts (ELA) test. The state test was selected because it is the measure connected to high-stakes for school districts and it is viewed as a comprehensive measure of reading skill. Various predictive factors of student performance on the PASS ELA test were analyzed. Demographic factors included sex, race/ethnicity, socioeconomic status (SES), special education, and degree of exposure to CAT. The specific predictive factors of the MAP that were analyzed included the MAP Reading Rasch Unit (RIT) score, the MAP Understanding and Using Literary Texts Goal Performance area, the MAP Understanding and Using Informational Texts Goal Performance area, and the MAP Building Vocabulary Goal Performance area. From among the demographic variables and MAP scores analyzed, the aim was to identify the best predictors of student performance on the PASS ELA. In order to accomplish this, the following research questions were explored:

1. What is the relationship between and among the demographic variables and student performance on the PASS ELA? How do these variables interrelate?
    - 1.1. Is there a difference in student performance between males and females on the PASS ELA?
    - 1.2. Is there a difference in student performance between American Indian or Alaska Native, Asian or Other Pacific Islander, Black, Hispanic, and White students on the PASS ELA?

1.3. Is there a difference in student performance between students who receive free/reduced lunch and students who are not eligible for free/reduced lunch on the PASS ELA?

1.4. Is there a difference in student performance between disabled and nondisabled students on the PASS ELA?

1.5. Is there a difference in student performance between students in grades 3-8 on the PASS ELA?

2. Is a student's performance on the MAP Reading able to predict his or her performance on the PASS ELA?

2.1. Is there any association between the MAP Understanding and Using Literary Texts Goal Performance area and the PASS ELA?

2.2. Is there any association between the MAP Understanding and Using Informational Texts Goal Performance area and the PASS ELA?

2.3. Is there any association between the MAP Building Vocabulary Goal Performance area and the PASS ELA?

3. From among the demographic variables and MAP scores identified, what is/are the best predictor(s) of student performance on the PASS ELA?

Hypotheses

To test the following hypotheses, demographic information and archival MAP Reading and PASS ELA test data were collected from students in grades 3-8 from five elementary and three middle schools located in the target school district.

The hypotheses of this study were as follows:

Hypothesis H_1 – There is a clinically significant difference in student performance between and among sub-categories of the various demographic variables on the PASS ELA.

Hypothesis $H_{1.1.}$ – There is a clinically significant difference in student performance between males and females on the PASS ELA. Data from the National Assessment of

Educational Progress (NAEP) indicated female students scored higher on average nationally than their male counterparts in reading (Lee, Grigg, & Donahue, 2007; National Center for Education Statistics [NCES], 2009a; Perie, Grigg, & Donahue, 2005). Within South Carolina, results from the 2009 PASS ELA indicated that a larger percentage of females than males scored *Met* or *Exemplary* in ELA (SCDOE, n.d.c).

Hypothesis $H_{1.2.}$ – There is a clinically significant difference in student performance between American Indian or Alaska Native, Asian or Other Pacific Islander, Black, Hispanic, and White students on the PASS ELA. Data from the NAEP indicated that White and Asian/Pacific Islander students scored higher on average nationally in reading than their Black, Hispanic, and American Indian/Alaskan peers (Lee et al., 2007; NCES, 2009a; Perie et al., 2005). Within South Carolina, results from the 2009 PASS ELA indicated that a larger percentage of White and Asian/Pacific Islander students scored *Met* or *Exemplary* in ELA in comparison to Black, Hispanic, and American Indian/Alaskan students (SCDOE, n.d.c).

Hypothesis $H_{1.3.}$ – There is a clinically significant difference in student performance between students who receive free/reduced lunch and students who are not eligible for free/reduced lunch on the PASS ELA. Data from the NAEP indicated students who are not eligible for free/reduced lunch scored higher on average nationally in reading than their free/reduced lunch counterparts (Lee et al., 2007; NCES, 2009a; Perie et al., 2005). Within South Carolina, results from the 2009 PASS ELA indicated that a larger percentage of students who are not eligible for free/reduced lunch scored *Met* or *Exemplary* in ELA than students who receive free/reduced lunch (SCDOE, n.d.c).

Hypothesis $H_{1.4.}$ – There is a clinically significant difference in student performance between disabled and nondisabled students on the PASS ELA. According to Bielinski and Ysseldyke (2000), the difference in pass rates between students with disabilities and students without disabilities on statewide reading assessments is approximately 23%. These achievement gaps have proven to be consistent across grade levels. Data from the NAEP indicated nondisabled students scored higher on average nationally in reading than their disabled counterparts (Lee et al., 2007; NCES, 2009a; Perie et al., 2005). Within South Carolina, results from the 2009 PASS ELA indicated that a larger percentage of nondisabled students scored *Met* or *Exemplary* in ELA than disabled students (SCDOE, n.d.c).

Hypothesis $H_{1.5.}$ – There is a clinically significant difference in performance between students in grades 3- 8 on the PASS ELA. Many states set the bar significantly lower in elementary school than in middle school. The differences between third-grade and eighth-grade cut scores in reading are 20 percentile points or greater in South Carolina, New Jersey, and Texas, and there are similar disparities in math in New Jersey, Michigan, Minnesota, North Dakota, and Washington (Cronin, Dahlin, Adkins, & Kingsbury, 2007). Within South Carolina, results from the 2009 PASS ELA indicated that percentage of students scoring *Met* or *Exemplary* in ELA decreased in higher grades (SCDOE, n.d.c).

Hypothesis H_2 – A student's performance on the MAP Reading is able to predict his or her performance on the PASS ELA. In terms of predicting proficiency status, Cronin (2004) found the MAP score estimates for each grade correctly predicted proficiency for 80% to 82% of the cases when using the MAP reading to predict previous South Carolina high-stakes testing in ELA. A slight decrease in prediction was found in terms of predicting specific performance levels with cut score estimates correctly assigning performance levels for 59% to 67% of the cases.

Hypothesis $H_{2.1.2.3.}$ – There is an association between the MAP Understanding and Using Literary Texts Goal Performance, the MAP Understanding and Using Informational Texts Goal Performance, and the MAP Building Vocabulary Goal Performance and the PASS ELA. According to the alignment documentation between the MAP Reading and the South Carolina Academic Standards for ELA, 74% - 85% of the MAP items were aligned to South Carolina grade level standards with all standards being included in the test item pool (SCDOE, n.d.b). The PASS ELA assesses a student's mastery of the 2008 South Carolina Academic Standards in grades 3-8. At each grade level, four broad standards are assessed: Standard 1 – Literary Texts, Standard 2 – Informational Texts, Standard 3 – Vocabulary, and Standard 6 – Research. Depending on the student's grade level, 8-18 items per Standard are included on the PASS ELA (Refer to Table 4).

Hypothesis H_3 – From among the demographic variables and MAP scores identified, it is believed that a student's degree of exposure to CAT and a student's MAP Reading score will be the strongest predictors of student performance on the PASS ELA. Research has shown that students who are more familiar with computers perform better on computer-based tests in math and reading (Choi & Tinkler, 2002). Therefore, as the degree of exposure to CAT increases as a result of the student's advancement through school, the level of comfort with CAT should increase resulting in a more reliable and valid assessment of current instructional levels. Previous research (Cronin, 2004) has shown that MAP Reading scores correctly predicted proficiency on previous South Carolina high-stakes testing in ELA for 80% to 82% of the cases.

Figure 1 depicts the research path diagram of latent variables.

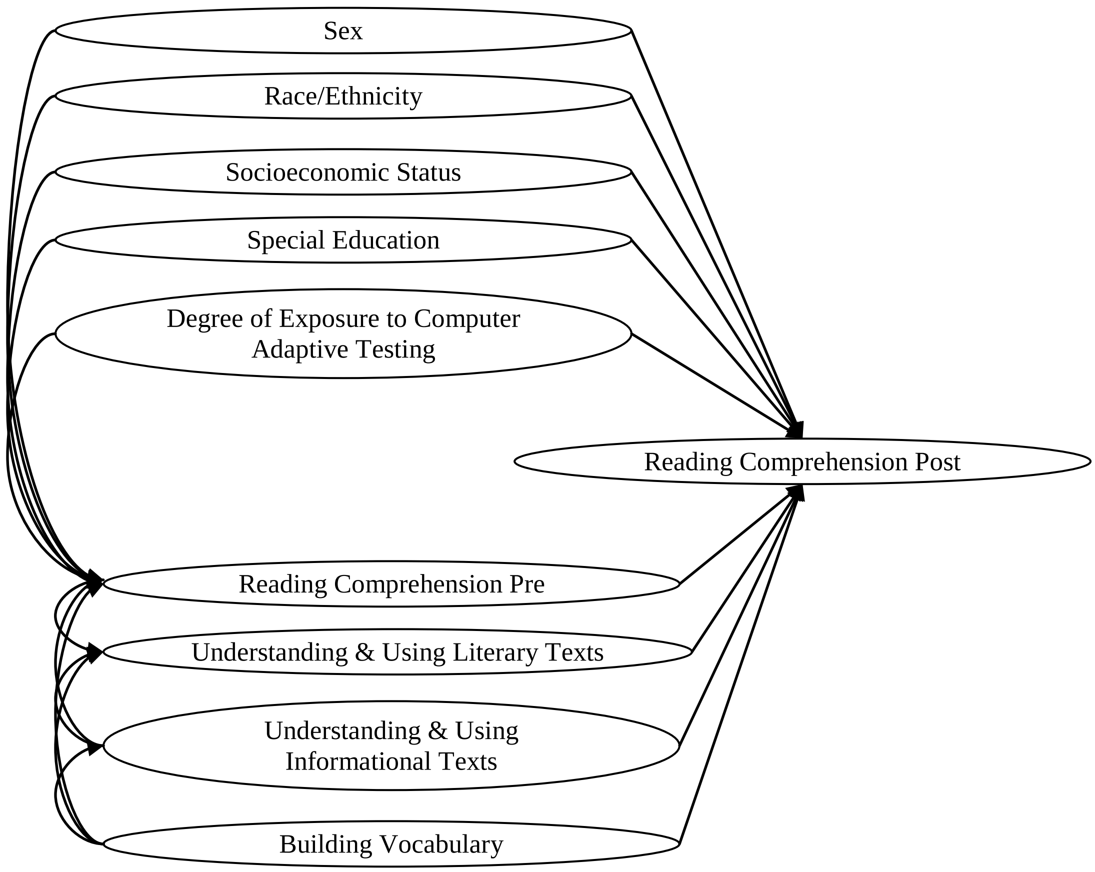

Figure 1. Research path diagram of the latent variables.

Problem Significance

Currently, there is no national assessment; however, many states have similar statewide, high-stakes testing procedures that include testing students in a number of grades and subject areas. The Reading First initiative under NCLB targeted accountability in reading. The Reading First Program is the largest and most focused early reading initiative the United States of America has ever undertaken to ensure that all children in America learn to read well by the end of third grade (U.S. Department of Education - Office of Elementary and Secondary Education, 2002). Proficient readers remain a minority as evidenced by the continuing concerns over illiteracy rates (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993; Marshall, 2009; NCES, 2006; Roman, 2004) and those identified as Learning Disabled (Flugum & Reschley, 1994; Meredith, Steele, & Dawson, 1997; Scruggs and Mastropieri, 2002; Van Haren, Russ, Chiang, & Fiedler, 2006). Over the past several decades, there have been various techniques and methods adopted in hopes of unlocking the mystery of how a young child best learns to read and write. Despite the various instructional techniques adopted over the last forty years, the percentage of children who have trouble in reading has consistently hovered around twenty percent (Fletcher & Lyon, 1998). As a result, illiteracy is a trend that is growing at an alarming rate throughout the United States with approximately 2.3 million individuals added to the illiterate population annually (Roman, 2004). For this reason, reading is one of the target subject areas accountability legislation hopes to improve through the monitoring of high-stakes testing.

Due to the sanctions that may be imposed on school districts if lack of progress is evident in reading based on test scores, many school districts across the nation have begun to use Curriculum-Based Measurement-Reading (CBM-R) to detect those students not on track to be proficient readers from an early age. CBM-R, which uses the number of words read correctly, enables educators to monitor the effectiveness of interventions designed to address the needs of these students (Howe, Scierka, Gibbons, & Silberglitt, 2003). It has been accepted as an empirically valid and reliable index of reading (Foegen, Jiban, & Deno, 2007; Good & Jefferson, 1998; Hintze, Owen, Shapiro, & Daly, 2000; McMaster & Espin, 2007; Wayman, Wallace, Wiley, Tichá, & Espin, 2007). Not surprisingly, with the increased attention to accountability and high-stakes assessment, the relationship between CBM-R and state-mandated testing programs has been a recent topic of interest. In particular, the criterion and predictive validity of CBM-R has been examined as the basis for making judgment about whether students will achieve mandated levels of performance on such high-stakes tests. Research has indicated that there is a moderate-to-strong relation between performance on measures of oral reading fluency (ORF) and state achievement tests within the same year (Barger, 2003; Buck & Torgeson, 2003; Good, Simmons, & Kame'enui, 2001; Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Roehrig, Petscher, Nettles, Hudson, & Torgeson, 2008; Shapiro, Keller, Edwards, Lutz, & Hintze, 2006; Shaw & Shaw, 2002; Stage & Jacobsen, 2001; Vander Meer, Lentz, & Stollar, 2005). Based on this, ORF could assist school districts substantially in targeting students at risk for not passing the state test and provide them the opportunity for additional interventions in reading.

A pitfall in the use of CBM-R is that empirical findings of CBM-R research across grades have demonstrated a deceleration in growth of reading fluency skills as grade levels increased. Silberglitt, Burns, Madyun, & Lail (2006) conducted a study to analyze the relationship between CBM-R and state accountability tests to examine if the strength of this relationship changed as a function of grade. Data were consistent with previous research showing a relationship between CBM-R and test scores (Barger, 2003; Buck & Torgeson, 2003; Good et al., 2001; Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shaw & Shaw, 2002; Stage & Jacobsen, 2001; Vander Meer et al., 2005). The magnitude of the relationship, however, between CBM-R and the state accountability test, declined significantly with advancing grade levels. Correlations dropped from strong to only moderate at increasing grade levels. Similar results were found by Keller-Margulis, Shapiro, and Hintze (2008) who assessed the long-term diagnostic accuracy of CBM-R and found that as grade-level increased, oral reading fluency decreased in relation to performance on state achievement tests. This diminishing relationship could have significant implications for educators. They may assume the strong relationship between CBM-R and state tests in the early grades remains as strong in the later grades. Recent research (Keller-Margulis et al., 2008; Silberglitt et al., 2006) suggests the possibility that such an extrapolation is erroneous, and CBM-R may not have as great a value for predicting statewide achievement test scores in later grades.

Alternatives to CBM-R that are reported to have greater face validity among educators include oral and written retell of stories, cloze passages, and Maze tasks (Fuchs, Fuchs, & Maxwell, 1988). Maze tasks involve students determining which words best fit in paragraphs that have the first sentence complete and every "X" word thereafter having to be filled in by the student. Three choices are provided for each missing word. Earlier research (Fuchs & Fuchs, 1992) revealed that the Maze task was sensitive to change in performance over time. In addition, teachers rated their satisfaction with Maze tasks highly, reporting that they believed the Maze tasks reflected multiple dimensions of reading, including decoding, comprehension, and fluency. In a direct comparison of the technical adequacy of CBM-R and Maze tasks selection measures, Jenkins and Jewell (1993) examined the validity of the two measures across Grades 2-6. Within-grade correlations were moderate to strong for both measures, ranging from .58 to .88. In Grades 2-4, correlations tended to be stronger for CBM-R than for Maze tasks, but in Grades 5 and 6, this pattern of differences disappeared. Looking across grades, correlations between the CBM-R and the criterion measures dropped from the .80s in Grades 2-4 to .60s to .70s in Grades 5 and 6. In contrast, correlations for the Maze tasks remained consistent across the grade levels, with most between .65 and .75. Finally, both measures revealed increases across Grades 2-6, and from fall to spring within grade. For the CBM-R, change was greatest from Grades 2-3, after which it leveled off. Maze tasks, in contrast, reflected more even rates of change across the grades. Thus, if progress is to be monitored across school years, Maze tasks might prove to be a better choice. It has been shown to have reasonable validity and reliability for students across Grades 2-8, and the growth rates across grades have shown greater consistency than those for CBM-R (Wayman et al., 2007).

While more research needs to be conducted to determine the reasons for the age-related differences seen between CBM-R and Maze tasks, perhaps the teachers from the Fuchs and Fuchs (1992) study were correct: Perhaps Maze tasks reflect multiple aspects of reading proficiency to a greater extent than CBM-R does. Furthermore, Wayman et al. (2007) emphasized additional advantages offered by Maze tasks in terms of group administration and appropriateness for computerized administration verses CBM-R which pose logistical issues in assessing large groups of students because it requires individual administration.

In the K-12 school system, educational stakeholders are exploring more efficient measurement tools in place of traditional paper-and-pencil tests (PPTs). In 2006–2007, 23 states were reported to offer computer-based assessments to measure achievement in U.S. schools (Bausell & Klemick, 2007). By 2008, almost all states were expected to have some form of online testing in place (Williams, n.d.). Many educational stakeholders foresee the promise of using computer-based testing (CBT) in their state due to the advantages of CBTs over traditional PPTs in terms of immediate scoring and reporting of students' test results, greater test security, test administration efficiency, flexible test administration schedules, reduced costs compared to handling PPTs, the use of multimedia innovative item types that are not feasible in the PPT format, audio and large-print accommodations for visually impaired students, and the ability to measure response time (Bennett, 2001, 2002; Chaney & Gilman, 2005; Fichten, Asuncion, Barile, Ferraro, & Wolforth, 2009; Folk & Smith, 1998; Klein & Hamilton, 1999; Parshall, Spray, Kalohn, & Davey, 2002).

To meet the demands of NCLB (2001), the need for reliable and valid progress monitoring tools that reflect the multiple dimensions of reading (i.e., decoding, comprehension, and fluency) and allows for assessment of large group of students has become imperative. According to Kingsbury and Hauser (2004), CAT, which has been used very successfully in the military (Sands, Waters, & McBride, 1997) and in professional certification and licensure (Zara, 1992) may be the answer. In this paradigm, the test adapts to match the difficulty of the questions administered to the performance of each student as the student takes the test. The advantages of this adaptive-testing paradigm include increased testing efficiency and tests that are challenging but not frustrating for each student (Weiss, 1982). Although there is substantial evidence of CAT validity, a relationship to state assessments may provide school districts the practical incentive to adopt this sound, research-based practice. If CAT was sensitive in this regard, it could be used to monitor progress toward, and predict future performance on, the state assessment and to assist in establishing appropriate benchmarks. Once these benchmarks are established at each grade level, ongoing progress monitoring for struggling students may assist teachers in adjusting instruction as needed to prepare students for the eventual state test.

Definition of Terms

Achievement Gap: Differences in academic performance among groups of students who are identified by race/ethnicity and income level.

Adequate Yearly Progress (AYP): AYP is the minimum level of performance that school districts and schools must achieve each year as determined under the federal No Child Left Behind (NCLB) act.

Computer Adaptive Testing (CAT): A sequential computer-based test in which successive items are selected from a pool of items based on previous items to which the student has responded. Based in item response theory (IRT), this type of testing is intended to select items that are of appropriate difficulty for the test-taker.

Curriculum-Based Measurement (CBM): CBM is direct observation and recording of a student's performance in the local curriculum as a basis for gathering information to make instructional decisions (Hintze, Owen, Shapiro, & Daly, 2000).

Goal Performance Areas: Goal Performance areas are goal reporting categories on the MAP reports that are aligned to the content of the individual state standards or benchmarks. Goal performance is reported in ranges rather than specific scores. A goal performance of LO means that the student is performing at the 33^{rd} percentile or lower. AV means that the student is performing between the 33^{rd} and 66^{th} percentile while HI means that the student is performing at or above the 66^{th} percentile (NWEA, 2003).

Maze: Maze reading is a multiple-choice cloze task that students complete while reading silently. The first sentence of a word passage is left intact. Thereafter, every 7^{th} word is replaced with three words inside parenthesis. One of the words is the exact one from the original passage.

Measures of Academic Progress (MAP): The MAP are state-aligned computerized adaptive tests that assess the instructional level of each student in the areas of reading, math, language, and science (NWEA, 2003).

No Child Left Behind (NCLB): NCLB is the most recent authorization of the federal Elementary and Secondary Education Act.

Palmetto Assessment of State Standards (PASS): The PASS is South Carolina's standards-based, criterion-referenced test administered to students in grades 3-8. PASS measures students' mastery of grade-level curriculum standards in writing, English/language arts (reading and research), mathematics, science, and social studies. For each PASS test, three overall performance levels are reported: *Exemplary* – The student demonstrated exemplary performance in meeting the grade level standard; *Met* – The student met the grade level standard; *Not Met* – The student did not meet the grade level standard.

RIT Scale: The RIT Scale is a scale that uses individual item difficulty values to estimate student achievement on the MAP. The RIT scale is an equal interval scale meaning that the difference between scores is the same regardless of whether a student is at the top, bottom, or middle of the RIT scale, and it has the same meaning regardless of grade level.

Assumptions

This study was designed to determine the link between ongoing, formative evaluation with high-stakes testing. Specifically, the relationship between the MAP Reading and the PASS ELA was examined. It is assumed that the MAP Reading test questions are aligned to the South Carolina ELA academic standards as the test manufacturer reports. Therefore, the MAP should be a reliable indicator of student progress in achieving these standards, as measured by the PASS. In addition, since the impact certain factors (i.e., sex, race/ethnicity, SES, special education, and degree of exposure to CAT) have on the correlation between MAP and PASS is being analyzed, it is assumed that students meet federal and state criteria for the free/reduced lunch program and/or an Individualized Education Program (IEP).

Limitations

Threats to Internal Validity

Threats to the internal validity could arise from testing effects. Student performance can fluctuate on a daily basis due to various physical and emotional states. In addition, while research indicates minimal differences, performances on the two tests may vary based on the mode by which it was administered (i.e., computer vs. paper-and-pencil). Predictive validity may also increase with students who are more familiar with CAT.

Threats to External Validity

Threats to external validity include the limited and homogenous sample size used. All participants were from the same school district located in a suburban South Carolina community. Therefore, these results may not generalize to a broader population. In addition, although this study investigated the correlation between a CAT and a traditional PPT, it is limited to the testing of reading. Further research is needed to research the comparability of CATs in other core subjects, such as math, writing, and science.

Summary

The effectiveness of public education in the United States has long been scrutinized in our society. As a result, the ability to measure school effectiveness has become a major issue in educational policy and legislation. The most recent attempt at such legislation, the No Child Left Behind (NCLB, 2001) act, includes accountability provisions that mandated states to test all students. Scores from these tests are used for a variety of purposes, from identifying whether individual students are proficient with respect to designated academic standards, to helping determine whether schools are achieving AYP. In an atmosphere of educational reform and federally mandated high-stakes testing, demands have increased for progress-monitoring strategies that reliably predict outcomes on statewide assessments (McGlinchey & Hixson, 2004).

Each year, students in South Carolina participate in testing as part of South Carolina's accountability program. In previous years, statewide testing accountability requirements were met in the state through the administration of the PACT. In June 2008, the South Carolina legislature amended the Education Accountability Act of 1998 (Act 282) to incorporate a new assessment in grades 3-8 known as the PASS. In addition to the PASS, the use of formative assessments such as the MAP is increasing in schools throughout South Carolina. In order to use the MAP as a reliable progress-monitoring tool, an alignment of the scores from the state and the MAP tests is essential. Regular state alignment studies have been conducted by the manufacturers of the MAP to examine the correspondence between the MAP and state standardized tests used to measure student achievement. Four prior studies investigating the alignment of cut scores between the spring administrations of the PACT and the MAP tests have suggested that the MAP tests are reliable and valid predictors of student performance on the PACT (Hauser, 2002; Cronin & Hauser, 2003; Cronin, 2004; Cronin, 2007).

This study investigated the predictive validity of demographic variables and the MAP in relation to student performance on the South Carolina's PASS ELA test. The state test was selected because it is the measure connected to high-stakes for school districts and it is viewed as a comprehensive measure of reading skill. Establishing further support for this relationship could improve teachers' acceptability of the MAP for the purpose of clear goal setting with regards to student progress. The relationships between five latent variables (i.e., sex, race/ethnicity, SES, special education, and degree of exposure to CAT) and student performance on the PASS ELA were examined as well as the validity of the MAP Reading to predict performance on the PASS ELA. Finally, from among the demographic variables and MAP scores analyzed, the aim was to identify the best predictors of student performance on the PASS ELA.

CHAPTER II

LITERATURE REVIEW

The effectiveness of public education in the United States has long been scrutinized. Recent reform efforts, such as No Child Left Behind (NCLB, 2001), have identified statewide accountability testing as a strategy for answering school-effectiveness questions. Although there currently is no national assessment, many states have similar statewide, high-stakes testing procedures. Poor performance on such tests can result in school sanctions, state takeovers, state interventions, or reconstitution of schools. With the threat of such sanctions, a national premium has been placed on the importance of early identification of and intervention programs for academic skill weaknesses. It has become essential for schools to monitor effectively the progress of all students. According to Kingsbury and Hauser (2004), computer adaptive testing (CAT), may be the answer. This literature review will outline reading trends in America. Accountability legislation and the resulting development of summative assessments, otherwise known as high-stakes testing, as it relates to various subgroups and the achievement gap in reading, will also be discussed. A brief overview of South Carolina's high-stakes test, the Palmetto Assessment of State Standards (PASS), will be provided. In addition, specific formative assessment tools in reading will be reviewed, highlighting differences in test modes and individual characteristics on CAT. Finally, information about the use of a specific computer adaptive test, the Measure of Academic Progress (MAP), in predicting student performance on high-stakes testing will be presented. Figure 2 illustrates the flow of the literature review.

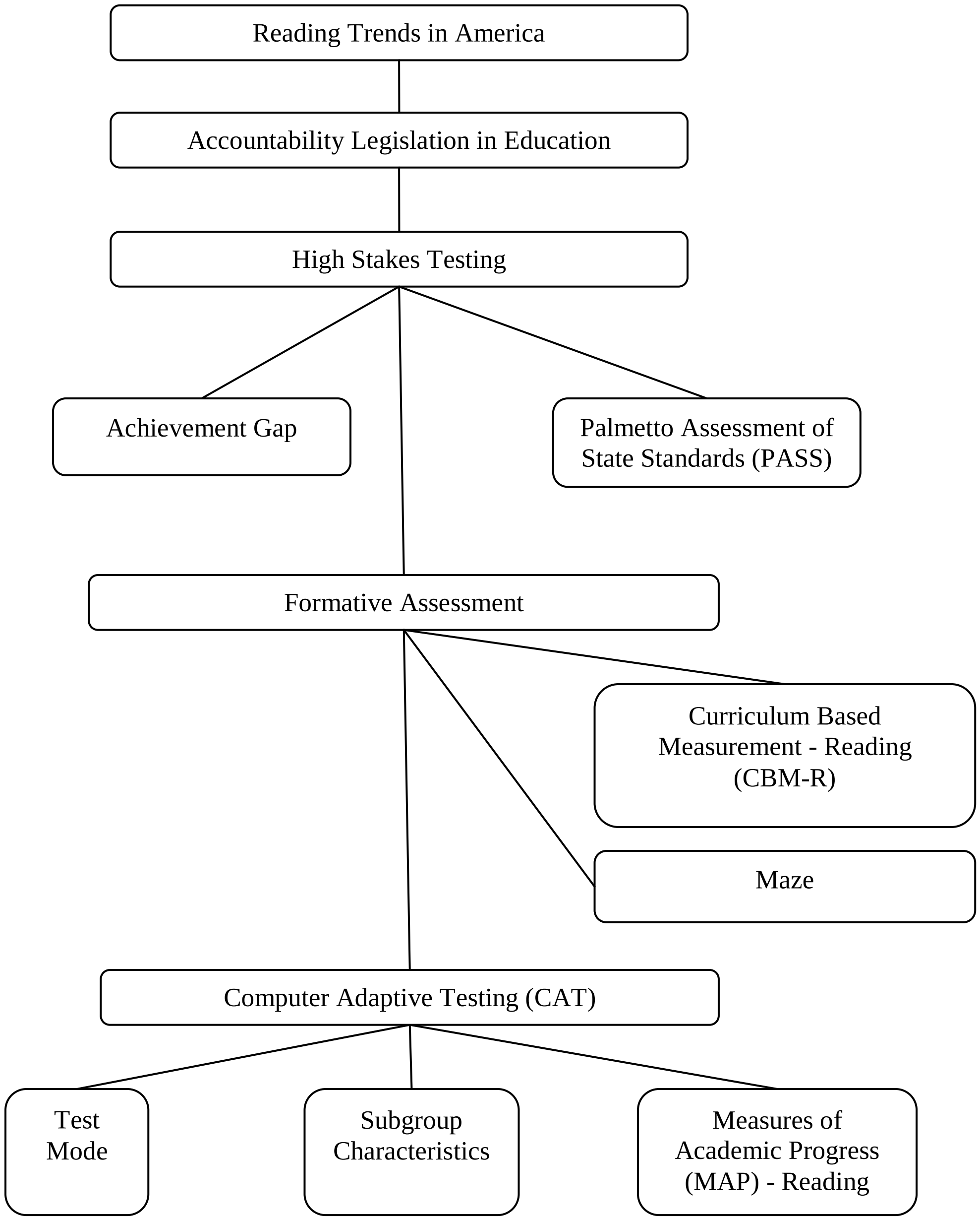

Figure 2. Literature review flow chart.

Reading Trends in America

Recent federal legislation within the United States has had an unprecedented effect on education, specifically on reading instruction. The increasing demands of the twenty-first century workplace have raised the literacy bar for students and subsequently have forced schools to review the effectiveness of their current reading instructional practices. Specific legislation, such as NCLB (2001), emphasizes the systematic and explicit teaching of reading in a manner that ensures student success. Research suggests enough is known about reading instruction to ensure at least 95 percent of the children in the United States can be taught to read at a level of proficiency to enjoy and engage in independent, age-appropriate reading activities (Moats, 1999).

Despite what is known about reading instruction, there has been little evidence of improved literacy outcomes in America during the last thirty years (Perie, Moran, & Lutkus, 2005). Academic at-risk youth with low literacy skills live in urban, suburban and rural communities and represent all social and racial/ethnic groups. However, certain student populations have significantly higher percentages of students reading below grade level. These student populations are becoming increasingly more apparent with the accountability mandates of NCLB (2001) which require detailed reporting of school achievement levels broken down by sex, race/ethnicity, socioeconomic status (SES), English Language Learners (ELL), and special education. Recent findings from the National Assessment of Educational Progress (NAEP) highlight some of the trends at the elementary level (National Center for Education Statistics [NCES], 2009a).

- Thirty-six percent of Male fourth graders read *below basic* level. Only 36 percent are reading at or above proficient level compared to 45 percent of Female fourth graders reading at or above proficient level.
- More than half of all Black and Hispanic fourth graders read *below basic* level. Only 16-17 percent are reading at or above proficient level compared to 42 percent of White fourth graders reading at or above proficient level.
- Forty-nine percent of fourth graders eligible for free/reduced lunch read *below basic* level. Only 19 percent are reading at or above proficient level compared to 57 percent of noneligible fourth graders reading at or above proficient level.

- Seventy-one percent of ELL fourth graders read *below basic* level. Only 6 percent are reading at or above proficient level compared to 44 percent of non ELL fourth graders reading at or above proficient level.
- Sixty-five percent of disabled fourth graders read *below basic* level. Only 15 percent are reading at or above proficient level compared to 43 percent of nondisabled fourth graders.

While some gains are evident based on the NAEP results, these trends continue into the secondary level (NCES, 2009a):

- Twenty-nine percent of Male eighth graders read *below basic* level. Only 30 percent are reading at or above proficient level compared to 41 percent of Female eighth graders reading at or above proficient level.
- Between 39-43 percent of all Black and Hispanic eighth graders read *below basic* level. Only 14-18 percent are reading at or above proficient level compared to 45 percent of White eighth graders reading at or above proficient level.
- Forty percent of eighth graders eligible for free/reduced lunch read *below basic* level. Only 17 percent are reading at or above proficient level compared to 46 percent of noneligible eighth graders reading at or above proficient level.
- Seventy-four percent of ELL eighth graders read *below basic* level. Only 3 percent are reading at or above proficient level compared to 37 percent of non ELL eighth graders reading at or above proficient level.
- Sixty-two percent of disabled eighth graders read *below basic* level. Only 8 percent are reading at or above proficient level compared to 38 percent of nondisabled fourth graders.

In increasingly challenging secondary curricula, reading well is a prerequisite for academic success (Kamil, 2003; Biancarosa & Snow, 2004). For the more than eight million struggling readers in Grades 4-12, “chances for academic success are dismal because they are unable to read and comprehend the material in their textbooks" (Kamil, p. 1). While being a capable reader will not ensure success, the inability to read at or above grade level seriously limits a student's potential in all academic areas. According to American College Testing (ACT, 2006), advanced literacy across content areas is the best available predictor of a student's ability to succeed in introductory college courses. At the nation's four-year colleges, nearly eight percent of all entering students are required to take at least one remedial reading course. From these students, only about one third of them are likely to graduate within eight years (Adelman, 2006). For many adolescent students, ongoing difficulties with reading and writing figure prominently in the decision to drop out of school (Ehren, Lenz, & Deshler, 2004).

Illiteracy is a trend that is growing at an alarming rate throughout the United States with approximately 2.3 million individuals added to the illiterate population annually (Roman, 2004). A recent federal study by the National Assessment of Adult Literacy (NAAL) found that an estimated 30 million adults in the United States possess no more than the most simple and concrete literacy skills (NCES, 2009b). Overall, the study found no significant changes in prose and document literacy between 1992 and 2003 with the only significant increase being in regards to quantitative literacy. Similar to the NAEP results (NCES, 2009a), the NAAL study indicated that several population groups were overrepresented in the *below basic* level (i.e., no more than the most simple and concrete literacy skills). Fifty-five percent of adults with *below basic* prose literacy did not graduate from high school, compared to 15 percent of adults in the general population. Forty-four percent with *below basic* prose literacy did not speak English before starting school, compared to 13 percent of adults in the general population. Thirty-nine percent of Hispanic adults and 20 percent of Black adults scored *below basic* for prose literacy, compared to 12 percent of the general population. Finally, 21 percent of adults with multiple disabilities scored *below basic* for prose literacy compared to 9 percent of the general population.

Illiterate individuals generally experience economic, physical, and emotional consequences. Research has shown that there is a positive relationship between literacy levels and employment stability and income. In fact, seventy-five percent of unemployed adults are functionality illiterate (Roman, 2004). This relationship is powerful because there is a definite effect of literacy on one's ability to become a contributing member of society. Unfortunately, individuals who are functionally illiterate may be unable to obtain gainful employment with a competitive wage. In 2000, average earnings ranged from $18,900 for high school dropouts to $25,900 for high school graduates and $45,400 for college graduates (U.S. Census Bureau, 2002). The societal effects of having a group of Americans unable to support themselves financially through gainful employment are extremely significant and potentially detrimental to society as a whole. According to a 2001 report of the Coalition of Juvenile Justice (CJJ), the nation is drained of more than $200 billion in lost earnings and taxes because of America's high dropout rate. Those who do find work are often stuck in minimum wage jobs that pay too little to support a family in today's society. Even more disturbing is the increased likelihood that high school dropouts, who enter society lacking work skills and life skills, will end up in a correction facility. According to CJJ, 82 percent of prison inmates are high school dropouts with more than one third of all juvenile offenders reading below the fourth grade level.

These statistics stand in stark contrast to what is now known about reading instruction and the prevention of reading failure, especially when only a small percentage, approximately three to five percent of the reading population, is severely reading disabled and unable to be helped with intensive intervention (Pogorzelski & Wheldall, 2002). Not surprisingly, foreboding statistics have increased the criticism of public education and lead to numerous school reform agendas resulting in increased accountability for educators (Popham, 2004).

Accountability Legislation

Accountability reform has been a part of the American education experience for more than a century. By the 1960s, there were significant changes in how educational accountability would be measured and who had control over educational policy-making (Airasian & Abrams, 2002; Heinecke, Moon, & Curry-Corcoran, 2003). The Civil Rights Act in 1964 marked the beginning of the War on Poverty (Hoffman, Assaf, & Paris, 2001). All aspects of society, including educators, were thinking about racial injustice, poverty, and educational inequality. The public had great faith in the American education system, expecting it to bring about social reform and prepare children for their future (Airasian & Abrams, 2002).

Coleman Report

Inequality in America's schools was evident. The questions were how to provide equality of education to everyone and how to measure the progress the nation was making towards this goal. Researchers attempted to answer these questions with the Equality of Education Survey, commonly referred to as the Coleman Report (Coleman et al., 1966). The goals of the study were to determine: (a) the extent of racial and cultural group segregation; (b) the extent to which education facilities and resources differed by racial and cultural groups; (c) the extent to which pupils were learning in school as measured by standardized tests; and (d) the resources and facilities in their schools (Airasian & Abrams, 2002; Heinecke et al., 2003; Hoffman et al., 2001).

The results of the first three questions of the Coleman Report documented that there were still racial and cultural inequalities in education. The researchers had expected that the results of the fourth question would be that the more resources a school had, the more the students learned. This was not what they found. The researchers concluded that the amount of resources a school had did little to overcome the inequalities in a student's background and social situation (Airasian & Abrams, 2002; Hoffman et al., 2001). As a result of the Coleman Report, the focus shifted from the amount of resources a school had to what a school was producing in terms of scores from standardized tests, specifically Minimum Competency Testing (Heinecke et al., 2003).

A Nation at Risk

The hope that Minimum Competency Testing was the answer to educational inadequacies was dashed when in 1983 the National Commission on Excellence in Education (NCEE) released the publication, *A Nation at Risk.* The commission was asked to: (a) assess the quality of teaching and learning in American schools; (b) compare American schools to those in other developed countries; (c) determine the degree to which social and educational changes in the last 25 years had affected student achievement; and (d) identify difficulties that must be overcome to achieve excellence in education (NCEE, 1983). The results were grim with the NCEE asserting that the American educational system was in jeopardy. It stated:

> Each generation of Americans has outstripped its parents in education, in literacy, and in economic attainment. For the first time in the history of our country, the education skills of one generation, will not surpass, will not equal, will not even approach those of their parents (Copperman, as cited in NCCE, 1983, p.4).

The report went on to state that approximately 23 million adults and 13% of 17 year-olds were functionally illiterate. Average achievement was lower than it had been 26 years earlier and SAT scores had dropped continually for 15 years. In addition, the commission found that secondary school curriculum was watered down, as were the textbooks. The overall finding of a *Nation at Risk* was that minimum competency had become the maximum and that America's schools were settling for mediocrity. The commission concluded that the decline would affect the United States' ability to compete in the world economy (Airasian & Abrams, 2002; Duke & Reck, 2003; Heinecke et al., 2003, NCEE, 1983; Valencia & Wixson, 1999).

A *Nation at Risk* is frequently identified as the impetus toward accountability and high-stakes testing over the past 16 years (Barksdale-Ladd & Thomas, 2000). The message of this report was that public schools were failing to address and meet ambitious academic standards and that such a failure threatened American economic competitiveness (Smith & Fey, 2000). In response, states began to set higher standards. Once standards were clearly identified and students were taught the material allowing them to meet the standards, testing seemed to be the most logical approach to identify students who did not meet expectations, as well as the teachers of these students (Barksdale-Ladd & Thomas, 2000). However, early reform efforts resulted in very little change (Duke & Reck, 2003; Valencia & Wixson, 1999).

Goals 2000: Educate America Act

During the 1980s, policymakers became impatient for educational outcomes to improve. They wanted to shift the focus even more heavily to accountability through assessments. In 1989, then President George H. W. Bush met with governors at an educational summit and came up with broad educational goals to be reached by 2000. In addition, the governors also promised to restructure their state's educational systems and develop an assessment system to measure student performance and the effectiveness of school reform efforts (Heinecke et al., 2003; Valencia & Wixson, 1999). The culmination of this movement came in 1994 when then President Clinton signed the Goals 2000: Educate America Act. The goals of this act were: (a) all children will start school ready to learn; (b) the high school graduation rate will increase to 90% or better; (c) all students will become competent in challenging subject matter; (d) teachers will have the knowledge and skills that they need; (e) U.S. students will be the first in the world in mathematics and science achievement; (f) every adult will be literate; (g) schools will be safe, disciplined and free of guns, drugs, and alcohol; and (h) schools will promote parental involvement and participation (Heinecke et al., 2003).

In addition to Goals 2000, under President Clinton's administration, the Elementary and Secondary Education Act of 1965 (ESEA) was reauthorized as the Improving America's Schools Act of 1994. This act was the first federally mandated school accountability plan requiring academic assessments. However, the law only applied to Title I students and there was no threat of federally mandated penalties for failure to meet the requirements of the law. In addition, many states indicated that they could not meet the requirements for these assessments and numerous waivers were granted (Peterson & West, 2003). Regardless, President Clinton's proposal to create a voluntary national test in reading and math catapulted testing to the top of the national educational agenda (Wise, Hauser, Mitchell, & Feuer, 1999).

No Child Left Behind (NCLB)

Improving America's Schools Act was the basis for NCLB, which was signed into law in January of 2002 by then President George W. Bush. The NCLB act expanded the accountability requirements to all Title I schools, as opposed to only Title I students (Peterson & West, 2003). The signing of the NCLB act was a seminal moment in educational reform, because it represented an unprecedented expansion of the role of the federal government in education (Popham, 2004). Prior to NCLB, schools typically reported on a few measures related to resource appropriation, but they did not report any indication of student achievement (Peterson & West, 2003). Legislators intended NCLB to enhance the state's assessment policy, not replace it (Lee, 2006). However, since so many states were granted waivers from the Improving America's Schools Act, most states did not have a comprehensive accountability program in place.

The accountability measures of NCLB require schools to ensure all students are proficient on state reading and math standards by 2014 (NCLB, 2001). Each state is expected to implement high academic standards along with reliable and accurate assessments. Schools are required to show more than just the average student attaining grade-level proficiency to prevent masking students who are not meeting expectations. Under NCLB, proficiency is reported by grade levels and subgroups. The subgroups include students from different racial and ethnic backgrounds, students who are economically disadvantaged, students with limited English proficiency, and students with disabilities (Yell, Katsiyannas, & Shiner, 2006). These subgroups were designated because research has indicated many of these students are falling behind their peers. Each grade level and subgroup must be proficient in order for the school to meet requirements for adequate yearly progress (AYP). AYP is a predetermined goal that sets the percent of students that must be proficient in a given year. The percentage required to attain AYP increases incrementally each year until 2014 or whenever each subgroup reaches 100% proficiency. Additionally to achieve AYP, 95% of all students continuously enrolled must participate in the NCLB assessments.

High-Stakes Testing

The high-stakes testing requirements of the NCLB legislation are clearly and unmistakably having a significant impact on public schools across the nation. Public schools are under significant pressure to improve student achievement on these state-controlled assessments because a school's determination of success is heavily dependent on how well the students perform on high-stakes tests. With these assessments come either rewards or public criticism, hence the term, "high-stakes". Sanctions, rewards, and accountability systems vary a great deal and are chosen by states or districts (Linn, 2001). Examples of current "high-stakes" include graduation, grade promotion or retention for students, personnel evaluations, performance-based pay and continued employment for teachers (Braden, 2002). According to the Advancement Project (2010), from 2001 until 2008, the number of states that used test results to sanction schools rose from 14 to 32. Furthermore, sanctions have become increasingly more severe. For example, from 2003 to 2008, the number of states that sanctioned schools by turning them over to private management increased from 6 to 16. During the same period, the number of states that threatened conversion of low-scoring schools into charter schools rose from 4 to 15. Finally, from 2002 to 2008, the number of states whose sanctions included the option of "reconstituting" schools (such as by firing everyone on staff) increased from 15 to 29.

High-stakes accountability testing has yielded both intended and unintended sets of outcomes. The most obvious, major intention is to improve education. This is meant to be accomplished through the alignment of curriculum with standards and tests, increasing the efficiency of education through better resource allocation, increasing student and staff motivation, and providing for educational equity. Despite the benefits of educational testing promoted by policymakers, there are multiple documented drawbacks. Studies discussing how educational testing reduces time for instruction, narrows the curriculum to only test items, and limits divergent and higher-order thinking have been published (Center on Education Policy [CEP], 2009; Nichols & Berliner, 2007 ; Pedulla, Abrams, Madaus, Russell, Ramos, & Miao, 2003;

Supovitz, 2009). In fact, the impact of testing on the classroom has become so great that some teachers report spending a quarter of their time, or even more than half their time, preparing for and administering standardized tests (Nichols & Berliner, 2007). In addition, increases in dropout rates have been attributed to high-stakes tests (Reardon, Atteberry, Arshan, & Kurlaender, 2009). One reason for this is the number of states requiring the passage of high-stakes exit exams for high school graduation has risen dramatically over the last 20 years. In 1981, only one state was using a high school exit exam (Warren, Jenkins, & Kulick, 2006). By 2008, that number increased significantly to 23 states, with another three states planning to implement exit exams by 2012 (Marchant & Paulson, 2005; Warren et al., 2006; Zabala, Minnici, McMurrer, & Briggs, 2008). Others are concerned about the negative impact on students who may be retained in a grade because of poor test performance even though grade retention has been shown to be the single largest predictor of student dropout (Edley & Wald, 2002; Nichols & Berliner, 2007).

Achievement Gap

The "achievement gap" in education refers to the disparity in academic performance between groups of students. The term is most often used to describe the troubling performance gaps between many Black and Hispanic students, at the lower end of the performance scale, and their White peers, at the higher end of the performance scale. A similar academic disparity is evident between students from low-income and high-income families. The achievement gap reveals itself in grades, standardized-test scores, course selection, dropout rates, and college-completion rates and, as a result, has become a focal point of educational reform efforts. The racial gap itself has been a major impetus for federal education policy as embodied in NCLB and has entered into countless state and local debates regarding school finance equalization, academic tracking, and school testing and accountability programs.

Racial minority subgroup and high-stakes testing. No topic goes to the heart of American concerns about equity in K-12 education more than the racial achievement gap. Since the publishing of the Coleman Report (Coleman et al., 1966), researchers have known that average achievement test scores of Black students lag well behind those of White students, but this issue has been taken up with renewed energy in the last decade. Starting in the 1990s, researchers used nationally representative samples to document the extent of and change in racial and ethnic gaps in achievement test scores. Several studies examined data from the NAEP, which is a nationally representative assessment of what U.S. students know and can do in various subject areas. One of the major indicators of the NAEP focuses on the results of fourth and eighth grade students in the subject of reading. In summarizing the NAEP over the past several decades, Phillips and Chin (2004) found that the Black-White gap narrowed during the 1970s and into the 1980s, after which it stagnated or grew again slightly. In 2000, this gap was 0.83 standard deviations in 4th grade reading with a slight increase to 0.85 standard deviations for 8th grade. Over the past decade, this gap has narrowed somewhat, but still persists. In 2009, the Black-White gap in reading was 26 points at the fourth grade level, which was not significantly different from the recorded gap in 2007 (i.e., 27 points). At the eighth grade level, significant score gaps persisted with the 26 point gap not being significantly different from its corresponding gaps in 1992 or 2007 (NCES, 2009a).

Hispanic students nationwide now comprise a larger minority group than Black students. Although the historical circumstances and policy issues may differ between these two groups, the issue related to measuring the test score gap with Whites is similar. As a general matter, the size of the Hispanic-White gap tends to be relatively smaller than the Black-White one. In their NAEP analysis, for example, Phillips and Chin (2004) found gaps to the order of 0.70 standard deviations in reading at the fourth grade level and 0.80 standard deviations at the eighth grade level. Although smaller than the Black-White gap, the Hispanic-White gap has been seen as a constant in the past decade. In 1992, the Hispanic-White gap in reading was 27 points at the fourth grade level, and 26 points at the eighth grade level (NCES, 2009a). In 2009, the Hispanic-White gap in reading was 25 points at the fourth grade level, which was not significantly different from the recorded gaps in 1992 or 2007. Similar trends were also found at the eighth grade level with the Hispanic-White 24 point gap not being significantly different from its corresponding gaps in 1992 or 2007 (NCES, 2009a).

In 2002, the CEP began an ongoing, formative assessment of NCLB's effect on student outcomes in which they analyzed a vast array of test data from 4th, 8th, and 10th grade students. On the state reading tests used for NCLB accountability, the CEP examined two indicators of achievement: the percent of students scoring at or above the "proficient" level (the main indicator of progress under NCLB) and effect size, indicating the strength of a relationship between two variables. For Blacks, 14 of 38 states had narrowed gaps in reading across all three grade levels, while no state showed evidence that gaps had widened (CEP, 2007). Results were similar for Hispanic and low-income subgroups (CEP, 2007). Furthermore, for the Black and Hispanic subgroups, many more states showed gaps narrowing than widening at all grade levels analyzed in both reading and math (CEP, 2008).

McCall, Hauser, Cronin, Kingsbury, and House (2006) found similar results when looking at high-stakes testing results across the country. Their research indicated that the gap still exists between Whites and their Black and Hispanic peers and is relatively consistent across all grades and subject areas studied. Furthermore, McCall et al. found that in all grades and subject areas, Black students grow less academically during the school year than students in other racial/ethnic groups. This difference was more noticeable in mathematics than in reading. Another notable finding McCall et al. reported was that minority students grow less, or lose more ground, over the summer than peers who start with the same score, particularly among high performers.

Economically disadvantaged subgroup and high-stakes testing. The disparities in achievement are often attributed to SES factors. Since SES characteristics, such as income and parental education, tend to correlate with race, it is likely that at least a portion of the observed gaps between racial and ethnic groups can be accounted for by non-racial factors. The research on achievement gaps has sought to determine just how large this portion is, in part because it is the portion of the currently observed gap that presumably will wither away over time as SES differences recede. According to the U.S. Census Bureau, of all children younger than 18 years of age living in families, 27 percent of Hispanic children and 30 percent of Black children live in poverty, compared with approximately 13 percent of White children (Proctor & Dalaker, 2002). Data from the U.S. Department of Education's Early Childhood Longitudinal Study states that the average cognitive score of pre-kindergarten children in the highest SES bracket was significantly higher than the average score of students in the lowest SES bracket. The composition of these SES brackets was closely tied to race; 34 percent of Black children and 29 percent of Hispanic children were in the lowest SES bracket, compared with just nine percent of White students (Lee and Burkam, 2002). Research has also shown that dropout rates tend to be higher for children who live in poverty. In 2000, young adults living in families with incomes in the lowest 20 percent of all family incomes were six times more likely than their peers from families in the top 20 percent of income distribution to drop out of high school (Kaufman, Alt, & Chapman, 2001).

Similar to the racial/ethnic subgroup, McCall et al. (2006) found that an achievement gap exists between students in low-poverty schools and those in high-poverty schools. The gap was relatively consistent across all grades, indicating that the groups of students in schools with high levels of poverty are no closer to students in low-poverty schools in eighth grade than they were in the third grade. In their study covering grades K-3, Fryer and Levitt (2006) found that by adding a small set of controls (i.e., age, sex, birth weight, mother's age at first birth, and indicators of SES) estimated Black-White gaps were reduced by more than half and actually eliminated the pure racial component at the beginning of kindergarten, after which it grew at the rate of about a tenth of a standard deviation through third grade. Additional support for the impact SES has on achievement comes for the Longitudinal Evaluation of School Change and Performance (LESCP) in Title I Schools (Westat, 2001). Students in the Title I Schools had average reading Stanford Achievement Test-9 (SAT-9) scores of 602 in third grade and 640 in fifth grade, compared with national norms of 614 in third grade and 654 in fifth grade. Also, students who lived in poverty (i.e., were eligible for free/reduced lunch) did significantly worse on reading tests than other students in the sample. They had a third grade reading score 6.1 points below the average for the sample. Interestingly enough, the study found that these students made gains at an average pace, neither closing the gap nor falling behind further in reading.

As previously mentioned, a student's eligibility for free/reduced lunch is often used as an indicator of SES. Students from low-income families are typically eligible while students from higher-income families typically are not. Recent results from the NAEP indicate that students in the fourth and eighth grade who were not eligible for free/reduced lunch scored higher on average than those who were eligible. Students eligible for reduced-price lunch scored higher than those eligible for free lunch (NCES, 2009a). The scores for all three groups in fourth grade showed no significant change from 2007 to 2009. Average scores for eighth grade were higher in 2009 than in 2007 both for students who were eligible for free school lunch and for students who were not eligible. However, the scores in 2009 for students eligible for reduced-price lunch were not significantly different from 2007 (NCES, 2009a).

Researchers have tried to pinpoint why race and class are such strong predictors of educational attainment. In the 1990s, the controversial *The Bell Curve* (Herrnstein & Murray, 1994) claimed that gaps in achievement were the natural result of variation in genetic makeup and natural ability. The book drew severe criticism from various research fields. Many experts highly contested the findings and asserted that achievement gaps were the result of more subtle environmental factors. Growing up in a low-income family, for example, often means having fewer educational resources at home, in addition to poor health care and nutrition-factors that can contribute to lower academic performance. Others point directly to factors within schools, such as peer pressure, student tracking, negative stereotyping, and test bias (Viadero, 2000).

More recently, scholars have analyzed the effect that certain in-school factors have on student achievement. While it is difficult to isolate the variables that directly impact student achievement, research has shown that good teaching matters (Rivkin, Hanushek, & Kain, 2005).

The Education Trust, a Washington-based research and advocacy organization, found that many minority students attend inner-city schools, which are often underfunded. As a result, those students tend to receive poorer-quality instruction, have fewer high-caliber teachers, and have access to fewer resources (Education Trust, 2002).

In principle, the public is behind closing the achievement gap. In a 2003 national opinion poll on Americans' attitudes toward public education, conducted by Phi Delta Kappan and Gallup, 90 percent of those polled believed closing the achievement gap between White, Black, and Hispanic students was somewhat or very important. Although most think the gap is a result of factors unrelated to the quality of schooling. A 2001 poll revealed that more than half of those polled thought it was the responsibility of public schools and educators to close the gap (Rose & Gallup, 2001, 2003).

Students with disabilities subgroup and high-stakes testing. Closing the achievement gap for students with disabilities has emerged as one of the greatest challenges presented by the accountability movement. A report published by the National Center on Educational Outcomes, Interpreting Trends in the Performance of Special Education Students (Bielinski & Ysseldyke, 2000), portrayed the grim performance of students with disabilities. According to Bielinski and Ysseldyke, the difference in pass rates between students with disabilities and students without disabilities on statewide reading assessments is approximately 23%. These achievement gaps have proven to be consistent across grade levels. In 2003, reading assessment results from the NAEP showed that only 57% of fourth graders with disabilities scored at the basic or above level in reading compared to 83% of general education students. Results for eighth grade revealed that students with disabilities performed at a 32% pass rate and students without disabilities at a 73% pass rate (Donahue, Daane, & Jin, 2003). The most recent results from the NAEP indicate that the average reading score for fourth-grade students with disabilities was higher in 2009 than in 1998, but was not significantly different from the average score in 2007. Student with disabilities in the fourth grade scored on average 34 points lower than students without disabilities. The average reading score for eighth-grade students with disabilities in 2009 was higher when compared to 2007, but was not significantly different from the score in 1998. Student with disabilities in the eighth-grade scored on average 37 points lower than students without disabilities (NCES, 2009a).

Sex subgroup and high-stakes testing. Academic achievement differences between males and females have long captured the attention of educational researchers. Educational studies from the 1930s and 1940s found no differences in the reading levels of males and females (Hogrebe, Nist, & Newman, 1985). However, in the 1960s, a trend began that has continued on for decades with research indicating that the reading abilities of girls exceeded those of boys (e.g., Gates, 1961; Hogrebe et al. 1985; Stanchfield 1973; Sheridan 1976). In more recent years, studies continue to support that females outperform their male counterparts in the area of reading (Donahue et al., 2003; Grigg, Daane, Jin, & Campbell, 2003; Mullis, Martin, Gonzalez, & Kennedy, 2003). A 2004 study by the NCES provided an analysis of sex differences in reading achievement for the 1992-2003 administration of the NAEP. This analysis revealed that females in 4th, 8th, and 12th grade consistently performed better than their male counterparts in reading achievement. This trend continued in 2005 and 2007 as female in fourth and eighth grade both scored higher, on average, than their male peers (Lee, Grigg, & Donahue, 2007; Perie, Grigg, & Donahue, 2005). The most recent results from the NAEP indicate that female students scored 7 points higher on average than male students in fourth grade and 9 points higher on average in the eighth grade. Important to note though is the average reading score for male students was higher in 2009 than in both 2007 and 1992, while the score for female students was not significantly different from either year (NCES, 2009a).

Citing data from the NAEP, Kleinfeld (2006) suggests that all male students, not just minorities or those living in poverty, are lagging behind their female peers in reading. Kleinfeld's analysis of the data indicated that at the end of high school, 23% of the White males of college educated parents scored *below basic* in reading achievement, compared to 7% of their female peers. Black male students fared even worse, as 43% of the Black males of college-educated parents performed at the *below basic* level in reading, compared to 33% of their female peers. The same trend is also true for Hispanic students, as 34% of Hispanic males of college-educated parents performed at the *below basic* level, compared to 19% of their female peers.

The sex differences in reading are further evident in the research surrounding those students diagnosed with reading disabilities. In a recent review of sex differences in reading disabilities, Rutter et al. (2004) reported the sex ratios in four independent epidemiological studies indicated that significantly more males than females are diagnosed with reading disabilities. Similar findings were reported by two other studies (i.e., Hawke, Wadsworth, Olson, & DeFries, 2007; Olson, 2002).

Not all experts and researchers agree that there is a crisis related to the reading achievement of boys. A recent report released by the Education Sector (Mead, 2006) presents evidence suggesting that the real issue is that girls are performing better; not that boys are doing worse. Acknowledging that some sex differences are real and that poor Black and Hispanic boys do demonstrate poor achievement, the report identifies several factors contributing to the recent increase in attention to sex differences in achievement. According to Mead's report, one factor is the media visibility surrounding the realization that privileged boys are also at-risk. Others include research conducted on the differences between boys and girls and the emergence of theories purporting to explain why boys of all nationalities are not performing at the same levels as girls.

Subgroups and high-stakes testing. A number of states have excluded certain subgroups from accountability reporting by raising the N-size of the subgroup. Under NCLB, AYP does not need to be met in a case in which the number of students in a subgroup is insufficient to yield statistically reliable information or if the results would reveal personally identifiable information about an individual student. Schools and districts are allowed to exclude test scores from subgroups of students small enough to be statistically unreliable when counted towards the performance of a school. While this number must be sanctioned by the Federal Department of Education, it varies from state to state (Davis, 2006). States are required to "consider only whether a given size effectuates the collection of reliable information and ensures student anonymity" (Stephenson, 2006, p. 171). Thus, states may choose the minimum number of students that a school must have in a subgroup in order to trigger the requirement for reporting to the public and tracking student achievement for the subgroup. The impact of such variation from state to state can be significant. For example, a school in Maryland failed to meet AYP because ten special education students did not test proficient; however, during the same year, a school in Virginia met AYP goal although twenty-four disabled students in the school were not proficient. The difference between AYP achievements in the two states was the N-size of the disabled student subgroup. The minimum number for a subgroup was five students in Maryland and fifty students in Virginia (Stephenson, 2006).

Test scores may also be counted multiple times, based on the number of designated subgroups for which they are counted. A single student may be considered as a part of the disabled students group, the economically disadvantaged group, the female group, and also the group of all students (Riddle, 2004). In a school with a small enrollment of students, "one student with the right demographic features who fails the test could swing a district's score as much as 3.3% (given a size of 30 in each subgroup)" (Jarrell, 2005, p. 68). As a result, the chances of a school or district meeting AYP may be impacted negatively by merely a few test scores. In general, the number of schools meeting AYP increases as the minimum subgroup size increases (Porter, Linn, & Trimble, 2005; Simpson, Gong & Marion, 2006). For this reason, more and more states have raised the minimum subgroup size in order to increase the number of schools meeting AYP. Chudowsky and Chudowsky (2005) found that 23 states have increased, but no states have decreased their minimum subgroup size since 2004. They also noted a trend toward states tending to choose larger minimum subgroup size, create different sizes for different subgroups, and use formulas to calculate subgroup sizes according to the size of the schools.

High-Stakes Testing in South Carolina

In the initial phases of NCLB accountability, the Palmetto Achievement Challenge Test (PACT) was developed to assess student proficiency in South Carolina. The PACT was aligned to the South Carolina academic standards developed for English language arts, mathematics, science, and social studies and was administered annually to students in grades 3-8. Results were reported as total scaled scores and four performance levels were established to reflect the knowledge and skills exhibited by students: *Advanced* **-** The student *exceeded expectations* for student performance based on the curriculum standards; *Proficient* **-** The student has *met expectations* for student performance based on the curriculum standards; *Basic* **-** The student has *met minimum expectations* for student performance based on the curriculum standards; and *Below Basic* - The student has *not met minimum expectations* for student performance based on the curriculum standards (South Carolina Department of Education [SCDOE], n.d.a). Since the implementation of PACT, though, concerns have arisen over the four performance levels established to reflect the knowledge and skills exhibited by students. Under NCLB, only three performance levels are required (i.e., *basic*, *proficien*t, and *advanced*). Initially, the SCDOE created the *Basic* level of performance to represent minimal preparation for the next grade level. NCLB requires all students to score *Proficient* in ELA and Mathematics by 2014. As a result, students in South Carolina were being held to a double standard. School districts were required to report the percentage of students meeting the standard of *Basic* and above on the PACT for state requirements, but the same school districts were also required under federal standards to report the percentage of *Proficient* or above on the PACT. This double standard caused much confusion between state comparisons at the federal level as well as confusion among the public at the state level.

To rectify the aforementioned issue, the Education Accountability Act of 1998 was amended (Act 282) during the 2008 state legislative session. Beginning with the 2008-2009 school year, the Palmetto Assessment of State Standards (PASS) test replaced the PACT and is now used for statewide accountability testing purposes. The test continues to be given on an annual basis with students being tested in writing, ELA (reading and research), mathematics, science, and social studies. For each PASS test, three new overall performance levels are reported: *Exemplary* – The student demonstrated exemplary performance in meeting the grade level standard; *Met* – The student met the grade level standard; *Not Met* – The student did not meet the grade level standard. Scores of *Met* and *Exemplary* will meet the federal requirement of proficiency under NCLB (SCDOE, n.d.a).

A new statewide accountability test was not the only change that was brought about by Act 282. Formative assessments will become mandatory for students in grades 1-9 in English language arts and mathematics beginning with the 2009-2010 school year for all public schools in South Carolina (South Carolina Education Oversight Committee, n.d.a). While there is some choice in what formative assessment is used by individual school districts, the act does require that the assessment must provide opportunities for periodic formative assessment during the school year as well as reports that are useful for informing classroom instruction, strand, or significant groupings of standards level information about individual students.

Formative Assessments

The pressure of NCLB is profound and challenges all schools to raise student achievement as measured by standardized assessments. In this quest of raising student achievement, schools are in desperate need of predictive assessments resulting in preventive actions to accurately gauge student achievement prior to students taking high-stakes assessments. Formative assessments may meet this need as they provide teachers the necessary feedback to modify their practices before the high-stakes test. Research has shown that when formative assessments are used effectively learning outcomes improve. In a meta-analysis of over 250 studies, Black and William (1998) found that formative assessment could be effective in basically all educational settings, across all content areas and grade levels, and across all levels of achievement. Black and William suggest that the use of high quality formative assessment can produce learning gains of ½ - 1 standard deviation on summative assessments. This is equivalent to raising an average student's score at the 50th percentile to the 85th percentile. Other researchers have found similar findings that formative assessments improve student learning, which subsequently is reflected in accountability testing (Fisher, Lapp, Flood, Frey, & Moore, 2006). Thus, the benefits of exploring the predictive qualities of formative assessments on high-stakes assessments are worthwhile.

Curriculum Based Measurement-Reading

Because of the stark and multiplicative nature of reading failure, formative assessments designed to identify children struggling to learn at an early age, and thus remediate the problem as soon as possible, has been the focus of much research. Currently, there exist formative assessment measures that allow educators to observe and quantify early literacy skills in the areas of phonemic awareness, letter-sound correspondences, and oral reading fluency. Curriculum Based Measurement (CBM) is one type of direct, formative assessment that provides efficient and reliable information of such skills (Foegen, Jiban, & Deno, 2007; Good & Jefferson, 1998; Hintze, Owen, Shapiro, & Daly, 2000; McMaster & Espin, 2007; Wayman, Wallace, Wiley, Tichá, & Espin, 2007).

The criterion and predictive validity of CBM-Reading (CBM-R) has been examined as the basis for making judgment about whether students will achieve mandated levels of performance on such high-stakes tests. Various studies have explored the relationship between oral reading fluency and specific statewide assessments. Numerous studies have shown that third grade fluency scores are excellent predictors of performance on a variety of state performance tests (Buck & Torgesen, 2003; Good, Simmons, & Kame-enui, 2001; McGlinchey & Hixson, 2004). Likewise, Stage and Jacobsen (2001) found that performance on CBM-R among fourth graders was found to significantly predict failure and success on the Washington Assessment of Student Learning (WASL) and increased predictive power by 30%. More recently, Silberglitt and Hintze (2005) examined the performance of over 2,000 students who were administered

CBM-R benchmark assessments in the Spring of Grades 1, 2, and 3 and the Minnesota Comprehensive Assessment (MCA) also in the Spring of Grade 3. Results again suggested a significant relationship between CBM-R and the MCA with the relationship being the strongest for those CBM-R assessments that were temporally closer to the administration of the MCA (i.e., Spring of Grade 3). Regardless, CBM-R was able to predict with a high degree of accuracy (greater than 80%) those students who were likely to pass the MCA as far back as the Spring of Grade 1.

As the aforementioned research shows, CBM-R is a reliable and valid indicator of overall reading competence with significant predictive power for students in third and early fourth grade. A pitfall does exist, however, in the use of CBM-R in later grades. Empirical findings of CBM-R research across grades have demonstrated a deceleration in growth of reading fluency skills as grade level increased. Silberglitt, Burns, Madyun, and Lail (2006) conducted a study to analyze the relationship between CBM-R and state accountability tests to examine if the strength of this relationship changed as a function of grade. The magnitude of the relationship between CBM-R and the state accountability test declined significantly with advancing grade levels. Correlations dropped from strong to only moderate at increasing grade levels. Given that reading growth decelerates as children progress through elementary school (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993), it seems reasonable that developmental increases in reading fluency levels would eventually reach an asymptote. The mean CBM-R scores from Silberglitt et al. data increased by almost 30 words per minute from Grades 3-5, and approximately 20 words per minute from Grades 5 to 7, but increased only approximately two words per minute between Grades 7 and 8. Thus, although CBM-R continues to account for substantial amounts of variance in student performance in the later grades, the overall value of this predictor diminishes substantially. Similar results were found by Keller-Margulis, Shapiro, and Hintze (2008) who assessed the long-term diagnostic accuracy of CBM-R and found that as grade-level increased, oral reading fluency decreased in relation to performance on state achievement tests. Based on past research demonstrating the strong relationship between CBM-R and state tests in the early grades, educators may be assuming that this relationship will remain strong in the later grades as well. Recent research (Keller-Margulis et al., 2008; Silberglitt et al., 2006) suggests the possibility that such an extrapolation is erroneous, and CBM-R may not have as great a value for predicting statewide achievement test scores in later grades.

Not surprisingly, the need for better reading outcomes for older students is clear. At the point when decoding is automatized for students and proficient reading can no longer be inferred from reading rate alone, production-based assessment tools are needed that allow educators to observe and quantify indicators of reading comprehension development in later grades. Data from the NAEP indicates the magnitude of such a need. In 2005, more than two-thirds of the eighth graders tested scored below the proficient level, and more than one-quarter were unable to read at even a basic level (Perie et al., 2005). Scores for 12^{th} graders in 2002 showed a significant decrease from the last assessment in 1998, with more than one-quarter of 12^{th} graders also reading *below basic* (Grigg et al., 2003). Moreover, it is estimated that approximately 8.7 million fourth through twelfth graders in America have dismal chances for academic success because they are unable to read and comprehend academic texts (Kamil, 2003).

Not only school performance is affected by poor reading skills. Increasingly high levels of literacy are required for living-wage jobs. Students who leave school without a level of literacy sufficient to enter the skilled workforce or successfully enter higher education are without means to access the economic and social capital necessary for personal autonomy. For older students with serious reading difficulties, the prospects are grim. Students who drop out or leave school with poor reading skills are disproportionately represented in the juvenile justice system, and large numbers of incarcerated juveniles are illiterate or marginally literate (Barton, 2000). Based on the aforementioned statistics, it is imperative to identify additional progress monitoring tools that are effective in identifying older students at risk in reading.

Maze

Alternatives to CBM-R that are reported to have greater face validity among educators include oral and written retell of stories, cloze passages, and Maze tasks (Fuchs, Fuchs, & Maxwell, 1988). From these three, Maze tasks have received the most attention in recent research. Maze tasks involve students determining which words best fit in paragraphs that have the first sentence complete and every "X" word thereafter having to be filled in by the student. Three choices are provided for each missing word. Earlier research (Fuchs & Fuchs, 1992) revealed that the Maze tasks were sensitive to change in performance over time. In addition, teachers rated their satisfaction with Maze tasks highly, reporting that they believed the Maze tasks reflected multiple dimensions of reading, including decoding, comprehension, and fluency. In a direct comparison of the technical adequacy of CBM-R and Maze selection measures, Jenkins and Jewell (1993) examined the validity of the two measures across Grades 2-6. Within-grade correlations were moderate to strong for both measures, ranging from .58 to .88. In Grades 2-4, correlations tended to be stronger for CBM-R than for Maze tasks, but in Grades 5 and 6, this pattern of differences disappeared. Looking across grades, correlations between the CBM-R and the criterion measures dropped from the .80s in Grades 2-4 to .60s to .70s in Grades 5 and 6. In contrast, correlations for the Maze tasks remained consistent across the grade levels, with most between .65 and .75. Thus, if progress is to be monitored across school years, Maze tasks might prove to be a better choice. Maze tasks have been shown to have reasonable validity and reliability for students across Grades 2-8, and the growth rates across grades have shown greater consistency than those for CBM-R (Wayman et al., 2007).

While more research needs to be conducted to determine the reasons for the age-related differences seen between CBM-R and Maze tasks, perhaps the teachers from the Fuchs and Fuchs (1992) study were correct: perhaps Maze tasks reflect multiple aspects of reading proficiency to a greater extent than CBM-R does. Furthermore, Wayman et al. (2007) emphasized additional advantages offered by Maze tasks in terms of group administration and appropriateness for computerized administration verses CBM-R, which pose logistical issues in assessing large groups of students because it requires individual administration.

To meet the demands of NCLB, the need for reliable and valid progress monitoring tools that reflect the multiple dimensions of reading (i.e., decoding, comprehension, and fluency) and allows for the assessment of large group of students has become imperative. According to Kingsbury and Hauser (2004), computer adaptive testing (CAT), which has been used very successfully in the military (Sands, Waters, & McBride, 1997) and in professional certification and licensure (Zara, 1992) may be the answer.

Computer Adaptive Testing (CAT)

Since its development in the 1970s, the use of CAT has increased substantially. Reckase (1988) defines CAT as a computerized testing procedure that selects items to match the ability of the examinee during test administration. The objective of CAT is to construct an optimal test for each examinee. To begin with, an initial item is selected with a difficulty level matches the examinee's current estimated ability. The examinee's response to this item is then scored, and their ability estimate is updated to incorporate this information. The next item presented will be more difficult if a correct answer is given. An incorrect response will result in the presentation of an easier item (Wainer, 2000).

Unlike traditional test format in which everyone is assessed on the same items, different examinees taking a CAT, in all likelihood, take different forms of the test. A very proficient examinee might have few (or even no) items in common with someone who was considerably less proficient. On a traditional test, proficiency may be measured in 'percentage correct.' On a CAT, though, 'percentage correct' would not work because (if the test is working properly) all examinees would get about half of the items presented to them correct. The more proficient examinees would get half of a rather difficult subset correct while the less proficient would get their half out of a much easier subset. This type of testing format is based on the psychometric theory called Item Response Theory (IRT).

Item response theory. IRT has developed over the last 40 years to extend concepts of classical test theory down to the level of individual examinees and test items. A central idea in IRT is the item response function, which relates to the probability of examinees answering particular items correctly to their general level of latent ability usually denoted by the Greek symbol theta (θ) (Lord, 1980, as cited by Parshall, Spray, Kalohn, & Davey, 2002). According to Parshall et al., IRT presents a mathematical characterization of what happens when an individual meets an item. Each individual is characterized by a proficiency parameter (θ), and each item by a collection of parameters – one of which is the item's difficulty. The IRT model compares the person's proficiency with the item's difficulty and predicts the probability of that person getting that item correct. If the person is much more proficient than the item is difficult, then this probability will be large. If the item is much more difficult than the person is proficient, then this probability will be small. The item-choice algorithm tries to pick items that yield the greatest amount of information while at the same time satisfying the variety of content specifications that are critical for a good test. Under IRT, the most is learned about the examinee when this expected probability is close to one-half (p=.5). An examinee's proficiency is calculated from the difficulty of the items that are presented to him or her.

A flowchart for a typical CAT algorithm is depicted in Figure 3. After a test item pool has been developed and calibrated (components one and two are completed), that part of the algorithm is fixed and as many students as desired can be tested (components three through seven). The heart of the process is components four, five, and six. In these components, a test item is selected, administered, and judged, a provisional estimate of ability is made and the termination criterion is tested. If the termination criterion is met, a final estimate of ability is made. If the termination criterion is not met, another test item is administered and the process repeats.

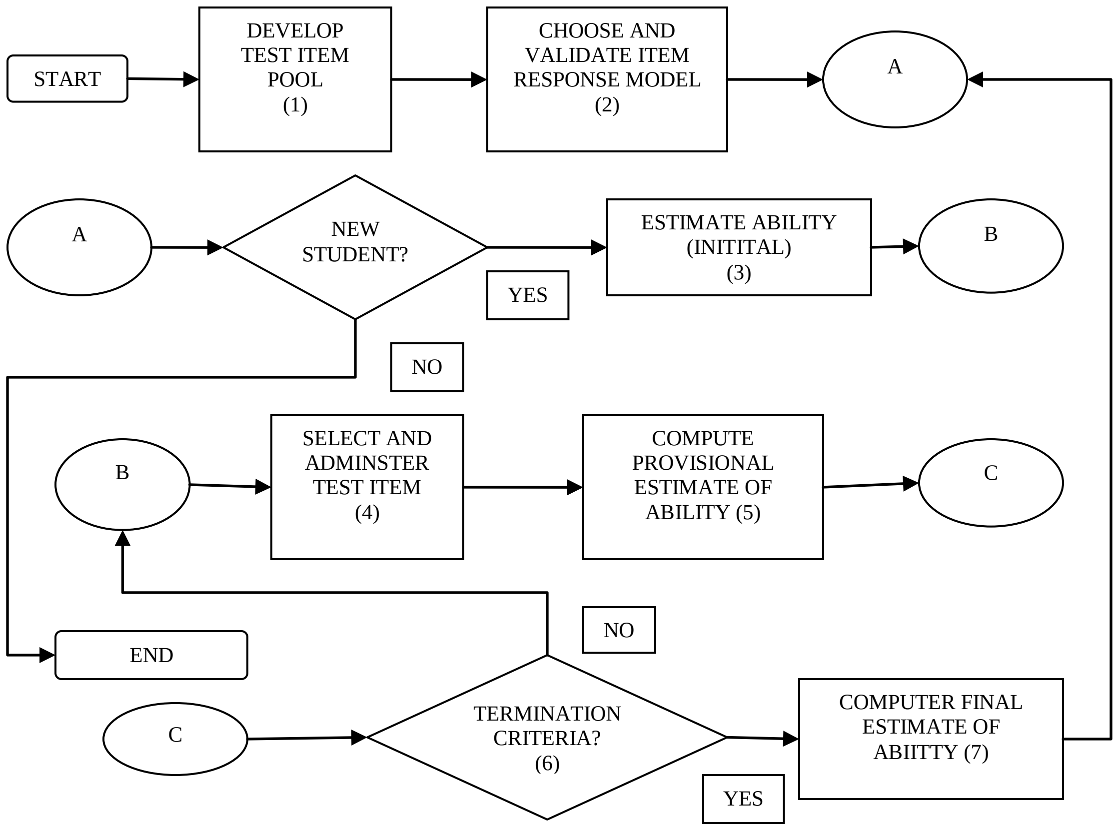

Figure 3. Typical computer adaptive test (CAT) algorithm (Carlson, 1993-94).

The potential of CAT to construct tests that are tailored to examinees' ability levels has generated considerable interest from various educational testing centers around the globe. According to Meijer & Nering (1999), the Graduate Record Examination (GRE) and the Computerized Placement Test (CPT) have operational CAT versions and several licensure boards have also implemented CAT versions of their tests, including the National Council of State Boards of Nursing and the National Board of Medical Examiners. Additionally, the U.S. Department of Defense has implemented a CAT version of the Armed Services Vocational Aptitude Battery (ASVAB; Sands, Waters, & McBride, 1997 as cited by Meijer & Nering, 1999). In the K-12 school system, educational stakeholders are also exploring more efficient measurement tools in place of traditional paper-and-pencil tests (PPTs). In 2006–2007, 23 states were reported to offer computer-based assessments to measure achievement in U.S. schools (Bausell & Klemick, 2007). By 2008, almost all states were expected to have some form of online testing in place (Williams, n.d.). Many educational stakeholders foresee the promise of using computer-based testing (CBT) in their state due to the advantages of CBTs over traditional PPTs in terms of immediate scoring and reporting of students' test results, greater test security, test administration efficiency, flexible test administration schedules, reduced costs compared to handling PPTs, the use of multimedia innovative item types that are not feasible in the PPT format, audio and large-print accommodations for visually impaired students, and the ability to measure response time (Bennett, 2001, 2002; Chaney & Gilman, 2005; Fichten, Asuncion, Barile, Ferraro, & Wolforth, 2009; Folk & Smith, 1998; Klein & Hamilton, 1999; Parshall, Spray, Kalohn, & Davey, 2002). As various agencies look to cross over from PPT to CBT, the issue of test comparability is at the forefront of concerns.

Test mode. Since the early 1900s the fixed length PPT has been the standard method for assessing student achievement. In public education, such tests (i.e., standardized achievement test) continue to be the norm given once or twice a year. While there are numerous benefits to CBTs, evidence is contradictory if identical PPTs and CBTs will obtain the same results. Such findings are often referred to as the test mode effect. While the preponderance of the evidence suggests that, for multiple-choice-only tests, student performance is not significantly different for different modes of administration; some studies suggest students might do better on computer, while others suggest they might do better on paper.

Many comparability studies found computer tests to be equivalent in difficulty or slightly easier than paper tests (Bridgeman, Bejar, & Friedman, 1999; Choi & Tinkler, 2002; Mead & Drasgow, 1993; Nichols & Kirkpatrick, 2005; Pearson Educational Measurement, 2002, 2003; Poggio, Glasnapp, Yang, & Poggio, 2005; Pommerich, 2004; Pomplun, Frey, & Becker, 2002; Russell, 1999; Russell & Haney, 1997, 2000; Russell & Platti, 2001; Schaeffer, Bridgeman, Golub-Smith, Lewis, Potenza, & Steffen, 1998; Taylor, Jamieson, Eignor, & Kirsch, 1998; Wang, 2004; Zandvliet & Farragher, 1997). Results from these studies indicated that student performance was similar across the demographics of sex, academic placement, and SES. Fewer studies found students in K-12 performed poorer on computer tests than paper tests (Cerrillo & Davis, 2004; Choi & Tinkler, 2002; O'Malley, Kirkpatrick, Sherwood, Burdick, Hsieh, & Sanford, 2005; Russell & Plati, 2001). There are a number of possible factors that might help explain some of this inconsistency. Although CBTs are capable of presenting many more items and more flexible formatting than PPTs, most test developers have attempted to avoid content equivalence issues by using identical types of questions in both testing modes. This practice has led to comparability issues especially in the area of reading. For example, Choi and Tinkler (2002) found student scores from the computer tests in reading on the Oregon state assessment were lower than the paper tests, especially for third-graders. Follow-up analyses indicated that it was easier for students to scan reading texts for key words or cued phrases on paper than on a computer screen. Research has shown that long reading passages on a computer tend to be more difficult than on paper (Murphy, Long, Holleran, & Esterly, 2003; O'Malley, et al., 2005).

For tests where all of the information for an item could be presented in its entirety on the screen, results of comparability studies often show small or insignificant mode effects (Bridgeman, Lennon, & Jackenthal, 2001; Choi & Tinkler, 2002; Hetter, Segall, & Bloxom, 1997; Spray, Ackerman, Reckase, & Carlson, 1989). For tests where all of the information for an item could not be presented in its entirety on the screen, and some form of navigation (typically scrolling) was necessary on the part of the examinee to view all of the information, results often showed more significant mode effects (Bridgeman et al., 2001; Choi & Tinkler, 2002). In the process of scrolling through text, sentences are often split across screens, requiring the reader to remember the information in the first part of the sentence or paragraph while paging or scrolling to reveal the rest of the sentence(s). This situation may cause what has been termed the split attention effect (Chandler & Sweller, 1991; Sweller & Chandler, 1991, 1994). According to the split attention effect, the requirements to mentally integrate noncontiguous material impose an unnecessary and heavy load on working memory. The cognitive resources required to integrate material on separate screens may compete with the more meaningful processes of reading comprehension, such as constructing the main idea, making inferences, and identifying important information. However, Kobrin and Young (2003) found that the only significant difference between the CBTs and PPTs in a group of college juniors and seniors was in the frequency of identifying important information in the passage. There was no evidence of any differences in search strategies or in overall test-taking strategies on the CBTs and PPTs. Furthermore, some of their findings indicated that CBTs might actually encourage more construct-relevant behaviors than PPTs.

Subgroup characteristics (Sex, Race, SES, Disabled, Computer Familiarity*)*. While equity is a critical concern, most studies do not focus on comparability for different subgroups of students, for a variety of logistical reasons. Gallagher, Bridgeman, and Cahalan (2000) examined data from several national testing programs to determine whether the change from paper and pencil to computer testing influences group differences in performance. Performance by sex, racial/ethnic, and language groups on the Graduation Record Examination (GRE) General Test, the Graduate Management Admissions Test (GMAT), the SAT I: Reasoning (SAT) test, the Praxis: Professional Assessment for Beginning Teachers (Praxis), and the Test of English as a Foreign Language (TOEFL) was analyzed. This study concluded the change is too small to pose a disadvantage to any of these subgroups. Additional research by Sandene, Horkay, Bennett, Allen, Kaplan, and Oranje (2005) and Nichols and Kirkpatrick (2005) also found no difference in administration mode comparability among various demographic subgroups whereas Sim and Horton (2005), Sandene et al. (2005), and McCann (2006) failed to find any effect associated with sex. With regards to SES, Sandene et al. (2005) found no significant difference in performance associated with parent's educational level, a common proxy of SES. A survey on computer use by students with disabilities in Germany (Ommerborn & Schuemer, 2001) reported more advantages than disadvantages to computer administration. Brown-Chidsey, Boscardin, and Sireci (2001) interviewed students with learning disabilities and found that the computer helped them with limitations that often interfered with the completion of their work. The research concluded, “Students’ beliefs about computers are likely to shape the extent to which instructional technology enhances their achievement” (Brown-Chidsey et al. 2001, p.4).

Concerns persist about whether there are subgroups of students who are disadvantaged because of lack of access, use, or familiarity with computers (Trotter, 2001). The gap in access to technology is continuing to grow. The gap has widened considerably for computer ownership among racial minorities when compared with Whites. In the context of the overall racial digital divide, low-income White children are three times more likely to have Internet access than their Black counterparts, and four times as likely as Latino children in the same socioeconomic category (Bolt & Crawford, 2000). One study (Choi & Tinkler, 2002) did find that computer familiarity was related to computer test performance; students who rarely used a computer tended to perform poorer in both math and reading than those students who had more computer experience.

After students took computerized tests, some studies ask participants whether they would prefer to take future tests on computer or paper. In an evaluation of testing experience, students overwhelmingly preferred computer tests to paper tests (Bridgeman et al., 2001; Glassnapp, Poggio, Poggio, & Ynag, 2005; Higgins, Russell, & Hoffman, 2005; Ito & Sykes, 2004; Johnson & Green, 2006; O’Malley et al., 2005; Pearson Educational Measurement, 2006; Sim & Horton, 2005; Wang, 2004). Most students, regardless of demographics or ability, believed that the computer version was easier, faster, and more fun. Students also responded that using a computer helped concentration by presenting only one question at a time.

In recent years, software companies have attempted to bridge the instructional challenges through packaged, formative assessment programs. The programs are intended to replace traditional paper/pencil assessments and provide immediate feedback along with instructional plans and/or interventions (Villano, 2006). Most companies hire consultants with curriculum development experience to write questions for formative assessment tools. Each vendor claims that the formative assessment tool is aligned to the state standards, but it is imperative that school systems evaluate the validity of the assessment.

In South Carolina, the South Carolina Code Ann. § 59-18-310 (E) (Supp. 2007) provided for the creation of a statewide adopt^2□□□list of formative assessments in English language arts (ELA) and mathematics. The legislation requires that the formative assessment satisfy professional measurement standards and align with the South Carolina Academic Standards. A panel of measurement experts reviewed studies submitted by seven companies to determine if the products positively impacted student achievement. To evaluate the submissions, the Education Oversight Committee and the SCDOE jointly developed evaluation criteria. Measures of Academic Progress (MAP), published by the Northwest Evaluation Association was approved for the Adoption List by the State Board of Education on March 14, 2007 (SCDOE, n.d.b).

Measures of Academic Progress (MAP)

Measures of Academic Progress (MAP) developed by the Northwest Evaluation Association (NWEA) is a computer adaptive test which adjusts to match the performance of the student after each item is given. NWEA was founded in 1976 by a group of school districts looking for practical answers to measure efficiently and accurately how much students have achieved and how quickly they are learning (NWEA, 2003). Since then, NWEA has created one of the most widely used computer adaptive tests. MAP is used by more than 2,340 school districts in the United States and in 61 other countries (Ash & Sawchuk, 2008). While most districts choose to use the MAP to assess reading, math and language arts, tests for science are also available. Most districts administer the MAP twice a year from grades 2-10; however, districts have the option to administer the tests up to three to four times within a school year to better monitor student growth. All of NWEA's tests are untimed, but supervised. The computerized MAP takes 45 minutes to an hour per subject for most students.

MAP tests report student performance on a single, cross-grade scale, which NWEA calls the RIT scale, which is short for Rasch Unit after the scaling theory's founder, Danish statistician

Georg Rasch. This scale is based on the same modern test theory used by the SAT, GRE, and LSAT (NWEA, 2003). The benefit of this test theory is that it aligns student achievement levels with item difficulties on the same scale. NWEA places all of the test items on the RIT scale according to their difficulty. Each increasing RIT is assigned a numeric value, or RIT score, that indicates a higher level of difficulty. As a student takes a MAP test, he or she is presented with items of varying RITs, or levels of difficulty. Once the MAP system determines the difficulty level at which the student is able to perform and the system collects enough data to report a student's abilities, the test ends and the student is assigned an overall RIT score.

According to NWEA (n.d.a), the characteristics of the RIT scale provide several benefits to educators:

> Grade-independent
>
> Because the tests are adaptive and the test items displayed are based on student performance, not age or grade, identical scores across grades mean the same thing. For example, a third grader who received a score of 210 and a fourth grader who received a score of 210 are learning at the same level. This allows growth to be measured independent of grade.
>
> Equal-interval
>
> The RIT scale is infinite, but most student scores fall between the values of 140 and 300. Like meters or pounds, the scale is equal-interval, meaning that the distance between 170 and 182 is the same as the distance between 240 and 252. This allows educators to apply simple mathematical equations to the scores to determine information such as the mean and median scores in a class or grade.
>
> Stability
>
> More than twenty years after it was first implemented, scores along the RIT scale mean the same thing. As a result, educators can confidently measure growth over many years (para. 8).

In addition to RIT scores, goal performance areas are also reported that are aligned to the content of the individual state standards or benchmarks. According to NWEA (2003), since there are so few items administered in a single goal, approximately seven items per goal, goal scores have a relatively high standard error of measurement. Thus, goal performance is more accurately reported in ranges rather than specific scores. A goal performance of LO means that the student is performing at the 33rd percentile or lower. AV means that the student is performing between the 33rd and 66th percentile, while HI means that the student is performing at or above the 66th percentile (NWEA, 2003).

To provide each student a challenging test with an accurate score, the MAP system requires the use of large pools of items with a difficulty range appropriate for all of the students being tested. MAP item pools generally contain 1,200-2,400 items (NWEA, 2003). Most MAP assessments have about four to eight goals with five to six sub-goals each, and contain between 40 and 50 items (NWEA, 2003). Each year hundreds of new items are developed by teachers who receive thorough training in the item-writing processes. Each potential item must pass a rigorous bias and content review, which is followed by field-testing with a minimum of 300 students. Only those items that pass the bias review, field-testing, and the subsequent strict statistical screening procedures are calibrated for difficulty and assigned the appropriate value on the RIT scale. These items become part of the continually expanding item bank.

Since the MAP is used across the country in various states, NWEA conducts regular state alignment studies to examine the correspondence between the MAP and state standardized tests used to measure student achievement. The alignment process begins with a thorough review of a state's standards document by NWEA's curriculum specialists. The general goal areas or strands within a state's standards that appear across grade levels become the goals in the goal structure. Areas in a state's standards documents that are determined to be sub-domains of the goals/strands become the sub-goals in the goal structure.

Reliability and validity. NWEA did an extensive study of validity and reliability of the MAP assessment in 2004. This study has been followed by yearly smaller studies which have yielded similar results. In the 2004 study, NWEA used a test-retest reliability method and obtained results between .76 and .93. It should be noted that due to the nature of the MAP assessments logic, students were exposed to none of the same items and, in fact, it would be expected that between the first administration and the second administration of the test, the student would show growth. Therefore, “it would not seem unreasonable to expect reliability to drop below .80” (NWEA, 2004, p.2).

When looking at the validity of the MAP assessment, NWEA notes “content validity of NWEA tests is assured by carefully mapping existing content standards from a district or a state to a test blueprint. Test items are selected for a specific test based on their match to the content standards as well as on the difficulty level of the test being created” (NWEA, 2004, p4). In the 2004 study, NWEA employed concurrent validity using a number of state and nationally normed tests in close proximity to an administration of the MAP test. Pearson correlation coefficients ranged from .69 to .88 and averaged approximately .85. These correlations are well within the acceptable range for both validity and reliability.

Summary

The effectiveness of public education in the United States has long been scrutinized in our society. Proficient readers remain a minority as evidenced by the continuing concerns over illiteracy rates (Kirsch et al., 1993; Marshall, 2009; NCES, 2006; Roman, 2004) and those identified as Learning Disabled (Flugum & Reschley, 1994; Meredith et al., 1997; Scruggs and Mastropieri, 2002; Van Haren et al., 2006). For this reason, reading is one of the target subject areas accountability legislation hopes to improve through the monitoring of high-stakes testing.

The "achievement gap" in education refers to the disparity in academic performance between groups of students. The term is most often used to describe the troubling performance gaps between many Black and Hispanic students, at the lower end of the performance scale, and their White peers, at the higher end of the performance scale. A similar academic disparity is evident between students from low-income and high-income families. The racial gap itself has been a major impetus for federal education policy as embodied in NCLB and has entered into countless state and local debates regarding school finance equalization, academic tracking, and school testing and accountability programs. In an atmosphere of educational reform and federally mandated high-stakes testing, demands have increased for progress monitoring strategies that reliably predict outcomes on statewide assessments (McGlinchey & Hixson, 2004).

Many school districts across the nation have begun to use Curriculum-Based Measurement-Reading (CBM-R) to detect those students not on track to be proficient readers from an early age. A pitfall exists in the use of CBM-R in that empirical findings of CBM-R research across grades have demonstrated a deceleration in growth of reading fluency skills as grade levels increased. To meet the demands of NCLB, the need for reliable and valid progress monitoring tools that reflect the multiple dimensions of reading (i.e., decoding, comprehension, and fluency) and allows for assessment of large group of students at all grade levels has become imperative. According to Kingsbury and Hauser (2004), computer adaptive testing may be the answer. In this paradigm, the test adapts to match the difficulty of the questions administered to the performance of each student as the student takes the test. Many educational stakeholders foresee the promise of using computer-based testing (CBT) in their state due to the advantages of CBTs over traditional PPTs in terms of immediate scoring and reporting of students' test results, greater test security, test administration efficiency, flexible test administration schedules, reduced costs compared to handling PPTs, the use of multimedia innovative item types that are not feasible in the PPT format, audio and large-print accommodations for visually impaired students, and the ability to measure response time (Bennett, 2001, 2002; Chaney & Gilman, 2005; Fichten et al., 2009; Folk & Smith, 1998; Klein & Hamilton, 1999; Parshall et al., 2002). Measures of Academic Progress (MAP) is one of the most widely used computer adaptive tests in school systems today.

CHAPTER III

METHODS

This study investigated the predictive validity of demographic variables and the Measures of Academic Progress (MAP) in relation to student performance on the South Carolina's Palmetto Assessment of State Standards (PASS) English language arts (ELA) test. The state test was selected because it is the measure connected to high-stakes for school districts and it is viewed as a comprehensive measure of reading skill.

Design

This research study was an exploration of various predictive factors for student performance on the PASS ELA test. This was a correlational design. Demographic factors included sex, race/ethnicity, socioeconomic status (SES), special education, and degree of exposure to computer adaptive testing (CAT). The specific predictive factors of the MAP that were analyzed included the MAP Reading Rasch Unit (RIT) score, the MAP Understanding and Using Literary Texts Goal Performance RIT score, the MAP Understanding and Using Informational Texts Goal Performance RIT score, and the MAP Building Vocabulary Goal Performance RIT score.

Archival test data were included in the sample when a student record had both a Spring 2009 MAP reading score and a Spring 2009 PASS ELA score. Demographic information and test data were collected from students in grades 3-8 from five elementary and three middle schools located in the target school district. Figure 4 depicts the design of the study.

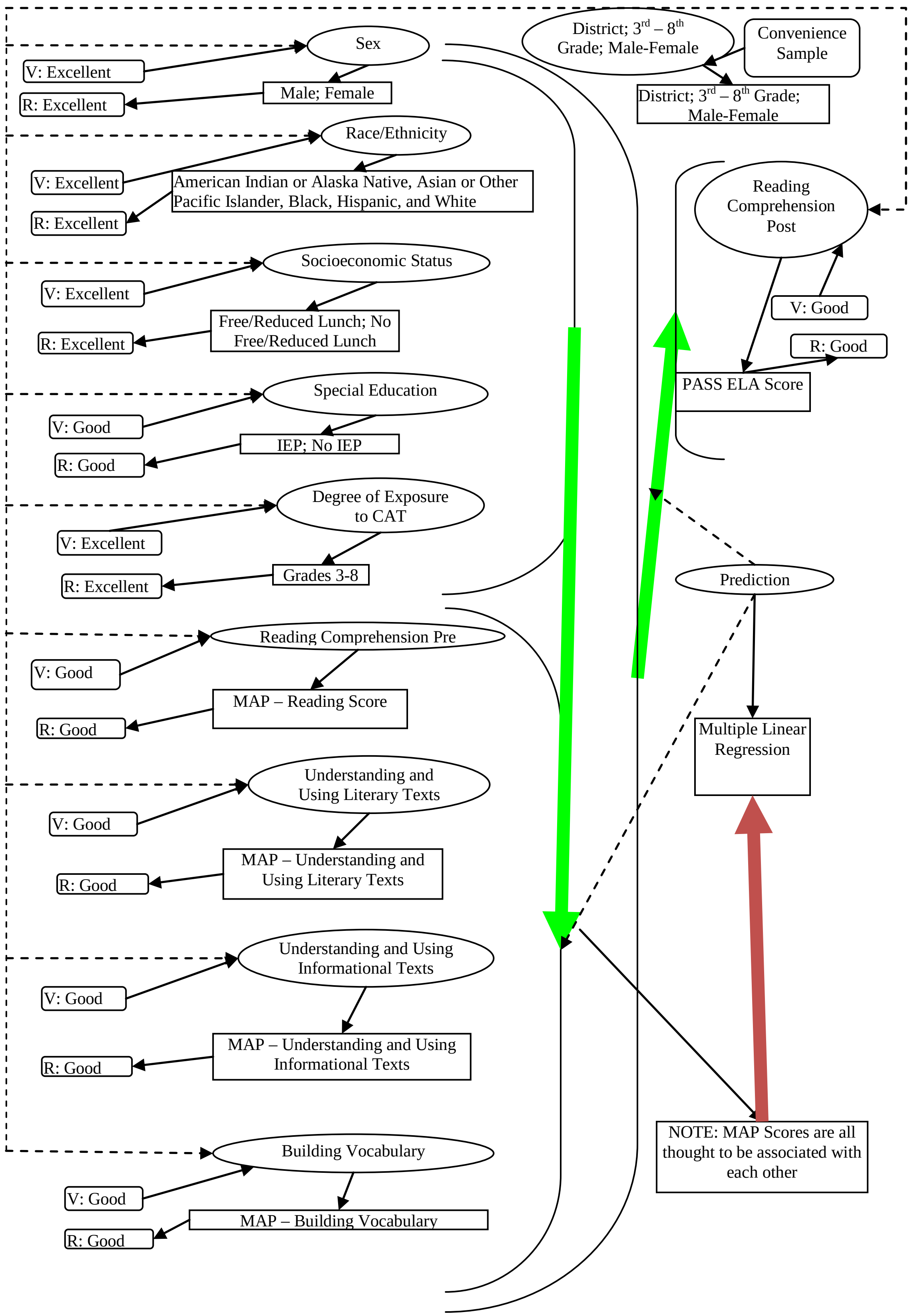

Figure 4. Research path diagram of the NWEA MAP and PASS ELA study.

Population

The population utilized in this study was students in grades 3-8 in the target school district who are mandated to take the PASS for statewide accountability testing purposes. The district is located in a predominantly Caucasian middle to upper middle-class community in upstate South Carolina. Based on the State Report Card for the target school district (South Carolina Department of Education [SCDOE], 2009), 5,113 students were administered the PASS ELA during the 2008-2009 school year. Of the 5,113 students, 51.6% were male (n = 2,637) and 48.4% were female (n = 2,476). The racial breakdown of the population included 80.3% White (n = 4,104), 11.6% Black (n = 593), and 8.1 % Other (n = 416). The Other category consisted of 4.8 % Hispanic (n = 246), 3.0 % Asian & Pacific Islander (n = 152) and 0.3% American-Indian or Alaska Native (n = 18). With regards to SES, 82.1% of the population did not qualify for free/reduced lunch (n = 4,199) and 17.9% qualified for free/reduced lunch (n = 914). With regards to special education, 10.3% of the sample received some form of special education (n = 525) while 89.7% did not receive special education services (n = 4,588). Information regarding the breakdown of the population by grade was not available at the time of this study. Table 1 contains a summary of the PASS ELA population demographics.

Table 1

PASS ELA Population Demographics

Demographics		Number of Students (N = 5,113)
Sex		
	Male	2,637 (51.6%)
	Female	2,476 (48.4%)
	Total	5,113 (100%)
Race/Ethnicity		
	White	4,104 (80.3%)
	Black	593 (11.6%)
	Other	416 (8.1%)
	Total	5,113 (100%)
Socioeconomic Status		
	No Free/Reduced Lunch	4,199 (82.1%)
	Free/Reduced Lunch	914 (17.9%)
	Total	5,113 (100%)
Special Education		
	Yes	525 (10.3%)
	No	4,588 89.7%)
	Total	5,113 (100%)

Sample

The sample comprised of archival test data from students in grades 3-8 in the target school district. Archival test data were excluded from the study if the student did not have both a Spring 2009 MAP reading score and a Spring 2009 PASS ELA score. In addition, students in grades 3-8 who participated in the South Carolina Alternate Assessment (SC-Alt) were excluded from the study. The SC-ALT is a state assessment designed for students with significant cognitive disabilities who participate in a school curriculum that includes functional and life skills as well as academic instruction. Under the South Carolina Education Accountability Act (1998), school districts are allowed to assess one percent of the student population with SC-ALT.

Assignment

No assignment was required for this study. Archival test data and demographic information were obtained from five elementary and three middle schools located in the target school district. The sample comprised of students in grades 3-8 with available demographic data who had scores on the Spring 2009 MAP Reading and Spring 2009 PASS ELA assessments. All students matching the aforementioned criteria were included in the study.

Measurement

All demographic information and test scores was exported from *TestView* (2004), which is the web-based program the school district currently uses to store and analyze various standardized test data. All demographic information on *Testview* is exported from the statewide School Administration Student Information (SASI*xp*™, 2003) system used by all South Carolina public schools. The information on SASI*xp* is obtained directly from student's parents during school enrollment.

Demographic Information

Sex. The latent variable of sex is represented by male and female. The reliability and validity of this variable is considered excellent because the information was obtained directly from the student's parents at the time of enrollment into the school district.

Race/Ethnicity. The latent variable of race/ethnicity is represented by American Indian or Alaska Native, Asian or Other Pacific Islander, Black, Hispanic, and White. Race/ethnicity categories are defined as follows:

American Indian or Alaska Native: A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

Asian or Other Pacific Islander: A person having origins in any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent, or the Pacific Islands. This includes, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, Vietnam, Hawaii, Guam, and Samoa.

Black: A person having origins in any of the black racial groups of Africa.

Hispanic: A person of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish culture or origin, regardless of race.

White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa. (National Center for Education Statistics, 2002, para. 5)

The reliability and validity of this variable is considered excellent because the information was obtained directly from the student's parents at the time of enrollment into the school district

Socioeconomic status (SES). The latent variable of SES is represented by those students qualifying for the National School Lunch Program. The National School Lunch Program is a federally assisted meal program that provides low-cost or free lunches to eligible students. It is sometimes referred to as the free/reduced-price lunch program. Free lunches are offered to those students whose family incomes are at or below 130 percent of the poverty level; reduced-price lunches are offered to those students whose family incomes are between 130 percent and 185 percent of the poverty level (U.S. Department of Agriculture, 2008). The reliability and validity of this variable is considered excellent because free and reduced-lunch is frequently used as a valid indicator of SES in educational research.

Special education. The latent variable of special education is represented by those students qualifying for special education and those who do not. To qualify for special education, the student must meet state and district eligibility criteria for special education in one of the 13 areas: Autism, Deafblindness, Deaf and Hard of Hearing, Developmental Delay, Emotional Disability, Mental Disability, Multiple Disabilities, Other Health Impairment, Orthopedic Impairment, Traumatic Brain Injury, Specific Learning Disability, Speech or Language Impairment, or Visual Impairment. The reliability and validity of this variable is considered good because eligibility criteria varies from state to state.

Degree of exposure to computer adaptive testing. The latent variable of the degree of exposure to CAT is represented by grades 3-8. Students in the target school district are initially administered the MAP in the winter of second grade. Starting in the third grade, students are administered the test in the fall and spring of each school year. As the student progresses in school, his/her exposure to CAT will increase by the function of his/her grade. The reliability and validity of this variable is considered excellent because access to computers has increase significantly over the past several decades both in homes and at schools.

Measures of Academic Progress (MAP)

The MAP developed by the Northwest Evaluation Association (NWEA) is a CAT that adjusts to match the performance of the student after each item is given. The data from the MAP system (the tests and student data) are downloaded via the Internet from NWEA's Web site to the district file server. An application to administer the tests is also downloaded from NWEA's Web site and distributed to each workstation used to administer tests. When the testing session is complete, student data is uploaded to NWEA for scoring and reporting services. The data remain only on the file server until test results are uploaded. The application for administering the tests remains on the workstations for future use. All of NWEA's tests are untimed, but supervised. The computerized MAP takes approximately 45 minutes to an hour per subject for most students.

The MAP reports student performance on a single, cross-grade scale, which NWEA calls the RIT scale, short for Rasch Unit after the scaling theory's founder, Danish statistician Georg Rasch (NWEA, 2003). NWEA conducts regular state alignment studies to examine the correspondence between the MAP and state standardized tests used to measure student achievement. The alignment process begins with a thorough review of a state's standards document by NWEA's curriculum specialists. The general goal areas or strands within a state's standards that appear across grade levels become the goals in the goal structure. Areas in a state's standards document that are determined to be sub-domains of the goals/strands become the sub-goals in the goal structure. Table 2 provides the South Carolina Framework for Reading based on NWEA state alignment studies.

Table 2

South Carolina Goal Structure

1. Understanding and Using Literary Texts

    a. Predictions, Conclusions, Inferences

    b. Point of View, Characters, Setting, Plot, Theme

    c. Devices of Figurative Language, Author's Craft

    d. Summarizing, Details, Main Idea

    e. Characteristics of Fiction, Nonfiction, Poetry

2. Understanding and Using Informational Texts

    a. Summarize Evidence that Supports the Central Idea

    b. Draw Conclusions and Make Inferences

    c. Understand Cause-and-Effect Relationships

    d. Facts and Opinions, Author Bias, Propaganda

    e. Text Elements, Graphic Features, Text Features

3. Building Vocabulary

    a. Using Context Clue to Determine Meaning

    b. Base Word, Prefix, Suffix, Idiom, Connotation

    c. Phonemic Awareness and Phonics

    d. Synonyms, Antonyms, and Homonyms

Adapted from "Using SC 2007 Academic Standards for English Language Arts" by NWEA, (2007b)

South Carolina Code Ann. § 59-18-310 (E) (Supp. 2007) provided for the creation of a statewide adoption list of formative assessments in English language arts (ELA) and mathematics. The legislation requires that the formative assessment satisfy professional measurement standards and align with the South Carolina Academic Standards. A two-stage process was implemented to select the products considered for adoption. In the first stage, a panel of measurement experts reviewed studies submitted by seven companies to determine if the products positively impacted student achievement. Using evaluation criteria developed jointly by the Education Oversight Committee and the SCDOE a panel of 14 measurement experts evaluated twelve submissions. The MAP was approved for the Adoption List by the State Board of Education on March 14, 2007. During the second stage of adoption list process, items from the MAP assessments were resubmitted to SCDOE for an alignment with the new state academic standards. Two committees of curriculum specialists (ELA and mathematics) convened in September 2008 to determine the alignment of each item to the South Carolina academic standards and indicators. Table 3 show the extent to which MAP items are aligned with the 2007 South Carolina English language arts and mathematics standards.

Table 3

Northwest Evaluation Association—MAP Reading and Language

Standard	3rd Grade # Items	4th Grade # Items	5th Grade # Items	6th Grade # Items	7th Grade # Items	8th Grade # Items
1	183	172	88	78	34	19
2	90	122	118	51	34	46
3	85	52	32	21	26	37
4	209	166	101	82	60	73
5	8	19	16	10	10	8
6	7	37	24	10	8	8
Not Aligned	90	46	69	97	61	54
% Aligned	87%	93%	85%	72%	74%	78%
Total Items	672	614	448	349	233	245

Adapted from “2008–09 Adoption List of Formative Assessments Alignment by Standards for Each Subject and Grade” by SCDOE (n.d.b)

Goal performance areas are also reported and are aligned to the content of the individual state standards or benchmarks. According to NWEA (2003), since there are so few items administered in a single goal, approximately seven items per goal, scores have a relatively high standard error of measurement. Thus, goal performance is more accurately reported in ranges rather than specific scores. A goal performance of LO means the student is performing at the 33rd percentile or lower. AV means the student is performing between the 33rd and 66th percentile. HI means the student is performing at or above the 66th percentile (NWEA, 2003).

Reliability. The reliability of the MAP has been calculated in two different manners. One method used marginal reliability which may be applied to any tests constructed using item response theory (IRT). Marginal reliability uses the test information function to determine the expected correlation between the scores of two hypothetical tests taken by the same student. The marginal reliability for MAP Reading across grades 3-8 ranged between r = .93 to .95 (NWEA, 2003). The test-retest reliability of the MAP has also been investigated. The correlation between the pairs of scores of students from Spring to Fall, Spring to Spring, and Fall to Spring were calculated. Test-retest values from grades 3-8 ranged from r = .85 to .91 (NWEA, 2003).

Validity. Due to the implementation of a new statewide accountability testing program (i.e. Palmetto Assessment of State Standards [PASS]) for the 2008-2009 school year, no statistical data is currently available for the alignment between the MAP and the PASS. NWEA has conducted prior studies to investigate the alignment of cut scores between South Carolina's previous high-stakes test (i.e., Palmetto Achievement Challenge Test [PACT]) and the MAP tests (e.g., Cronin & Hauser, 2003). These studies have suggested that the MAP tests are reliable and valid predictors of student performance on the PACT. Cronin and Hauser (2003) first analyzed the accuracy of the RIT scale in correctly predicting whether students are likely to be proficient or advanced on the PACT (which was referred to as proficiency status). The accuracy of reading proficiency status predictions ranged from a low of about 77% for grade 3 to a high of about 85% for grade 8. They also analyzed how accurately the RIT scale predicted proficiency level assignment on PACT (i.e., estimating whether students would finish in the *below basic, basic, proficient*, or *advanced level*). Using second order regression, prediction indexes range from a low of R^2 = .656 in sixth grade to a high of R^2 = .824 in the eighth grade.

In 2004, Cronin conducted another study of the ongoing alignment of the NWEA RIT Scale with the South Carolina PACT. Over 22,000 test records of students were included in this study. Concurrent validity was tested by examining same subject Pearson correlations between the MAP and the PACT. Same subject correlations for the PACT ELA and the MAP Reading ranged from .76 to .79. Discriminant validity was tested by examining same subject Pearson correlations next to correlations for the alternate subject (math against reading and language usage). In all cases, the same subject correlations were higher than correlations against the alternate subject (Cronin, 2004).

Palmetto Assessment of State Standards (PASS)

In June 2008, the South Carolina legislature amended the Education Accountability Act of 1998 (Act 282) to incorporate a new assessment in grades 3-8 known as the Palmetto Assessment of State Standards (PASS). According to the SCDOE (n.d.a), beginning with the 2008-2009 school year, the PASS test results will be used for statewide accountability testing purposes and include tests in writing, ELA (reading and research), mathematics, science, and social studies. For each PASS test, three overall performance levels are reported: *Exemplary* – The student demonstrated exemplary performance in meeting the grade level standard; *Met* – The student met the grade level standard; *Not Met* – The student did not meet the grade level standard.

For the purposes of this study, the PASS ELA will only be examined. The PASS ELA assesses a student's mastery of the 2008 South Carolina Academic Standards in grades 3-8. At each grade level, four broad standards are assessed: Standard 1 – Literary Texts, Standard 2 – Informational Texts, Standard 3 – Vocabulary, and Standard 6 – Research. Table 4 shows test item distribution per grade level.

Table 4

PASS ELA Test Blueprint for Grades 3-8

Grade	Total Number of Items	Standard	Item Number Ranges Per Standard
		1	8-12
3	36	2	8-11
		3	8-9
		6	8-10
		1	8-12
4	36	2	8-11
		3	8-9
		6	8-10
		1	8-14
5	38	2	8-12
		3	8-9
		6	8-10
		1	10-15
6	40	2	9-12
		3	8-10
		6	8-10
		1	12-18
7	45	2	9-14
		3	8-10
		6	8-12
		1	12-18
8	50	2	10-16
		3	8-10
		6	10-14

Reliability and validity. Information regarding the reliability and validity of the PASS is currently not available. The test was administered for the first time during the 2008-2009 school year.

Procedures

During the 2008-2009 school year, students in grades 3-8 in the target school district participated in the MAP Reading during the designated Spring test term window to assess the current instructional level of each student and measure growth over time. Upon completion of the test term window, and after 24 hours has passed, district level-reports are ordered and uploaded into *TestView* (2004) by district-level administrators. Score reports are available at the individual, class, school, and district level. Individual reports are sent home with each student's report card. Beginning with the 2008-2009 school year, students in the target district were also administered a new assessment in grades 3-8 known as the PASS. The PASS is used for statewide accountability testing purposes and includes tests in writing, English language arts (reading and research), mathematics, science, and social studies. The students were administered the English language arts (reading and research) portion of the PASS on May 12, 2009.

Written permission to conduct the study was obtained from the district's Superintendent of Schools. A copy of the letter requesting permission to conduct the study appears in Appendix A. Archival test data and demographic information were obtained from five elementary and three middle schools located in the target school district, where the researcher was employed as the Coordinator of Special Services at the time of the study.

All archival test data were held in strict confidence. No individual test scores were reported. All demographic information, MAP, and PASS test scores were exported from *TestView* (2004) into a Microsoft Excel spreadsheet. To ensure the anonymity of individual test data, all data were encoded by number with the student's name and student's ID having been removed by a third party prior to analysis by this investigator. In addition, to further guarantee that no personally identifiable information about an individual student was revealed, all individual test scores were combined and only results by demographic groups (i.e., sex, race/ethnicity, SES, special education, degree of exposure to CAT) were analyzed and reported. Because of the anonymity of the student data, no parental consent was required, but this project was reviewed, and approved in June 2009, under the methods and procedures applied to human subjects by the Indiana University of Pennsylvania's Institutional Review Board. Table 5 shows the timeline from preparation to conclusion of the study.

Table 5

Timeline for Research Study

March/April 2009	Students in Grades 3-8 participated in MAP Reading
May 2009	Students in Grades 3-8 participated in PASS ELA
	Permission obtained from superintendent of schools to conduct study
June 2009	IRB approved by the Indiana University of Pennsylvania's Institutional Review Board
December 2009	PASS scores uploaded into *TestView* (2004) by administrators at the district level. Score Reports for the PASS were available at the individual, school, and district level. Individual score reports were sent home with students.
January 2010	All demographic information and test scores were exported from *TestView* (2004) and analyzed with the Statistical Package for the Social Sciences (SPSS) statistical program.
June 2010	Defense of study

Sample Size

The number of predictors, power, and effect size were considered when determining an adequate sample size. The power of a study is the probability of not overlooking an effect or a relationship that exists between variables (Rosenthal & Rosnow, 1984). Eighty percent is a conventional figure for the minimum power when conducting a study (Cohen, Cohen, West, &

Aiken, 2003). The effect size refers to the degree to which the dependent variable is related to the predictor variable. Analysis of Variance (ANOVA), correlation, and regression analysis require a large sample size in order to generate stable coefficients and provide sufficient power to reject false null hypotheses. The minimum number of cases recommended to construct a Multiple Linear Regression (MLR) model varies with respect to the effect size and the number of predictor variables in the model. Assuming a small effect size (i.e., an R^2 value < 0.1) the number of cases required to construct an MLR model with up to 8 predictor variables is 757 (Cohen, 1988). The expected sample size of over 3,000 subjects used in this study was, therefore, more than adequate to provide sufficient statistical power for purposes of statistical analysis.

Statistical Analyses

This study investigated the predictive validity of demographic variables and the MAP in relation to student performance on the South Carolina's PASS ELA test. The relationships between assorted variables for student performance on the PASS ELA test were analyzed. The levels of measurement of the observed variables determine the types of statistical analysis that should be performed (Field, 2009). The variables measured in this study with respect to their levels of measurement are specified in Table 6.

Table 6

Research Variables

Scale/interval level variables
PASS ELA (score)
MAP Reading (score)
MAP Understanding and Using Literary Texts (score)
MAP Understanding and Using Informational Texts (score)
MAP Building Vocabulary (score)
Sex (1 = Male, 0 = Female)
Race/Ethnicity (1 = White, 0 = Not White)
(1 = Black, 0 = Not Black)
Special Education (1 = Special Education, 0 = No Special Education)

Ordinal variables	Hierarchical codes
Socioeconomic status	1 = Free lunch
	2 = Reduced lunch
	3 = No free/reduced lunch
Degree of exposure to CAT	Grade 3, 4, 5, 6, 7, or 8

The variables were coded in the SPSS data editor, and statistical analyses were performed using SPSS following methods described by Field (2009). The data were analyzed using associational measures based on Cross-tabulation, Multi-factorial ANOVA, Pearson correlation, and MLR leading to the construction of a hypothetical path model. The validity of null hypotheses was tested using inferential statistics. The p value of the inferential statistic indicated whether the null hypothesis should be accepted. Although it was not possible to prove that a null hypothesis was true; it was possible to reject a null hypothesis, and conclude that the results of the investigation did not happen by chance. Rejecting the null hypothesis was equivalent to disproving it, thereby providing objective evidence for the acceptance of the alternative hypotheses as the only logical choice. Accepting an alternative hypothesis did not imply that it was proven unequivocally. It meant only that there was not enough statistical evidence to disprove it.

The decision rule was to reject the null hypotheses and assume that the tests were statistically significant, if the p value of the test statistics was less than the selected significance level $\alpha = .05$. The results of the statistical tests were interpreted assuming that statistical significance and scientific significance were not equivalent. The issue of statistical versus scientific significance must be considered here with respect to the controversy concerning whether or not the use of null hypothesis significance tests can be justified for the purposes of this research. Some statisticians assert that null hypothesis significance tests are meaningless and should be banned in educational research (Hunter, 1997; Kline, 2004). They argue that dichotomous decisions based upon whether or not the p value of a test statistic is less than or greater than a pre-determined arbitrary level such as $\alpha = 05$ may indicate whether the observed data deviate from that which might be expected by chance. However, they provide no useful information whatsoever about the scientific significance, practical implications, and meaningfulness of observed data in reality, particularly if the sample size is large. A survey of American Educational Research Association (AERA) members indicated that 19% agreed (Mittag & Thomson, 2000). However, the APA Task Force on Statistical Inference did not support such radical views, but instead recommended that researchers should always provide effect size estimates when reporting p values for null hypothesis significance tests (Wilkinson & Task Force on Statistical Inference, 1999). This study followed the APA recommendation, so the effect sizes, indicating the proportion(s) of the variance in the dependent variable explained by the variance in the predictor variable(s) were computed and interpreted. High scientific significance was accredited to results with a large effect size, and low scientific significance was accredited to small effect sizes, irrespective of the magnitudes of the p values. The conventional distinction between small, medium, and large categories of effect size defined by Field (2009) based on the values of η^2 (for ANOVA), R^2 (for correlation and regression analysis), and Cramer's V coefficient (for χ^2 analysis) were applied in this study. Nevertheless, the view of mandating effect size in preference to statistical significance is not supported by all statisticians. Effect sizes must be interpreted with caution and cannot be compared directly between one analysis and another. Field's (2009) categories are only guidelines and cannot be interpreted as accurate and precise indicators of practical significance. Consequently, the use of effect sizes without recognition of their inherent limitations does not always add to a better understanding of research findings (Olejnik & Algina, 2000).

Research Question #1

What is the relationship between and among the demographic variables and student performance on the PASS ELA? How do these variables interrelate?

It was hypothesized that there is a significant difference in student performance between and among sub-categories of the various demographic variables on the PASS ELA. To answer Research Question #1, the five demographic variables were partitioned into sub-questions.

1.1 Is there a difference in student performance between males and females on the PASS ELA?

It was hypothesized that a difference in student performance existed between males and females on the PASS ELA. Based on previous data from the National Assessment of Educational Progress (NAEP) and the target school district, females on average score higher in reading than their male counterparts (Lee, Grigg, & Donahue, 2007; National Center for Education Statistics [NCES], 2009a; Perie, Grigg, & Donahue, 2005; SCDOE, n.d.c).

1.2 Is there a difference in student performance between American Indian or Alaska Native, Asian or Other Pacific Islander, Black, Hispanic, and White students on the PASS ELA?

It was hypothesized that a difference existed in student performance between White, Black, and Other students on the PASS ELA. Based on previous data from the NAEP and the target school district, White and Asian/Pacific Islander students on average score higher in reading than their Black, Hispanic, and American Indian/Alaskan peers (Lee et al., 2007; NCES, 2009a; Perie et al., 2005; SCDOE, n.d.c).

1.3 Is there a difference in student performance between students who receive free/reduced lunch and students who are not eligible for free/reduced lunch on the PASS ELA?

It was hypothesized that a difference existed in student performance between students who receive free/reduced lunch and students who are not eligible for free/reduced lunch on the PASS ELA. Based on previous data from NAEP and the target school district, students who were not eligible for free/reduced lunch on average score higher in reading than their free/reduced lunch counterparts (Lee et al., 2007; NCES, 2009a; Perie et al., 2005; SCDOE, n.d.c).

1.4 Is there a difference in student performance between disabled and nondisabled students on the PASS ELA?

It was hypothesized that a difference existed in student performance between disabled and nondisabled students on the PASS ELA. Based on previous data from NAEP and the target school district, students who were nondisabled on average score higher in reading than their disabled counterparts (Lee et al., 2007; NCES, 2009a; Perie et al., 2005; SCDOE, n.d.c).

1.5 Is there a difference in student performance between students in grades 3-8 on the PASS ELA?

It was hypothesized that a difference existed in student performance between students in grades 3-8 on the PASS ELA as many states set the bar significantly lower in elementary school than in middle school (Cronin, Dahlin, Adkins, & Kingsbury, 2007).

To measure the strengths of the associations between nominal and ordinal variables (i.e., sex, race/ethnicity, socioeconomic status, special education, and degree of exposure to CAT), Cramer’s V statistics were computed by cross-tabulation analysis. Cramer's V is a correlation coefficient that indicates the relationship amongst two categorical variables. Like Pearson's coefficient, Cramer's V ranges from -1 to +1, with 0 indicating no relationship and -1 or +1 indicating a perfect relationship. Also like Pearson's coefficient, the square of Cramer's V indicates the proportion of the total possible association (i.e., the maximum possible value of the chi-square statistic) that is present in the data. The conventional interpretation of Cramer’s V coefficient was applied (i.e., < 0.1 = little, if any, association; 0.1 to 0.3 = weak or low association; 0.3 to 0.5 = moderate association; > 0.5 = high or strong association) (Field, 2009).

Multi-factorial ANOVA was applied to determine if the mean values of the dependent variable measured at the scale/interval level (i.e., PASS ELA scores) varied with respect to the effects of categorical predictor variables (i.e., sex, race/ethnicity, socioeconomic status, special education, and degree of exposure to CAT). An assumption upon which ANOVA is based was that the variances of the scores within each combination of predictor variables were equal, otherwise known as homogeneity of variance. Violation of the assumption of homogeneity of variance compromises the results of ANOVA; therefore, it must be checked before ANOVA is performed. Another assumption is that the residuals (the differences between the mean values and the observed values) are normally distributed and scattered randomly and evenly around their mean (zero) value. To test for normality, a frequency distribution of the residuals was observed. Scatter plots of the residuals versus the predicted values were also constructed. Homogeneity of variance was concluded if the residuals were randomly and relatively evenly scattered around their mean (zero) value. Scatter plots displaying a geometric pattern (e.g. a line, a curve, a cloud, or a wedge shape) indicated that the variances varied systematically with respect to the predicted values, and non-homogeneity of variance was concluded.

The outcomes of ANOVA were the F (variance ratio) statistics and associated p (probability) values. The decision rule was that a predictor variable had a statistical relationship with a dependent variable if the p value of the computed F statistic was $\leq .05$. If the p value was $> .05$, then no effect was concluded. In addition to p values, Eta^2 values were computed as measures of effect size. Eta^2 indicates the proportion of the variance in the dependent variable that is statistically explained by the predictor variable(s).

Research Question #2

Is a student's performance on the MAP Reading able to predict his or her performance on the PASS ELA?

It was hypothesized that a student's performance on the MAP Reading does predict his/her performance on the PASS ELA. Prior studies investigating the alignment of cut scores between the spring administrations of the PACT (South Carolina's previous high-stakes test) and the MAP tests have suggested that the MAP tests are reliable and valid predictors of student performance on the PACT (Hauser, 2002; Cronin & Hauser, 2003; Cronin, 2004; Cronin, 2007).

In addition to analyzing the predictive validity of MAP Reading score in relation to the PASS ELA, the association between the three MAP Goal Performance areas (i.e., Understanding and Using Literary Texts, Understanding and Using Informational Texts, and Building Vocabulary) and PASS ELA were also calculated. It was hypothesized that there was a significant linear relationship between the three MAP Goal Performance areas. A large percentage of MAP items (i.e., 74% - 85%) are aligned to South Carolina grade level standards. In addition, all standards are included in the test item pool (South Carolina Department of Education, n.d.b). Furthermore, the PASS ELA assesses a student's mastery of the 2008 South Carolina Academic Standards in grades 3-8. At each grade level, four broad standards are assessed: Standard 1 – Literary Texts, Standard 2 – Informational Texts, Standard 3 – Vocabulary, and Standard 6 – Research. Depending on the student's grade level, 8-18 items per Standard are included on the PASS ELA (Table 4).

Pearson's correlation coefficients were used to measure the strengths of the linear relationships between normally distributed variables measured at the scale/interval level (i.e., PASS ELA, MAP Reading, MAP Understanding and Using Literary Texts, MAP Understanding and Using Informational Texts, and MAP Building Vocabulary). The decision rule was to reject the null hypothesis that significant relationships did not exist among the MAP variables if the p value of Pearson's r statistic was $\leq .05$. The nature of the causal mechanism underlying a significant zero order correlation between two variables may sometimes be the joint influence of a common cause, known as a control variable, that mediates between the two variables in question (Cohen et al., 2003). A correlation involving a third control variable, that jointly causes the correlation between two other variables, is termed partial correlation. In this study, partial correlation analysis was used to identify if there was an overlap in correlation between two variables due to the influence of a control variable. The influence of a control variable was indicated if the magnitudes of the partial correlation coefficients changed substantially relative to the zero-order coefficients.

Research Question #3

From among the demographic variables and MAP scores identified, what is/are the best predictor(s) of student performance on the PASS ELA?

It was hypothesized that demographic variables and the MAP scores are related to the overall PASS scores. Research has shown that students who are more familiar with computers perform better on computer-based tests in math and reading (Choi & Tinkler, 2002). Therefore, as the degree of exposure to CAT increases as a result of the student's advancement through school, the level of comfort with CAT should increase resulting in a more reliable and valid assessment of current instructional levels. Furthermore, previous research (Cronin, 2004) has shown that MAP Reading scores correctly predicted proficiency on previous South Carolina high-stakes testing in ELA for 80% to 82% of the cases. Multiple Linear Regression Analysis (MLR) was conducted to answer Research Question #3.

Multiple linear regression analysis. The aim of performing MLR was to construct mathematical models of the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_n X_n + \varepsilon$$

Where: Y = the predicted average value of the dependent variable (e.g., the MAP or the PASS ELA score); β_0 = the intercept (the theoretical predicted value of the dependent variable when all the predictor variables were zero); β = the regression coefficient for predictor variable X; n = the number of the predictor variables, ε = residual error. Partial regression coefficients can only be compared with each other if they are measured on the same scale. β weights (standardized regression coefficients) were used to numerically compare the relative importance of partial regression coefficients in this study when the predictor variables were measured on different scales. The β weight of a regression coefficient $\beta_i = S_{Xi} / S_Y$ where S_{Xi} = standard deviation of X_i, S_Y = standard deviation of Y (Cohen et al., 2003).

An assumption upon which MLR is based is that the error terms at each value of X are normally distributed and measured at the interval/scale level. The predictor variables were measured at the interval/scale level or were ordinal level variables coded numerically in rank order of magnitude (See Table 6). Nominal variables with binary categories were coded with 0 or 1 (e.g. male = 0 female = 1). Nominal categories coded with more than two categories (e.g., Race/Ethnicity) could not be coded with a series of numbers because they did not represent a numerical hierarchy (e.g. Black subjects coded with 2 and Other subjects coded with 3 were not two or three times higher than White subjects coded with 1). Consequently, the nominal variables such as race/ethnicity were decomposed and coded with dummy binary codes (using 1 to represent a subject within a category and 0 to represent a subject not within a category. The rule was followed that the number of binary codes must be one less than the total number of categories in the variable. The reason for this rule is that if all of the categories in a nominal variable with more than two groups were coded, then no unique estimators of the regression coefficients could be computed (Neter, Kutner, Wasserman, & Nachsteim, 1996).

A second assumption of MLR is homogeneity of variance, (i.e., that the variability of the error terms at any point on the regression line should be the same as at any other point on the regression line). If heteroscedasticity or non-homogeneity of variance is present, then the standard errors are biased, so that hypothesis tests and confidence intervals may be invalid (Cohen et al., 2003). Regression analysis also assumes that the residual error, (i.e., the differences between the predicted and observed values), should not deviate from normality. Tests for residual normality and homogeneity of variance were therefore performed similar to those described above for ANOVA.

An MLR model by definition also assumes linearity between the dependent variable and each of the predictor variables. Linearity implies that the average change in the dependent variable associated with a unit change in each predictor variable is constant. Bivariate scatterplots for each set of variables were examined to determine linearity. Partial correlation analysis was used to identify control variables.

The correlation and regression coefficients were recorded, together with the t and F statistics and their p values. The p value of the t statistic was used to test the null hypothesis that the regression coefficient was zero. The R^2 value was used to determine the effect size, (i.e., the proportion of the variability in the dependent variable explained by the variability in the predictor variable(s)). The p value of the F statistic was used to test the null hypothesis that R^2 was not significant. The tolerance level permitted for the significance of the Pearson's r, t and F statistics was chosen to be $p \leq .05$.

The predictor variables in MLR should not be collinear, (i.e., they must not be strongly correlated with each other). Collinearity induces undesirable changes in multiple regression statistics and must be identified and eliminated. Collinearity increases the values of the standard errors, which reduces the significance levels of the regression coefficients. Consequently, the regression coefficients of collinear predictor variables may not be statistically significant at the $p \leq .05$ level, even when they are linearly related to the dependent variable and even if the R^2 and F statistics are significant at $p \leq .05$ (Cohen et al., 2003). Collinearity was checked using Pearson's correlation coefficients and VIF (variance inflation factor) statistics. Pearson's correlation coefficients > 0.8 and a threshold VIF statistic of ≥ 3.3 were applied to indicate collinearity (Field, 2009). No collinear predictor variables were included in the models. If needed, inter-correlated variables were aggregated to construct non-collinear variables for the purposes of MLR. Table 7 provides a summary of the research questions, hypotheses, variables, statistical analyses, and statistical assumptions of the study.

Table 7

Research Questions, Hypotheses, Variables, Statistical Analyses, and Statistical Assumptions

Research Questions	Hypotheses	Variables	Statistical Analyses	Statistical Assumptions
1. What is the relationship between and among the demographic variables and student performance on the PASS ELA? How do these variables interrelate?	There is a significant difference in student performance between and among sub-categories of the various demographic variables on the PASS ELA.	Sex, Race/Ethnicity, SES, Special Education, Degree of Exposure to CAT, and PASS ELA score	ANOVA Cramer's V Statistic	1. Interval/scale data 2. Normality 3. Homogeneity of variance 1. Nominal/ordinal data
1.1. Is there a difference in student performance between males and females on the PASS ELA?	Females will score higher than their male counterparts.	Sex and PASS ELA score	ANOVA	1. Interval/scale data 2. Normality 3. Homogeneity of variance
1.2. Is there a difference in student performance between American Indian or Alaska Native, Asian or Other Pacific Islander, Black, Hispanic, and White students on the PASS ELA?	White and Asian/Pacific Islander students will score higher than their Black, Hispanic, and American Indian/Alaskan peers.	Race/Ethnicity and PASS ELA score	ANOVA	1. Interval/scale data 2. Normality 3. Homogeneity of variance
1.3. Is there a difference in student performance between students who receive free/reduced lunch and students who are not eligible for free/reduced lunch on the PASS ELA?	Students who are not eligible for free/reduced lunch will score higher than their free/reduced lunch counterparts.	Socioeconomic status and PASS ELA score	ANOVA	1. Interval/scale data 2. Normality 3. Homogeneity of variance
1.4. Is there a difference in student performance between disabled and nondisabled students on the PASS ELA?	Nondisabled students will score higher than their disabled counterparts.	Special Education and PASS ELA score	ANOVA	1. Interval/scale data 2. Normality 3. Homogeneity of variance

Table 7 (Continued)

Research Questions, Hypotheses, Variables, Statistical Analyses, and Statistical Assumptions

Research Questions	Hypotheses	Variables	Statistical Analyses	Statistical Assumptions
1.5. Is there a difference in student performance between students in grades 3-8 on the PASS ELA?	Students in elementary grades will score higher than students in middle school.	Degree of Exposure to CAT and PASS ELA score	ANOVA	1. Interval/scale data 2. Normality 3. Homogeneity of variance
2. Is a student's performance on the MAP Reading able to predict his or her performance on the PASS ELA?	A student's performance on the MAP Reading is able to predict his or her performance on the PASS ELA.	MAP Reading RIT Score and PASS ELA score	Pearson's Correlation	1. Interval/scale data 2. Normality 3. Linearity
2.1. Is there any association between the MAP Understanding and Using Literary Texts Goal Performance area and the PASS ELA?	There is an association between the MAP Understanding and Using Literary Texts Goal Performance area and the PASS ELA.	MAP Understanding and Using Literary Texts score and PASS ELA score	Pearson's Correlation	1. Interval/scale data 2. Normality 3. Linearity
2.2. Is there any association between the MAP Understanding and Using Informational Texts Goal Performance area and the PASS ELA?	There is an association between the MAP Understanding and Using Informational Texts Goal Performance area and the PASS ELA.	MAP Understanding and Using Information Texts score and PASS ELA score	Pearson's Correlation	1. Interval/scale data 2. Normality 3. Linearity
2.3. Is there any association between the MAP Building Vocabulary Goal Performance area and the PASS ELA?	There is an association between the MAP Building Vocabulary Goal Performance area and the PASS ELA.	MAP Building Vocabulary score and PASS ELA score	Pearson's Correlation	1. Interval/scale data 2. Normality 3. Linearity

Table 7 (Continued)

Research Questions, Hypotheses, Variables, Statistical Analyses, and Statistical Assumptions

Research Questions	Hypotheses	Variables	Statistical Analyses	Statistical Assumptions
3. From among the demographic variables and MAP scores identified, what is/are the best predictor(s) of student performance on the PASS ELA?	A student's degree of exposure to CAT and his/her MAP Reading score will be the strongest predictors of student performance on the PASS ELA.	Sex, Race/Ethnicity, SES, Special Education, Degree of Exposure to CAT, MAP Reading RIT score, MAP Understanding and Using Literary Texts score, Understanding and Using Information Texts score, MAP Building Vocabulary score and PASS ELA score	Multiple Linear Regression	1. Interval/scale data 2. Normality 3. Homogeneity of variance 4. Linearity 5. Non-collinearity

Summary

This study was designed to determine the link between demographic variables and ongoing, formative evaluation (i.e., MAP Reading) to the newly implemented high-stakes testing in South Carolina (i.e., PASS English Language Arts [ELA]). Establishing further support for this relationship could improve teachers' acceptability of the MAP for the purpose of clear goal setting with regards to student progress. Archival test data and demographic information were obtained from five elementary and three middle schools located in the target school district. Test data were included in the sample when a student record had both a Spring 2009 MAP reading score and a Spring 2009 PASS ELA score. All demographic information, MAP, and PASS test scores were exported from *TestView* (2004) into a Microsoft Excel spreadsheet. To ensure the anonymity of individual test data, all data were encoded by number with the student's name and student's ID having been removed by a third party prior to analysis by this investigator.

The relationships between five latent variables (i.e., sex, race/ethnicity, socioeconomic status, special education, and degree of exposure to computer adaptive testing) and student performance on the PASS ELA were examined as well as the validity of the MAP Reading to predict performance on the PASS ELA. Finally, from among the demographic variables and MAP scores analyzed, the aim was to identify the best predictors of student performance on the PASS ELA.

The variables were coded in the SPSS data editor, and statistical analyses were performed using SPSS following methods described by Field (2009). The data were analyzed using associational measures based on Cross-tabulation, Multi-factorial ANOVA, Pearson correlation, and Multiple Linear Regression leading to the construction of a hypothetical path model.

CHAPTER IV

RESULTS

This study investigated the predictive validity of demographic variables and the Measures of Academic Progress (MAP) in relation to student performance on the South Carolina's Palmetto Assessment of State Standards (PASS) English language arts (ELA) test. The state test was selected because it is the measure connected to high-stakes for school districts and it is viewed as a comprehensive measure of reading skill. Various predictive factors of student performance on the PASS ELA test were analyzed. Demographic factors included sex, race/ethnicity, socioeconomic status (SES), special education, and degree of exposure to computer adaptive testing (CAT). The specific predictive factors of the MAP that were analyzed included the MAP Reading Rasch Unit (RIT) score, the MAP Understanding and Using Literary Texts Goal Performance area, the MAP Understanding and Using Informational Texts Goal Performance area, and the MAP Building Vocabulary Goal Performance area. From among the demographic variables and MAP scores analyzed, the aim was to identify the best predictors of student performance on the PASS ELA.

Complications

There was a minor complication during this study. Due to the limited numbers of American Indian or Alaska Native, Asian or Other Pacific Islander, and Hispanic students included in the sample, the five categories of race/ethnicity initially identified at the beginning of the study were aggregated into three categories (i.e., Black, White, and Other) when conducting statistical analysis. The collapsing of the these three groups into one permitted sufficiently larger frequencies in the cells of cross-tabulations to compute Cramer's V and sufficiently larger sample sizes in the cells of a crossed design matrix to conduct multi-factorial analysis of variance (ANOVA). If the smallest race/ethnicity groups were not collapsed, some of the cells would include zero values or numbers less than 5, which might compromise the results of the statistical analyses. For purposes of Multiple Linear Regression Analysis (MLR), race/ethnicity was represented by two binary dummy variables (0 or 1 for White students and 0 or 1 for Black students). The Other race/ethnicity group was not included in the regression analysis. This was essential because the number of categories in a nominal predictor variable used in MLR must be one less than the total number of groups in the category otherwise no unique solution is possible (Tabachnik & Fidell, 2007). The logic is that since the three race/ethnicity groups are mutually exclusive, if a student is not White and not Black then he/she must be a member of the Other group. Furthermore, to improve the statistical and practical significance of the model as a predictor of the PASS ELA scores, MLR was repeated on a random sample of conditioned data with a sample size of $N = 2738$ from which multi-variate outliers, identified by their Mahalanobis distance values, were excluded.

Computer Program

The computer program used to analyze the data in this study was SPSS 16.0.

Sample

The sample was comprised of 3,861 students in grades 3-8 with available demographic data who had scores on the Spring 2009 MAP Reading and Spring 2009 PASS ELA assessments from the target school district's five elementary and three middle schools. Of the 3,861 students, 50.6% were male (n = 1,954) and 49.4% were female (n = 1,907). The racial breakdown of the sample included 82.5% White (n = 3,184), 10% Black (n = 388), and 7.5 % Other (n = 289). The Other category consisted of Hispanic (5.0%), Asian & Pacific Islander (2.5%), and American-Indian or Alaska Native (< 0.1%). With regards to SES, 81.7% of the sample did not qualify for free/reduced lunch (n = 3,156), 4% qualified for reduced lunch (n = 153), and 14.3% qualified for free lunch (n = 552). With regards to special education, 11.1% of the sample received some form of special education (n = 429) while 88.9% did not receive special education services (n = 3,432). Finally, exposure to CAT was measured as a function of grade placement. Within the sample, 17.2% of the students were in the 3rd grade (n = 665), 17.8% were in the 4th grade (n = 687), 15.7% were in the 5th grade (n = 605), 16.9% were in the 6th grade (n = 652), 16.7% were in the 7th grade (n = 646), and 15.7% were in the 8th grade (n = 606).

The frequencies in each demographic category are summarized in cross-tabulations in Appendix B. The largest representations in the sample consisted of White females (n = 1317 - 34.11%) and males (n = 1259 - 32.61%) who were not eligible for free/reduced lunch and in general education. Groups that represented less than 1% of the sample included: (1) Females from all race/ethnicity groups who were eligible for free/reduced lunch and in special education; (2) Females from all race/ethnicity groups who were eligible for reduced lunch and in general education; (3) Black and Other females not eligible for free/reduced lunch and in special education; (4) Black and Other males eligible for free lunch and in special education; (5) Black and Other males eligible for reduced lunch and in general education; (6) Males from all race/ethnicity groups who were eligible for reduced lunch in special education; and (7) Black and Other males not eligible for free/reduced lunch in special education.

Analyses

This study was designed to explore the possible links between ongoing, formative evaluation (i.e., MAP Reading) and the newly implemented high-stakes testing in South Carolina (i.e., PASS English Language Arts [ELA]). Archival test data and demographic information were obtained from five elementary and three middle schools located in the target school district. The relationships between five demographic variables (sex, race/ethnicity, SES, special education, and degree of exposure to CAT) and student performance on the PASS ELA were examined as well as ability of the MAP Reading scores to predict performance on the PASS ELA. From among the demographic variables and MAP scores analyzed, the aim was to identify the best predictors of student performance on the PASS ELA.

Research Question #1

What is the relationship between and among the demographic variables and student performance on the PASS ELA? How do these variables interrelate?

It was hypothesized that a difference in student performance existed between and among the demographic variables. To answer Research Question #1, the five demographic variables were partitioned into sub-questions.

1.1 Is there a difference in student performance between males and females on the PASS ELA?

It was hypothesized that a difference in student performance existed between males and females on the PASS ELA. Based on previous data from the National Assessment of Educational Progress (NAEP) and the target school district, females on average score higher in reading than their male counterparts (Lee, Grigg, & Donahue, 2007; National Center for Education Statistics [NCES], 2009a; Perie, Grigg, & Donahue, 2005; South Carolina Department of Education[SCDOE], n.d.c).

1.2 Is there a difference in student performance between American Indian or Alaska Native, Asian or Other Pacific Islander, Black, Hispanic, and White students on the PASS ELA?

It was hypothesized that a difference existed in student performance between White, Black, and Other students on the PASS ELA. Based on previous data from the NAEP and the target school district, White and Asian/Pacific Islander students on average score higher in reading than their Black, Hispanic, and American Indian/Alaskan peers (Lee et al., 2007; NCES, 2009a; Perie et al., 2005; SCDOE, n.d.c).

1.3 Is there a difference in student performance between students who receive free/reduced lunch and students who are not eligible for free/reduced lunch on the PASS ELA?

It was hypothesized that a difference existed in student performance between students who receive free/reduced lunch and students who are not eligible for free/reduced lunch on the PASS ELA. Based on previous data from NAEP and the target school district, students who are not eligible for free/reduced lunch on average score higher in reading than their free/reduced lunch counterparts (Lee et al., 2007; NCES, 2009a; Perie et al., 2005; SCDOE, n.d.c).

1.4 Is there a difference in student performance between disabled and nondisabled students on the PASS ELA?

It was hypothesized that a difference existed in student performance between disabled and nondisabled students on the PASS ELA. Based on previous data from NAEP and the target school district, students who were nondisabled on average score higher in reading than their disabled counterparts (Lee et al., 2007; NCES, 2009a; Perie et al., 2005; SCDOE, n.d.c).

1.5 Is there a difference in student performance between students in grades 3-8 on the PASS ELA?

It was hypothesized that a difference existed in student performance between students in grades 3-8 on the PASS ELA as many states set the bar significantly lower in elementary school than in middle school (Cronin, Dahlin, Adkins, & Kingsbury, 2007).

To measure the strengths of the associations between nominal and ordinal variables (i.e., sex, race/ethnicity, SES, special education, and degree of exposure to CAT), Cramer’s V statistics were computed by cross-tabulation analysis. Cramer's V is a correlation coefficient that indicates the relationship amongst two categorical variables. Like Pearson's coefficient, Cramer's V ranges from -1 to +1, with 0 indicating no relationship and -1 or +1 indicating a perfect relationship. Also like Pearson's coefficient, the square of Cramer's V indicates the proportion of the total possible association (i.e., the maximum possible value of the chi-square statistic) that is present in the data. The conventional interpretation of Cramer’s V coefficient was applied (i.e., < 0.1 = little, if any, association; 0.1 to 0.3 = weak or low association; 0.3 to 0.5 = moderate association; > 0.5 = high or strong association) (Field, 2009).

Multi-factorial analysis of variance (ANOVA) was applied to determine if the mean values of the dependent variable measured at the scale/interval level (i.e., PASS ELA scores) varied with respect to the effects of categorical predictor variables (i.e., sex, race/ethnicity, SES, special education, and degree of exposure to CAT).

Cramer's V coefficients. The matrix of Cramer's V coefficients (Table 8) provided evidence to conclude little or no association between most of the demographic variables. However, there was statistical evidence to indicate low levels of correlations between SES and race/ethnicity, indicated by Cramer's V = .285, and between SES and special education, indicated by Cramer's *V* = .185.

Table 8

Matrix of Cramer's V Coefficients to Indicate the Strengths of the Associations between Categorical Predictor Variables (N = 3861)

	Grade	Sex	Race/Ethnicity	SES
Sex	.037			
Race/Ethnicity	.048	.032		
SES	.037	.044	.285	
Special education	.042	.086	.063	.185

The highest value of Cramer's V reflected a difference between the White race/ethnicity group and the other two race/ethnicity groups with respect to SES. The observed frequencies of Black and Other race/ethnicity groups who received a free/reduced lunch were higher than expected. The observed frequencies of White students who received a free/reduced lunch were lower than expected. The implications were that a higher proportion of students in the Black and Other race/ethnicity groups were of lower SES (indicated by free/reduced lunch) than in the White race/ethnicity group (Table 9).

Table 9

Observed and Expected Frequencies in the Cross-Tabulation of Race/Ethnicity and SES

		Race/Ethnicity			
SES	Frequency	Black	White	Other	Total
Free lunch	Observed	189	264	99	552
	Expected	55.5	455.2	41.3	
Reduced lunch	Observed	35	100	18	153
	Expected	15.4	126.2	11.5	
No Free/Reduced Lunch	Observed	164	2820	172	3156
	Expected	317.2	2602.6	236.2	
Total	Observed	388	3184	289	3861

The observed frequencies of students in special education who received a free/reduced lunch were higher than expected while the observed frequencies of students in general (not special) education who received a free/reduced lunch were lower than expected. The implications were that a higher proportion of students in special education were of lower SES (indicated by free/reduced lunch) than those in general education (Table 10).

Table 10

Observed and Expected Frequencies in the Cross-Tabulation of SES and Special Education

		SES			
Special Education	Frequency	Free Lunch	Reduced Lunch	No Free/Reduced Lunch	Total
No	Observed	420	120	2892	3432
	Expected	490.7	136.0	2805.3	
Yes	Observed	132	33	264	429
	Expected	61.3	17.0	350.7	
Total	Observed	552	153	3156	3861

Descriptive statistics. The frequency distribution of the PASS ELA scores approximated a normal bell-shaped curve (Figure 5). The mean score = 653.3 (standard deviation = 49.7). The skewness statistic was .093, indicating that the frequency distribution was close to normality, and not skewed; nevertheless the distribution included outliers, observed in Figure 5 as extremely low or high scores, isolated in the left and right hand tails.

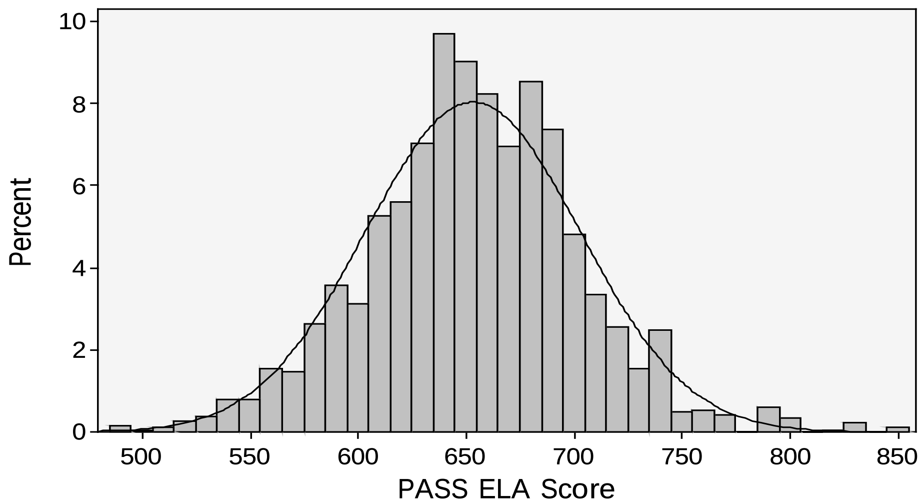

Figure 5. Frequency distribution of PASS ELA scores.

Visual examination of the frequency distributions histograms of the PASS ELA scores with respect to each of the demographic variables indicated that they also approximated normality. Deviations from normality were associated with the presence of outliers and not due to skewness. The minimum, maximum, and mean PASS ELA scores varied with respect to the demographic variables (Table 11). The mean PASS ELA score for females across the grades were usually higher than those for the males. The mean scores for White students were always higher across the grades than those of Black and Other students. The mean scores for students not eligible for free/reduced lunch were generally higher than those for students with free/reduced price lunches. The mean scores for students with no special education were in general greater than those of the special education students. Within each demographic group, however, the mean PASS ELA scores remained relatively consistent across the six grades (Table 11). The sample size, minimum, maximum, mean, and standard deviation for each demographic category are summarized in cross-tabulations in Appendix C.

Table 11

Descriptive Statistics for PASS ELA Scores

Variable	Sample size	Minimum	Maximum	Median	Mean	Standard Deviation
Female	1907	489	849	656	655.0	50.4
Male	1954	487	851	653	651.7	49.0
Total	3861	487	851	653	653.3	49.7
White	3184	489	851	654	655.3	49.7
Black	388	489	801	644	640.5	48.2
Other	289	487	767	649	648.1	49.2
Total	3861	487	851	653	653.3	49.7
Free lunch	552	487	832	639	638.2	48.8
Reduced lunch	153	497	774	641	640.7	52.7
No free/reduced lunch	3156	489	851	656	656.5	49.1
Total	3861	487	851	653	653.3	49.7
No special education	3432	487	851	654	654.7	49.3
Special education	429	489	832	644	642.3	51.8
Total	3861	487	851	653	653.3	49.7
Grade 3	665	487	832	660	656.3	48.4
Grade 4	687	523	832	653	653.3	50.0
Grade 5	605	497	849	653	654.7	48.5
Grade 6	652	489	832	652	651.2	50.2
Grade 7	646	492	849	653	652.1	52.6
Grade 8	606	489	851	653	652.1	48.3
Total	3861	487	851	653	653.3	49.7

Multi-factorial analysis of variance (ANOVA). The results of multi-factorial ANOVA (Table 12) indicated no significant interactions between the predictor variables at $\alpha = .05$. The higher order interactions are not included in Table 12 for this reason. The p values $\leq .05$ for the F statistics corresponding to sex ($p = .042$) and grade ($p = .021$) indicated that sex and grade had a statistically significant effect on the PASS ELA scores; however, the η^2 values between .001 and .004 reflected negligible effect sizes (Table 12). Since the effect sizes were so small, it is inferred that the statistically significant results at $p \leq .05$ have little scientific significance or practical implications. An investigation was performed to determine if the results of ANOVA were biased by the inclusion of outliers. Outliers were identified as cases with Mahalanobis distance statistics with p values < .001 (Tabachnik & Fidell, 2007). When ANOVA was performed excluding the outliers, the interpretation of the statistics was not different to that recorded above. The inclusion of outliers was therefore not considered to compromise the results of the ANOVA. Statistical evidence is provided to conclude that there was no significant variation of practical importance in the PASS ELA scores with respect to sex, race/ethnicity, SES, special education, or grade.

Table 12

Results of Multi-Factorial ANOVA on the Effects of Sex, Race/Ethnicity, SES, Special Education, and Grade on PASS ELA Scores (including outliers)

Source of variance	Type III Sum of Squares	Degrees of Freedom	Mean Square	F	p value	Effect size η2
MAIN EFFECTS:						
Sex	10019.6	1	10019.6	4.14	.042*	.001
Race/Ethnicity	7581.3	2	3790.7	1.57	.209	.001
SES	11838.9	2	5919.5	2.45	.087	.001
Special education	3958.3	1	3958.3	1.64	.201	<.001
Grade	32047.3	5	6409.5	2.65	.021*	.004
INTERACTIONS:						
Sex * Race/Ethnicity	5295.0	2	2647.5	1.09	.335	.001
Sex* SES	555.3	2	277.7	.12	.892	<.001
Sex * Special Education	1981.3	1	1981.3	.82	.366	<.001
Sex * Grade	16106.5	5	3221.3	1.33	.248	.002
Race/Ethnicity * SES	7683.2	4	1920.8	.79	.529	.001
Race/Ethnicity * Special Education	2225.4	2	1112.7	.46	.632	<.001
Race/Ethnicity * Grade	27497.6	10	2749.8	1.14	.331	.003
SES * Special education	1724.1	2	862.0	.36	.700	<.001
SES * Grade	30701.9	10	3070.2	1.27	.243	.003
Special education * Grade	23469.9	5	4694.0	1.94	.085	<.001

* Significant at α = .05

Research Question #2

Is a student's performance on the MAP Reading able to predict his or her performance on the PASS ELA?

It was hypothesized that a student's performance on the MAP Reading does predict his/her performance on the PASS ELA. Prior studies investigating the alignment of cut scores between the spring administrations of the PACT (South Carolina's previous high-stakes test) and the MAP tests have suggested that the MAP tests are reliable and valid predictors of student performance on the PACT (Hauser, 2002; Cronin & Hauser, 2003; Cronin, 2004; Cronin, 2007).

In addition to analyzing the predictive validity of MAP Reading score in relation to the PASS ELA, the association between the three MAP Goal Performance areas (i.e., Understanding and Using Literary Texts, Understanding and Using Informational Texts, and Building Vocabulary) and PASS ELA were also calculated. It was hypothesized that there was a significant linear relationship between the three MAP Goal Performance areas. A large percentage of MAP items (i.e., 74% - 85%) are aligned to South Carolina grade level standards. In addition, all standards are included in the test item pool (South Carolina Department of Education, n.d.b). Furthermore, the PASS ELA assesses a student's mastery of the 2008 South Carolina Academic Standards in grades 3-8. At each grade level, four broad standards are assessed: Standard 1 – Literary Texts, Standard 2 – Informational Texts, Standard 3 – Vocabulary, and Standard 6 – Research. Depending on the student's grade level, 8-18 items per Standard are included on the PASS ELA (Table 4).

Pearson's correlation coefficients. Pearson's correlation coefficients were used to measure the strengths of the inter-correlations between the scores for PASS ELA, MAP Reading, MAP Understanding and Using Literary Texts, MAP Understanding and Using Informational Texts, and MAP Building Vocabulary. The decision rule was to reject the null hypothesis that significant linear relationships did not exist among the MAP variables if the p value of Pearson's coefficient was $< .05$. The matrix of correlation coefficients with values ranging from .834 to .933 (Table 13) indicated that the MAP scores were strongly and significantly inter-correlated. The inter-correlations between the MAP scores and the PASS ELA scores were not, however, so strong as those between the MAP scores, indicated by Pearson's coefficients from .125 to .139 (Table 14). Although the correlations were statistically significant at the .05 level, the R^2 values (squares of the correlation coefficients) < 2% reflected very small effect sizes, implying that the linear relationships between the MAP and the PASS ELA scores have little practical importance.

Table 13

Matrix of Correlation Coefficients between MAP Scores

	MAP Reading	MAP Literary Texts	MAP Informational Texts
MAP Literary Texts	r = .929*		
MAP Informational Texts	r = .931*	r = .834*	
MAP Building Vocabulary	r = .933 *	r = .835*	r = .844*

*Significant at α = .05

Table 14

Matrix of Correlations Coefficients between MAP and PASS ELA Scores

	MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary
PASS ELA	r = .139*	r = .135*	r = .133*	r = .125*

*Significant at α = .05

Descriptive statistics. The frequency distributions of the MAP scores approximated normal bell-shaped curves (Figure 6); however, there were numerous outliers at the extreme left and right hand tails of the distributions, which could potentially interfere with statistical analysis.

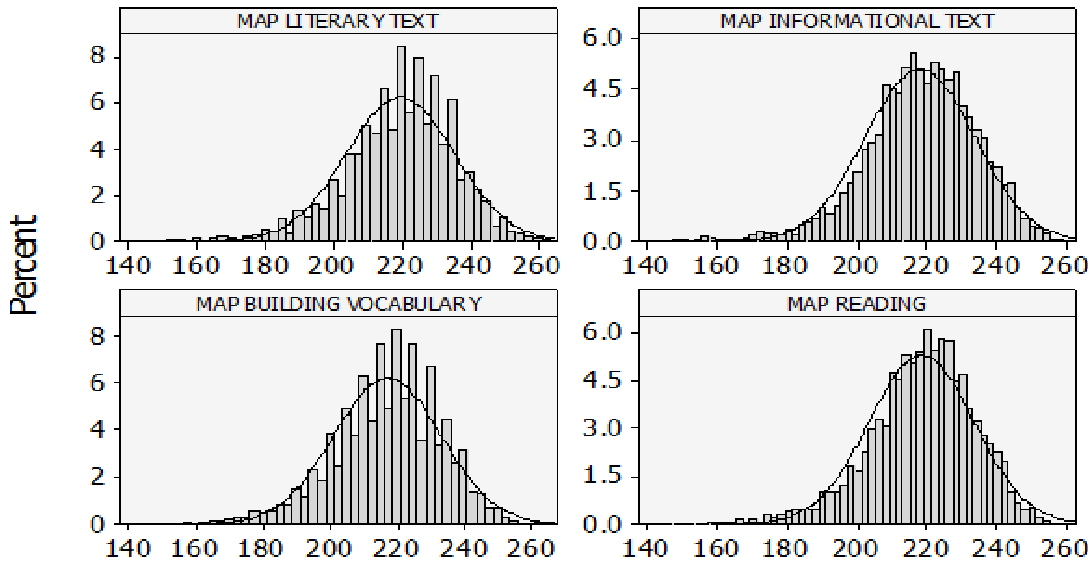

Figure 6. Frequency distributions of MAP scores.

The overall MAP Reading score varied with respect to sex, race/ethnicity, SES, special education, and grade (Table 15). The overall mean MAP Reading score for females was greater than that for males. The mean scores for the White and Other students were higher than the mean scores for the Black students. The mean score for students not eligible for free/reduced lunch was higher than the mean scores for students with free/reduced lunch. The mean score for students with no special education was greater than that of the special education students. A conspicuous feature of the overall mean MAP scores was that they increased systematically and progressively with respect to increasing exposure to CAT between grades 3-8 (Table 15). The sample size, minimum, maximum, mean, and standard deviation for each demographic category are summarized in cross-tabulations in Appendix C.

Table 15

Descriptive Statistics for MAP Scores

Variable	Sample Size	Minimum	Maximum	Median	Mean	Standard Deviation
Female	1907	155.0	254.3	220.8	219.6	14.4
Male	1954	147.3	253.5	218.3	216.9	15.4
Total	3861	147.3	254.3	219.8	218.3	15.0
White	3184	159.3	254.3	221.0	219.8	14.2
Black	388	147.3	248.3	210.3	209.7	16.1
Other	289	158.8	252.3	213.5	213.1	16.1
Total	3861	147.3	254.3	219.8	218.3	14.9
Free lunch	552	147.3	233.8	207.3	205.3	14.9
Reduced lunch	153	162.5	231.8	210.5	210.2	12.0
No free/reduced lunch	3156	162.0	254.3	221.8	220.9	13.7
Total	3861	147.3	254.3	219.8	218.3	15.0
No Special Education	3432	166.0	254.3	220.8	220.3	13.3
Special Education	429	147.3	240.0	205.0	201.9	17.5
Total	3861	147.3	254.3	219.8	218.3	15.0
3	665	149.8	233.0	206.3	205.2	13.0
4	687	154.0	243.5	214.0	212.1	13.0
5	605	147.3	253.8	219.0	217.4	13.0
6	652	167.0	249.0	222.0	220.9	13.0
7	646	170.0	252.8	226.8	225.4	12.0
8	606	178.0	254.3	230.8	230.0	11.9
Total	3861	147.3	254.3	219.8	218.3	15.0

Multi-factorial analysis of variance (ANOVA). The results of multi-factorial ANOVA (Table 16) indicated few significant interactions between the predictor variables at $\alpha = .05$. There were no higher order interactions. Although two-way interactions between race/ethnicity and SES, race/ethnicity and special education, and race/ethnicity and grade were indicated at $\alpha = .05$ by $p = .033$, .021, and .038 respectively, the effect sizes were negligible, indicating they had little scientific significance.

Table 16

Results of Multi-Factorial ANOVA on the Effects of Sex, Race/Ethnicity, SES, Special Education, and Grade on the MAP Reading Scores

Source	Type III Sum of Squares	Degrees of Freedom	Mean Square	F statistic	p value	Effect size η^2
MAIN EFFECTS:						
Sex	536.6	1	536.6	5.26	.022*	.001
Race/Ethnicity	3245.8	2	1622.9	15.91	<.001*	.009
SES	10429.5	2	5214.8	51.13	<.001*	.027
Special education	13740.4	1	13740. 4	134.73	<.001*	.035
Grade	27301.4	5	5460. 4	53.54	<.001*	.068
INTERACTIONS:						
Sex X Race/Ethnicity	591.2	2	295.5	2.89	.055	.002
Sex X SES	241.0	2	120.5	1.18	.307	.001
Sex X Special education	65.1	1	65.2	.64	.424	<.001
Sex X Grade	791.6	5	158.3	1.56	.170	.002
Race/Ethnicity X SES	1072.4	4	268.2	2.63	.033*	.003
Race/Ethnicity X Special education	789.0	2	394.6	3.87	.021*	.002
Race/Ethnicity X Grade	1960.5	10	196.1	1.92	.038*	.005
SES X Special education	36.2	2	18.1	.18	.837	<.001
SES X Grade	1342.7	10	134.4	1.32	.215	.004
Special education X Grade	2290.9	5	458.2	4.49	<.001*	.006

* Significant at $\alpha = .05$

The p values < .05 for the F statistics (Table 16) indicated that sex (p = .022), race/ethnicity (p = <.001), SES (p = <.001), special education (p = <.001), and grade (p = < .001) all had statistically significant effects on the MAP scores; however the η^2 values for sex, race/ethnicity, SES, special education, and grade < .1 reflected very small effect sizes (Table 16). It is inferred that the statistically significant differences between the MAP scores with respect to sex, race/ethnicity, SES, special education, and grade had little scientific significance or practical implications. When the outliers were excluded and the analysis repeated, the interpretation was the same, implying that the inclusion of outliers did not compromise the results of ANOVA.

Statistical evidence is provided to conclude that there was no significant variation of practical importance in the MAP scores with respect to sex, race/ethnicity, SES, special education, or grade.

Research Question #3

From among the demographic variables and MAP scores identified, what is/are the best predictor(s) of student performance on the PASS ELA?

It was hypothesized that demographic variables and the PASS ELA scores are related to the overall mean MAP scores. Research has shown that students who are more familiar with computers perform better on computer-based tests in math and reading (Choi & Tinkler, 2002). Therefore, as the degree of exposure to CAT increases as a result of the student's advancement through school, the level of comfort with CAT should increase resulting in a more reliable and valid assessment of current instructional levels. Furthermore, previous research (Cronin, 2004) has shown that MAP Reading scores correctly predicted proficiency on previous South Carolina high-stakes testing in ELA for 80% to 82% of the cases. MLR was conducted to answer Research Question #3.

Multiple linear regression analysis (MLR). The standardized regression coefficients computed by MLR are presented in Table 17. ANOVA indicated that the overall model was statistically significant, F (11, 3848) = 14.659 p = < .001; however, the results of MLR were severely compromised. The R^2 value = 3.7% was a very small effect size, indicative of a model with little or no practical importance. Furthermore, the high standard error = 48.7 indicated that the model was not precise and had limited predictive ability.

Table 17

Standardized Regression Coefficients to Predict the PASS ELA Scores

Variables	Standardize Coefficient	t statistic	p value
Intercept		24.42	<.001*
Sex	-.01	-.90	.369
White (Race/Ethnicity)	-.06	-.49	.623
Black (Race/Ethnicity)	-.08	-.78	.433
Hispanic (Race/Ethnicity)	-.04	-.56	.574
Asian/Pacific Islander (Race/Ethnicity)	-.04	-.79	.429
SES	.06	3.23	.001*
Special education	.00	.06	.953
Grade	-.12	-6.27	<.001*
MAP Literary Texts	.05	1.68	.093
MAP Informational Texts	.07	2.23	.026*
MAP Building Vocabulary	.06	1.76	.079

* Significant at α .05

Although SES and grade were significant predictors of the PASS ELA scores at $\alpha = .05$ (Table 17), the MAP Literary Texts and MAP Building Vocabulary scores, which were both significantly correlated with the PASS ELA scores (Table 14) were not, as might be expected, declared as significant predictors. The reason for this anomaly was that the model was deleteriously influenced by partial correlations and multi-collinearity (Table 18). The partial correlation coefficients for race/ethnicity, SES, special education, and the MAP scores were reduced almost to zero relative to the zero-order coefficients, indicating that the correlations were confounded. The VIF statistics between 4.2 and 70.1, corresponding to the binary dummy variables representing race/ethnicity, and the three MAP scores, reflected collinearity which violated the assumption of MLR. Statistical evidence is provided to indicate that the model defined in Table 17 was not a practically significant predictor of the PASS ELA scores. The construction of a path model based on the standardized regression coefficients in Table 17 was not considered to be justifiable.

Table 18

Correlation and Collinearity Statistics for the Model to Predict the PASS ELA Scores

Variables	Correlation		Collinearity
	Zero-order	Partial	VIF
Sex	-.03	-.01	1.0
White (Race/Ethnicity)	.09	-.01	70.1
Black (Race/Ethnicity)	-.09	-.01	44.9
Hispanic (Race/Ethnicity)	-.03	-.01	24.0
Asian/Pacific Islander (Race/Ethnicity)	-.01	-.01	11.7
SES	.14	.05	1.4
Special education	-.08	-.01	1.3
Grade	-.03	-.10	1.7
MAP Literary Texts	.14	.03	4.2
MAP Informational Texts	.13	.04	4.5
MAP Building Vocabulary	.13	.03	4.7

In view of the very small effect size, and the statistical violations of MLR discussed above, it was deemed desirable to improve the statistical and practical significance of the model as a predictor of the PASS ELA scores. MLR was repeated on a sample of conditioned data with a sample size of N = 2738 from which multi-variate outliers, identified by their Mahalanobis distance values, were excluded. To eliminate the strong multi-collinearity between the PASS ELA scores of the race/ethnicity groups, the dummy variable representing Race/Ethnicity was collapsed into two binary dummy variables specifically White (1 = White; 0 = Not white) and Black (1 = Black, 0 = Not Black). To represent an association between SES and Race/Ethnicity, SES (1 = free lunch, 2 = reduced lunch, or 3 = no free/reduced lunch) was multiplied by Race/Ethnicity to construct a new composite ordinal variable named SES X White, where 0 = Non-White (the reference baseline); 1 = White with free lunch; 2 = White with reduced lunch, and 3 = White with no free/reduced lunch. To represent an association between Special Education and Race/Ethnicity, Special Education (0 = No special education; 1 = with special education) was multiplied by Race/Ethnicity (0 = Not Black, 1 = Black) to construct a new composite ordinal variable named Special Education X Black, where 0 = No special education (the reference baseline) 1 = Black with special education.

When the three MAP scores were entered into the MLR analysis as separate independent variables, the respective R^2 values were 12.3% (MAP Building Vocabulary) 13.5% (MAP Literary Texts) and 13.6% (MAP Informational Texts). When the MAP Reading score was entered into the analysis, as a single substitute for the three separate MAP standard scores, the R^2 value increased to 15.3%. When the three MAP scores were entered into the model together, the R^2 value was also 15.3% and the three MAP scores were all significant predictors of the PASS ELA scores at the .05 level. The VIF statistics of 3.6 to 4.1 for the three MAP scores reflected lower levels of collinearity between the scores than recorded in the model based on unconditioned data (Table 20). It was decided to retain the three standard MAP scores as separate independent variables, so that their relative individual importance as predictors of the PASS ELA scores could be defined in the path model.

ANOVA indicated that the overall model including the three MAP scores was statistically significant, $F (7, 2724) = 71.35$ $p = <.001$. However, the R^2 =15.3% was a small effect size, indicative of a model with limited practical significance. In addition, the high standard error = 35.5 indicated that the estimate was not precise, so the model had limited mathematical ability to accurately predict the PASS ELA scores.

Grade, Sex, and SES X White were also significant predictors of the PASS ELA scores at the .05 level. The only variable that was not a significant predictor of the PASS ELA scores at $\alpha = .05$ was Special Education X Black which was significant at $p = .121$ (Table 19). Nevertheless, Special Education X Black was retained in the model since it was considered to contribute towards an understanding of the relationship between race/ethnicity, special education, and the PASS ELA scores, more than Special Education could do if entered into the model alone. The partial correlations did not generally decline to zero relative to the zero-order coefficients, apart for the partials for the MAP scores, which reflected their inter-correlation. The residuals were normally distributed (Figure 7) and relatively evenly distributed on either side of their mean (zero) value (Figure 8). Evidence is provided to conclude that the assumptions of MLR were not seriously violated. However, even based on the conditioned data, the relationships between the variables had limited practical significance.

Table 19

Regression Coefficients for the Model to Predict Conditioned PASS ELA Scores

Variables	Standardized regression coefficients	t statistic	p value
Intercept	422.80	34.06	<.001*
MAP Informational Texts	.17	5.00	<.001*
MAP Literary Texts	.16	4.67	<.001*
MAP Building Vocabulary	.12	3.30	.001*
Grade	-.14	-6.31	<.001*
Sex	-.04	-2.18	.029*
SES X White	.04	2.11	.035*
Special Education X Black	-.03	-1.55	.121

* Significant at $\alpha = .05$

Table 20

Correlation and Collinearity Statistics for the Model to Predict Conditioned PASS ELA Scores

Variables	Correlations		Collinearity
	Zero-order	Partial	VIF
MAP Informational Texts	.35	.10	3.7
MAP Literary Texts	.35	.09	3.6
MAP Building Vocabulary	.33	.06	4.1
Grade	.08	-.12	1.5
Sex	-.09	-.04	1.0
SES X White	.17	.14	1.2
Special education X Black	-.12	-.13	1.0

Path model. A path model (Figure 9) was formulated based on the MLR equation:

$$Y = 422.8 + .171\, X_1 + .157\, X_2 + .118\, X_3 - .137\, X_4 + .039\, X_5 + .040\, X_6 - .029\, X_7$$

Where: X_1 = MAP Informational Texts score; X_2 = MAP Literary Texts score; X_3 = MAP Building Vocabulary score; X_4 = Grade (3, 4, 5, 6, 7, or 8); X_5 = Sex (0 = Female, 1 = Male); X_6 = Race/Ethnicity x SES (0 = Non-White, 1 = White with free lunch; 2 = White with reduced lunch, and 3 = White with no free/reduced lunch); and X_7 = Race/Ethnicity X Special Education (0 = Non-Black; 1 = Black with free lunch; 2 = Black with reduced lunch, and 3 = Black with no free/reduced lunch).

The magnitudes of the standardized partial regression coefficients reflected the relative importance of the variables as predictors of the PASS ELA scores. The interpretation of the MLR model is that the PASS ELA scores:

- Increased by .171 for every unit increase in the MAP Informational Texts score.
- Increased by .157 for every unit increase in the MAP Literary Texts score.
- Increased by .118 for every unit increase in the MAP Building Vocabulary score.
- Decreased by -.137 for every unit increase in Grade from 3 to 8.
- Decreased by -.039 x 1 when the student was male (Sex = 1) relative to a female (Sex = 0).
- Increased by .04 x 1 for a White student with free lunch (White = 1 and SES = 1) relative to a Non-White (White = 0).
- Increased by .04 x 2 for a White student with reduced lunch (White = 1 and SES = 2), relative to a Non-White (White = 0).
- Increased by .04 x 3 for a White student with no free/reduced lunch (White = 1 and SES = 3) relative to a Non-White (White = 0).
- Decreased by -.029 x 1 for a Black special education student (Black = 1 and Special Education = 1), relative to a student without special education (Special Education = 0).

Figure 9 illustrates the Path model, including all the potential predictors of student performance on the PASS ELA, using standardized regression coefficients to denote their relative importance.

Evidence is provided to conclude that the three standard MAP scores and the grade had the most important direct effects on the variance in the PASS ELA scores. The effects of sex, SES, special education, and race/ethnicity, were relatively less important.

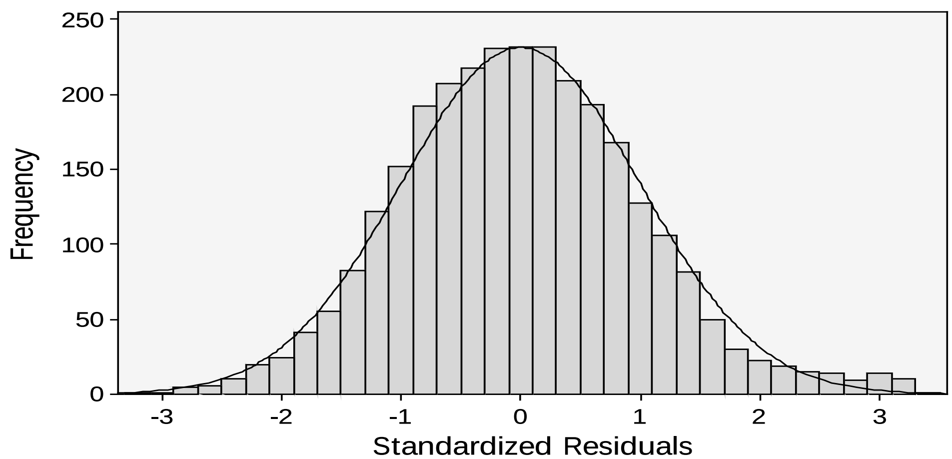

Figure 7. Frequency distribution of standardized residuals for the model to predict conditioned PASS ELA scores.

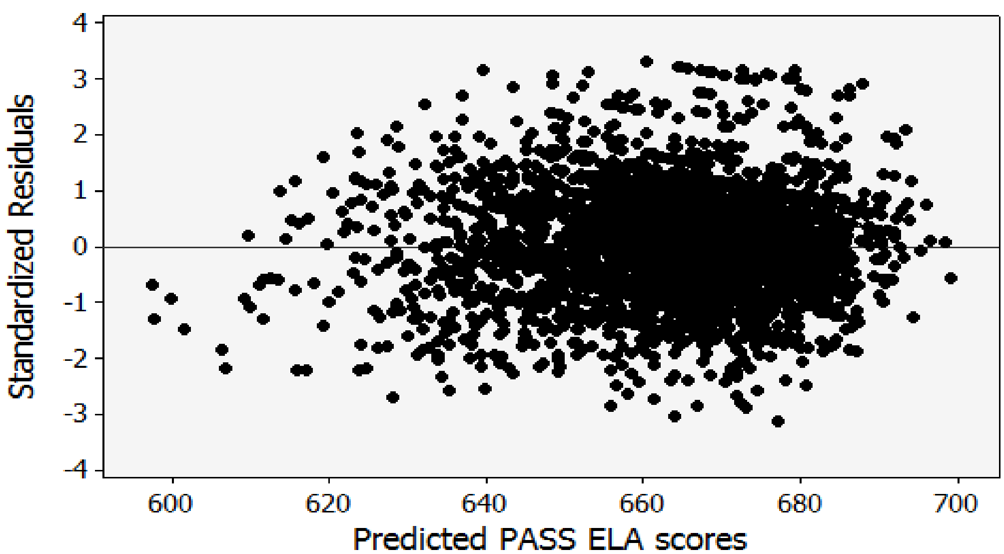

Figure 8. Plot of residuals versus predicted conditioned PASS ELA scores.

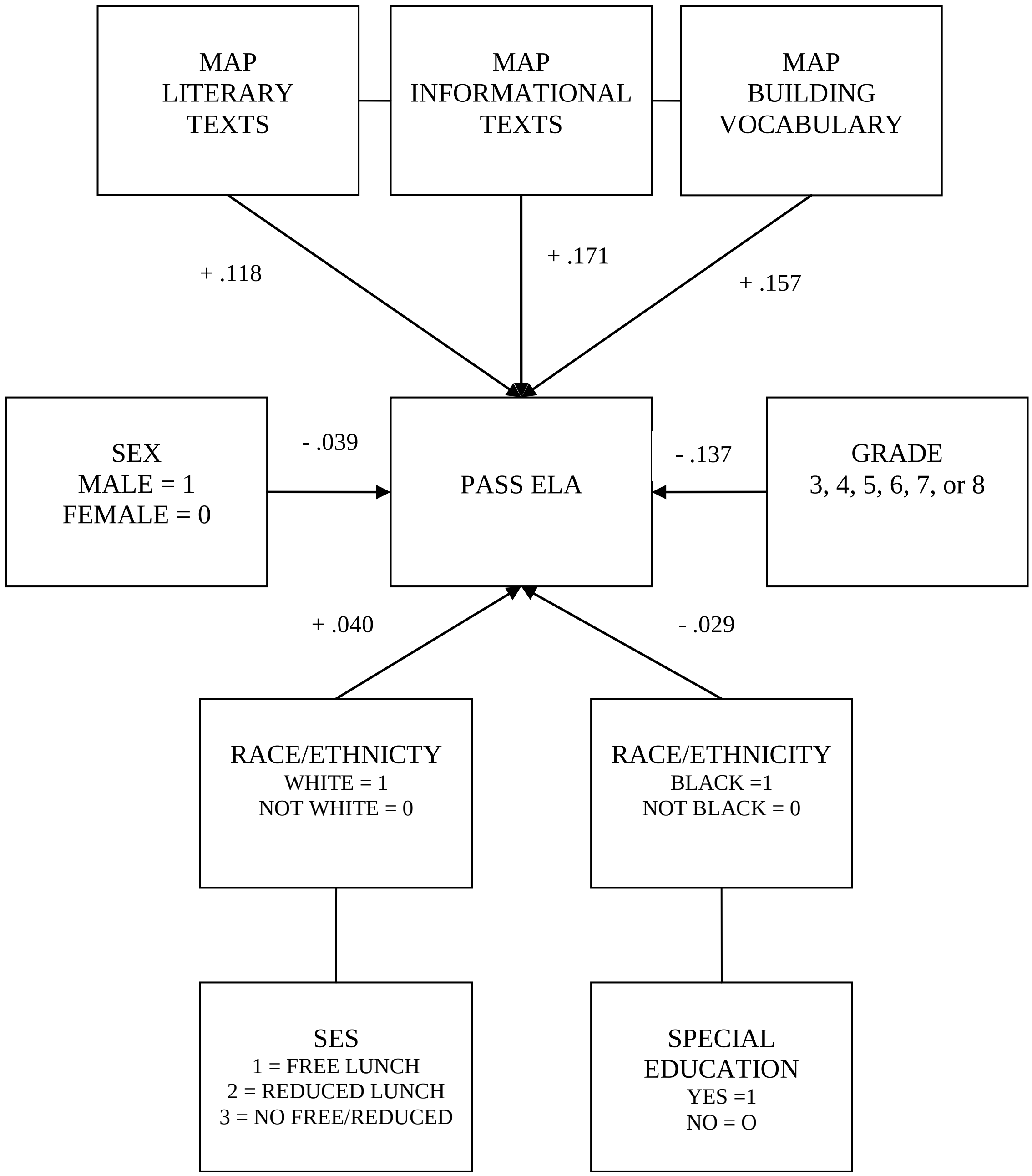

Figure 9. Path model representing the relationships between the predictors of student performance on the PASS ELA.

Summary

This study investigated the predictive validity of demographic variables and the Measures of Academic Progress (MAP) in relation to student performance on the South Carolina's Palmetto Assessment of State Standards (PASS) English language arts (ELA) test. Various predictive factors of student performance on the PASS ELA test were analyzed. Demographic factors included sex, race/ethnicity, SES, special education, and degree of exposure to CAT. The specific predictive factors of the MAP that were analyzed included the MAP Reading RIT score, the MAP Understanding and Using Literary Texts Goal Performance area, the MAP Understanding and Using Informational Texts Goal Performance area, and the MAP Building Vocabulary Goal Performance area. From among the demographic variables and MAP scores analyzed, the main conclusions of the statistical analysis are that:

The matrix of Cramer's V coefficients provided evidence to conclude little or no association between most of the demographic variables. However, there was statistical evidence to indicate low levels of correlations between SES and race/ethnicity and between SES and special education. A higher proportion of students in the Black and Other race/ethnicity groups were of lower SES than in the White race/ethnicity group while a higher proportion of students in special education were of lower SES than those in general education.

Based on the very small effect sizes, it is concluded that there were no relationships of practical significance between sex, race/ethnicity, SES special education, or grades of the students and the PASS ELA scores.

There were significant correlations between the MAP Reading, the MAP Understanding and Using Literary Texts, the MAP Understanding and Using Informational Texts, and the MAP Building Vocabulary. However, the correlations between the MAP scores and the PASS ELA scores were not as strong as those between the MAP scores. Although the correlations were statistically significant at the .05 level, the R^2 values $< 2\%$ reflected very small effect sizes, implying that the linear relationships between the MAP and the PASS ELA scores have little practical importance.

In view of the very small effect size, as well as statistical violations of MLR, it was necessary to improve the statistical and practical significance of the model as a predictor of the PASS ELA scores. MLR was repeated on a random sample of conditioned data with a sample size of $N = 2738$ from which multi-variate outliers, identified by their Mahalanobis distance values, were excluded. ANOVA indicated that the overall model including the three MAP scores was statistically significant. However, the small effect size was indicative of a model with little to no practical significance. In addition, the high standard error indicated that the estimate was not precise, so the model had limited mathematical ability to accurately predict the PASS ELA scores.

CHAPTER V

DISCUSSION

The ability to measure school effectiveness has become a major issue in educational policy and legislation. The most recent attempt at such legislation occurred in January of 2002 when then President Bush signed the No Child Left Behind (NCLB, 2001) act into law. NCLB set forth a broad spectrum of changes to the federal role in public education, including accountability provisions that mandated states to test all students. Scores from these tests are used for a variety of purposes, from identifying whether individual students are proficient with respect to designated academic standards, to helping determine whether schools are achieving adequate yearly progress (AYP). Under NCLB, schools consistently not meeting AYP may be placed under sanctions requiring additional educational opportunities for students. If a continuing lack of progress is evident, proposed sanctions include allowing students to transfer to schools with better performance and, eventually, the closing of schools that fail to meet standards. In an atmosphere of educational reform and federally mandated high-stakes testing, demands have increased for progress monitoring strategies that reliably predict outcomes on statewide assessments (McGlinchey & Hixson, 2004). This study investigated the predictive validity of demographic variables and the Measures of Academic Progress (MAP) in relation to student performance on the South Carolina's Palmetto Assessment of State Standards (PASS) English language arts (ELA) test. The state test was selected because it is the measure connected to high-stakes for school districts and it is viewed as a comprehensive measure of reading skill.

This discussion will summarize the procedures of the study as well as review the data analyses and findings for each research question and hypothesis. Possible reasons for these findings and implications for educators are offered. As with all research, a number of factors can limit the validity of a study. Cautions regarding the interpretation of this study's results will be identified and discussed. Finally, recommendations for further research will be presented.

A Review of Procedures and Analyses

During the Spring of the 2008-2009 school year, students in grades 3-8 in the target school district participated in the MAP Reading during the designated test term window to assess the current instructional level of each student and measure growth over time. Beginning with the 2008-2009 school year, students in the target district were also administered a new assessment in grades 3-8 known as the Palmetto Assessment of State Standards (PASS). The PASS is used for statewide accountability testing purposes and includes tests in writing, ELA (reading and research), mathematics, science, and social studies. The students were administered the ELA (reading and research) portion of the PASS on May 12, 2009.

Archival test data and demographic information were obtained from five elementary and three middle schools located in the target school district. The sample was comprised of 3,861 students in grades 3-8 with available demographic data who had scores on the Spring 2009 MAP Reading and Spring 2009 PASS ELA. Of the 3,861 students, 50.6% were male (n = 1,954) and 49.4% were female (n = 1,907). The racial breakdown of the sample included 82.5% White (n = 3,184), 10% Black (n = 388), and 7.5 % Other (n = 289). The Other category consisted of Hispanic (5.0%), Asian & Pacific Islander (2.5%), and American-Indian or Alaska Native (< 0.1%). With regards to socioeconomic status (SES), 81.7% of the sample did not qualify for free and reduced lunch (n = 3,156), 4% qualified for reduced lunch (n = 153), and 14.3% qualified for free lunch. (n = 552). With regards to special education, 11.1% of the sample received some form of special education (n = 429) while 88.9% did not receive special education services (n = 3,432). Finally, exposure to computer adaptive testing (CAT) was measured as a function of grade placement. Within the sample, 17.2% of the students were in the 3rd grade (n = 665), 17.8% were in the 4th grade (n = 687), 15.7% were in the 5th grade (n = 605), 16.9% were in the 6th grade (n = 652), 16.7% were in the 7th grade (n = 646), and 15.7% were in the 8th grade (n = 606).

The relationships between and among five latent variables (i.e., sex, race/ethnicity, SES, special education, and degree of exposure to CAT) and student performance on the PASS ELA were examined as well as the validity of the MAP Reading to predict performance on the PASS ELA. Finally, from among the demographic variables and MAP scores analyzed, the aim was to identify the best predictors of student performance on the PASS ELA. The data were analyzed using associational measures based on Cross-tabulation, Multi-factorial Analysis of Variance (ANOVA), Pearson correlation, and Multiple Linear Regression (MLR) leading to the construction of a hypothetical path model.

A Review of Results

This study investigated the predictive validity of demographic variables and the MAP in relation to student performance on the South Carolina's PASS ELA test. Various predictive factors of student performance on the PASS ELA test were analyzed. Demographic factors included sex, race/ethnicity, SES, special education, and degree of exposure to computer adaptive testing (CAT). The specific predictive factors of the MAP that were analyzed included the MAP Reading Rasch Unit (RIT) score, the MAP Understanding and Using Literary Texts Goal Performance area, the MAP Understanding and Using Informational Texts Goal Performance area, and the MAP Building Vocabulary Goal Performance area. From among the demographic variables and MAP scores analyzed, the aim was to identify the best predictors of student performance on the PASS ELA. In order to accomplish this, the following research questions were explored:

Research Question #1

What is the relationship between and among the demographic variables and student performance on the PASS ELA? How do these variables interrelate?

1.1 Is there a difference in student performance between males and females on the PASS ELA?

1.2 Is there a difference in student performance between American Indian or Alaska Native, Asian or Other Pacific Islander, Black, Hispanic, and White students on the PASS ELA?

1.3 Is there a difference in student performance between students who receive free/reduced lunch and students who are not eligible for free/reduced lunch on the PASS ELA?

1.4 Is there a difference in student performance between disabled and nondisabled students on the PASS ELA?

1.5 Is there a difference in student performance between students in grades 3-8 on the PASS ELA?

To measure the strengths of the associations between categorical predictor variables (i.e., sex, race/ethnicity, SES, special education, and degree of exposure to CAT), Cramer's V statistics were computed by cross-tabulation analysis. A few statistical associations were found between the categorical predictor variables specifically SES and race/ethnicity and SES and special education. A higher proportion of Black and Other students were of lower SES (indicated by free/reduced lunch) than White students and a higher proportion of students in special education were of lower SES than those in general education. However, the strengths of these associations were weak, indicated by Cramer's V coefficients < 0.3. Furthermore, the results of multi-factorial ANOVA indicated no significant interactions between the categorical predictor variables.

Similar to previous test data from National Assessment of Educational Progress ([NAEP] Lee, Grigg, & Donahue, 2007; National Center for Education Statistics [NCES], 2009; Perie, Grigg, & Donahue, 2005), females on average scored higher than males on the PASS ELA. The mean score for White students was higher than that for Black and Other students. The mean score for students not eligible for free/reduced lunch was higher than that for students with free/reduced lunches. The mean score for students with no special education was greater than that of the special education students. Finally, the PASS ELA scores remained relatively consistent across the six grades. While multi-factorial ANOVA confirmed the statistically significant effects sex and grade had on PASS ELA scores, the effect sizes were so small, it was inferred that the statistically significant results have little scientific significance or practical implications. Thus, with regards to Research Question #1, statistical evidence was provided to conclude that there was no significant variation of practical importance in the PASS ELA scores with respect to sex, race/ethnicity, socioeconomic status, special education, or degree of exposure to computer adaptive testing (CAT).

Statistical significance is a tool which can be used in quantitative research, but it is not necessarily the most important consideration in evaluating research results. In statistical analysis, a result is called statistically significant if it is unlikely to have occurred by chance. However, statistical significance does not evaluate whether results are important (Thompson, 2002). Effect size is an indicator of the strength and/or the importance of results. Effect sizes are especially important because statistical tests are so heavily influenced by sample sizes (Thompson, 2002). According to Hays (1981), "virtually any study can be made to show significant results if one uses enough subjects" (p. 23). Thus, for the purposes of this research, the results of the statistical tests were interpreted assuming that statistical significance and scientific significance were not equivalent. Some statisticians assert that null hypothesis significance tests are meaningless and should be banned in educational research (Hunter, 1997; Kline, 2004). They argue that dichotomous decisions based upon whether or not the p value of a test statistic is less than or greater than a pre-determined arbitrary level such as $\alpha = 05$ may indicate whether the observed data deviate from that which might be expected by chance. However, they provide no useful information whatsoever about the scientific significance, practical implications, and meaningfulness of observed data in reality, particularly if the sample size is large. A survey of American Educational Research Association (AERA) members indicated that 19% agreed (Mittag & Thomson, 2000). However, the APA Task Force on Statistical Inference did not support such radical views, but instead recommended that researchers should always provide effect size estimates when reporting p values for null hypothesis significance tests (Wilkinson & Task Force on Statistical Inference, 1999). This study supported the APA recommendation, so the effect sizes, indicating the proportion(s) of the variance in the dependent variable explained by the variance in the predictor variable(s) were computed and interpreted. High scientific significance was accredited to results with a large effect size, and low scientific significance was accredited to small effect sizes, irrespective of the magnitudes of the p values. Therefore, the results of this study do not support the hypothesis that a difference in student performance on the PASS ELA existed between and among the demographic variables.

Research Question #2

Is a student's performance on the MAP Reading able to predict his or her performance on the PASS ELA?

2.1. Is there any association between the MAP Understanding and Using Literary Texts Goal Performance area and the PASS ELA?

2.2. Is there any association between the MAP Understanding and Using Informational Texts Goal Performance area and the PASS ELA?

2.3. Is there any association between the MAP Building Vocabulary Goal Performance area and the PASS ELA?

A correlation analysis to determine the relationships between the MAP scores and the PASS ELA scores was performed. The Pearson's correlation coefficients were significant at $p < .001$, but statistical significance was merely a reflection of the large sample size. The very small effect size (i.e., .019) indicates that the statistically significant relationship between the MAP scores and the PASS ELA scores has little to no practical significance. Thus, current findings do not support the hypothesis that a student's performance on the MAP Reading is able to predict his or her performance on the PASS ELA.

Pearson's correlation coefficients were also used to measure the strengths of the linear relationships between the MAP Reading RIT score and the MAP Goal Performance areas. Pearson's r correlation coefficients ranged from 0.834 to 0.933 indicating that the scores for MAP Reading, the MAP Understanding and Using Literary Texts, the MAP Understanding and Using Informational Texts, and the MAP Building Vocabulary were very highly inter-correlated with each other.

Multi-factorial ANOVA was conducted to determine the relationship between and among the demographic variables and student performance on the MAP Reading. Similar to PASS ELA, the overall mean MAP score varied with respect to sex, race/ethnicity, socioeconomic status, special education, and grade. The overall mean MAP score for females was greater than that for males. The mean scores for White and Other students were higher than the mean scores for Black students. The mean score for students not eligible for free/reduced lunch was higher than the mean scores for students with free/reduced lunch. The mean score for students with no special education was greater than that of the special education students. A conspicuous feature of the overall mean MAP scores was that they increased systematically and progressively with respect to increasing exposure to CAT between grades 3-8. While multi-factorial ANOVA confirmed the statistically significant effects sex, race/ethnicity, SES, special education and grade had on MAP scores, the effect sizes were so small, it was inferred that the statistically significant results have little scientific significance or practical implications.

Research Question #3

From among the demographic variables and MAP scores identified, what is/are the best predictor(s) of student performance on the PASS ELA?

It was hypothesized that from among the demographic variables and MAP scores identified, that a student's degree of exposure to CAT and a student's MAP Reading score would be the strongest predictors of student performance on the PASS ELA. As previously discussed in Research Question #1, it was demonstrated using ANOVA that due to the very small effect sizes, that no relationship of practical significance existed between sex, race/ethnicity, SES, special education, or degree of exposure to CAT and the PASS ELA scores. Furthermore, correlation analysis conducted to determine the relationship between the MAP scores and the PASS ELA scores indicated that while the Pearson's correlation coefficient was significant at $p < .001$ that the statistical significance was merely a reflection of the large sample size. There was no direct relationship of practical significance between the MAP scores and the PASS ELA scores.

Multiple Linear Regression (MLR) analysis was performed to predict the mean overall PASS scores using sex, race/ethnicity, SES, special education, and grade as well as the three MAP Goal Performance Area scores (i.e., MAP Understanding and Using Literary Texts, the MAP Understanding and Using Informational Texts, and the MAP Building Vocabulary). Initial MLR analysis indicated that the overall model was statistically significant; however, the results of MLR were severely compromised. Although SES and grade were significant predictors of the PASS ELA scores, the MAP Literary Texts and MAP Building Vocabulary scores, which were both significantly correlated with the PASS ELA scores were not, as might be expected, declared as significant predictors. The reason for this anomaly was that the model was deleteriously influenced by partial correlations and multi-collinearity which violated the assumption of MLR.

In view of the very small effect size, and the statistical violations of MLR, it was deemed desirable to improve the statistical and practical significance of the model as a predictor of the PASS ELA scores. MLR was repeated on a random sample of conditioned data with a sample size of $N = 2738$ from which multi-variate outliers, identified by their Mahalanobis distance values, were excluded. In addition, to eliminate the strong multi-collinearity between the PASS ELA scores of the race/ethnicity groups, the dummy variable representing Race/Ethnicity was collapsed into two binary dummy variables specifically White (1 = White; 0 = Not white) and Black (1 = Black, 0 = Not Black). To represent an association between SES and Race/Ethnicity, SES (1 = free lunch, 2 = reduced lunch, or 3 = no free/reduced lunch) was multiplied by Race/Ethnicity to construct a new composite ordinal variable named SES X White, where 0 =

Non-White (the reference baseline); 1 = White with free lunch; 2 = White with reduced lunch, and 3 = White with no free/reduced lunch. To represent an association between Special Education and Race/Ethnicity, Special Education (0 = No special education; 1 = with special education) was multiplied by Race/Ethnicity (0 = Not Black, 1 = Black) to construct a new composite ordinal variable named Special Education X Black, where 0 = No special education (the reference baseline) 1 = Black with special education.

Based on the conditioned data, evidence was provided to conclude that the assumptions of MLR were not seriously violated. The partial correlations did not generally decline to zero relative to the zero-order coefficients, apart for the partials for the MAP scores, which reflected their inter-correlation. Furthermore, the residuals were normally distributed and relatively evenly distributed on either side of their mean (zero) value. The overall model including the three MAP scores was statistically significant. In addition, Grade, Sex, and SES X White were significant predictors of the PASS ELA scores at the .05 level. The only variable that was not a significant predictor of the PASS ELA scores at $\alpha = .05$ was Special Education X Black. However, even based on the conditioned data, the $R^2 = 15.3\%$ was a small effect size, indicative of a model with limited practical significance.

Implications

This research is significant because it adds new knowledge to the field of accountability and formative assessment. The accountability measures of NCLB require schools to ensure all students are proficient on state reading and math standards by 2014 (NCLB, 2001). Each state is expected to implement high academic standards along with reliable and accurate assessments. Schools are required to show more than just the average student attaining grade-level proficiency to prevent masking students who are not meeting expectations. Under NCLB, proficiency is reported by grade levels and subgroups. The subgroups include students from different racial and ethnic backgrounds, students who are economically disadvantaged, students with limited English proficiency, and students with disabilities (Yell, Katsiyannas, & Shiner, 2006). These subgroups were designated because research has indicated many of these students are falling behind their peers. Surprisingly, the results of the current study provided statistical evidence to conclude that there was *no significant variation* of practical importance in the PASS ELA scores with respect to sex, race/ethnicity, SES, or special education.

As the national debate continues about how to best measure school effectiveness, the results of the current study need to be taken into consideration. The reported intention of high-stakes accountability testing is to improve education. This is meant to be accomplished through the alignment of curriculum with standards and tests, increasing the efficiency of education through better resource allocation, increasing student and staff motivation, and providing for educational equity. With the results of the current study indicating that there was *no significant variation* in high-stakes testing with respect to sex, race/ethnicity, SES, or special education, some would suggest that the accountability provisions of NCLB have been effective in closing the achievement gap. Since the passing of NCLB's accountability provisions, similar increases in test scores have been observed in many school districts across the country. The degree to which those increases in scores reflect real improvements in student achievement has been the subject of intense debate (Cronin, Dahlin, Xiang, McCahon, 2009). In addition, there has been serious criticism about whether the instructional changes caused by the test-based accountability (i.e., reducing time for instruction, narrowing the curriculum to only test items, and limiting divergent and higher-order thinking) outweigh the benefits (Center on Education Policy [CEP], 2009; Nichols & Berliner, 2007 ; Pedulla, Abrams, Madaus, Russell, Ramos, & Miao, 2003; Supovitz, 2009).

Nonetheless, the pressures NCLB puts on school districts to raise student achievement is profound. In this quest of raising student achievement, schools are in desperate need of assessments that accurately gauge student achievement prior to students taking high-stakes assessments. Formative assessments may meet this need as they provide teachers the necessary feedback to modify their practices before the high-stakes test. Research has shown that when formative assessments are used effectively learning outcomes improve across all content areas, grade levels, and all levels of achievement (Black & Williams, 1998; Fisher, Lapp, Flood, & Moore, 2006). In the area of reading, the need for reliable and valid progress monitoring tools that reflect the multiple dimensions of reading (i.e., decoding, comprehension, and fluency) and allows for assessment of large group of students has become imperative. According to Kingsbury and Hauser (2004), CAT, which has been used very successfully in the military (Sands, Waters, & McBride, 1997) and in professional certification and licensure (Zara, 1992) may be the answer.

MAP developed by the Northwest Evaluation Association (NWEA) is a CAT which adjusts to match the performance of the student after each item is given. NWEA was founded in 1976 by a group of school districts looking for practical answers to measure efficiently and accurately how much students have achieved and how quickly they are learning (NWEA, 2003). Since then, NWEA has created one of the most widely used CATs. MAP is used by more than 2,340 school districts in the United States and in 61 other countries (Ash & Sawchuk, 2008).

NWEA did an extensive study of validity and reliability of the MAP assessment in 2004. In the 2004 study, NWEA employed concurrent validity using a number of state and nationally normed tests in close proximity to an administration of the MAP test. Pearson correlation coefficients ranged from .69 to .88 and averaged approximately .85. In addition, since the MAP is used across the country in various states, regular state alignment studies have been conducted by the manufacturers of the MAP to examine the correspondence between the MAP and state standardized tests used to measure student achievement. Four prior studies investigating the alignment of cut scores between the spring administrations of South Carolina's previous high-stakes test (i.e., Palmetto Achievement Challenge Test [PACT]) and the MAP tests have suggested that the MAP tests are reliable and valid predictors of student performance on the PACT (Hauser, 2002; Cronin & Hauser, 2003; Cronin, 2004; Cronin, 2007).

Surprisingly, the results from the current study indicated that there was no direct relationship of practical significance between the MAP scores and the PASS ELA scores. The

MAP scores could not be used in practice to predict the PASS ELA scores or *vice-versa* since less than 2% of the variability in the scores was explained by this relationship. While the Pearson's correlation coefficients between MAP and PASS ELA scores were significant at $p < .001$ that statistical significance was merely a reflection of the large sample size. Based on these results, statistical evidence is provided to hypothesize that the processes that cause variability in the MAP scores are not the same as those that cause variability in the PASS ELA scores. This conclusion implies that MAP and PASS ELA are not in necessarily in direct alignment with each other, and they appear to measure different dimensions of student performance. As a result, school district currently using the MAP to predict performance on the PASS ELA should do so with extreme caution.

A reason for the discrepancy between previous research findings and the current study may be a result of the introduction of the PASS during the 2008-2009 school year. As previously mentioned, NWEA conducts regular state alignment studies to examine the correspondence between the MAP and state standardized tests used to measure student achievement. The alignment process begins with a thorough review of a state's standards document by NWEA's curriculum specialists. The general goal areas or strands within a state's standards that appear across grade levels become the goals in the goal structure. Areas in a state's standards documents that are determined to be sub-domains of the goals/strands become the sub-goals in the goal structure. In addition, each alignment study identifies the specific RIT scale scores from the MAP that correspond to the various proficiency levels for each subject (reading, mathematics, etc.) and for each student grade. Because all states use different tests for measuring student achievement, alignment studies are necessary for each state.

In order to conduct alignment studies, it is necessary to study the performance of students who have completed both the state test and the MAP (NWEA, n.d.b). When updates or changes occur to state tests, such as with the introduction of PASS during the 2008-2009 school year, a new MAP/State test alignment study must be completed. Such updates cannot be attempted, however, until after the state test has been administered at least once (NWEA, n.d.b). With the SCDOE introducing the PASS in Spring 2009, NWEA could not attempt an alignment study until after the Spring 2009 PASS test data were published. The Spring 2009 PASS test data were released in December 2009. Therefore, it could be speculated that a discrepancy exists between previous research findings, which suggest that the MAP is a reliable and valid indicator of student performance on state tests, and the current study because NWEA was unable to conduct the recommended alignment study on the MAP and PASS test data on which this study was based.

Limitations

Potential threats to the validity of results can occur within a study to interfere with attempts to build casual relationships (internal validity), and others may occur more extraneously, which can limit the generalization to other populations and settings (external validity). Some threats to the validity of this study are noted below as cautions of interpretations of current findings.

The ability of MAP Reading to reliability predict performance on the PASS ELA may change with future administrations based on changes to the MAP as a result of the recommended alignment study. According to NWEA, when updates or changes occur to state tests such as with the introduction of PASS during the 2008-2009 school year, a new MAP/State test alignment study must be completed. Recently, NWEA completed an alignment study to connect the scale of the PASS with NWEA's RIT scale (NWEA, 2010). Information from the PASS assessments was used to establish performance-level scores on the RIT scale that would indicate a good chance of success on these tests. Thus, the ability of the MAP to predict performance on future administrations of PASS may increase as a result of the changes made to the MAP based on the recent state alignment study conducted by NWEA.

While research indicates minimal differences, performances on the two tests may have varied based on the mode by which they were administered (i.e., computer vs. paper-and-pencil). Many comparability studies found computer tests to be equivalent in difficulty or slightly easier than paper tests (Bridgeman, Bejar, & Friedman, 1999; Choi & Tinkler, 2002; Mead & Drasgow, 1993; Pearson Educational Measurement, 2002, 2003; Poggio, Glasnapp, Yang, & Poggio, 2005; Pommerich, 2004; Pomplun, Frey, & Becker, 2002; Russell, 1999; Russell & Haney, 1997, 2000; Russell & Platti, 2001; Schaeffer, Bridgeman, Golub-Smith, Lewis, Potenza, & Steffen, 1998; Nichols & Kirkpatrick, 2005; Taylor, Jamieson, Eignor, & Kirsch, 1998; Zandvliet & Farragher, 1997; Wang, 2004). Results from these studies indicated that student performance was similar across the demographics of sex, academic placement, and SES. Fewer studies found students in K-12 performed poorer on computer tests than paper tests (Cerrillo & Davis, 2004; Choi & Tinkler, 2002; O'Malley, Kirkpatrick, Sherwood, Burdick, Hsieh, & Sanford, 2005; Russell & Plati, 2001). Research has shown that long reading passages on a computer tend to be more difficult than on paper (Murphy, Long, Holleran, & Esterly, 2003; O'Malley, et al., 2005). For tests where all of the information for an item could be presented in its entirety on one screen, results of comparability studies often show small or insignificant mode effects (Bridgeman, Lennon, & Jackenthal, 2003; Choi & Tinkler, 2002; Hetter, Segall, & Bloxom, 1997; Spray, Ackerman, Reckase, & Carlson, 1989). For tests where all of the information for an item could not be presented in its entirety on one screen, and some form of navigation (typically scrolling) was necessary on the part of the examinee to view all of the information, results often showed more significant mode effects (Bridgeman, Lennon, & Jackenthal, 2003; Choi & Tinkler, 2002). With regards to MAP Reading, students are required to scroll down the computer screen to read longer passages which may have affected their performance.

Student maturation and attitude may have also changed over the course of the study because of the range of grades examined. Student performance can fluctuate on a daily basis due to various physical and emotional states. In addition to overall general physical and emotional well-being, this study did not take into account specific test anxiety. Computer anxiety, has been determined a cause of concern in computerized assessments, and may be a factor in tests outcomes, negatively impacting students with little or no experience in computerized assessments. Research has found that computer familiarity was related to computer test performance; students who rarely used a computer tended to perform poorer in both math and reading than those students who had more computer experience (Choi & Tinkler, 2002). Furthermore, test anxiety experienced during high-stakes testing was not taken into account. According to Triplett and Barksdale (2005), children overwhelmingly experienced negative emotions, including nervousness and anger, in relation to high-stakes testing.

The generalization of this study is limited since it was based on two measures that assess South Carolina ELA academic standards. The SCDOE (2008) in consultation with Mid-continent Research for Education and Learning developed the ELA standards and indicators utilizing a number of resources. Important among them are the English language arts standards documents of several other states as well as the national standards document *Standards for the English Language Arts*, published jointly in 1996 by the National Council of Teachers of English and the International Reading Association. While there is sure to be some overlap among the various state standards, caution needs to be given that the analyses conducted in this study was based on South Carolina ELA academic standards.

Although this study investigated the correlation between a CAT and a traditional paper-and-pencil test, it is limited to the testing of reading. While the preponderance of the evidence suggests that, for multiple-choice-only tests, student performance is not significantly different for different modes of administration; some studies suggest students might do better on computer, while others suggest they might do better on paper. The research reviewed for this study centered on the area of reading. Additional research is needed to compare student outcomes in other subject matter areas such as math, writing, science, and social studies.

Recommendations for Future Research

During the past decade, schools have invested heavily in putting technology—especially computers and their associated infrastructure—in the hands of students, teachers, and administrators. Based on a recent survey of U.S. Public Schools, an estimated 100 percent of public schools have one or more instructional computers with Internet access with the ratio of students to instructional computers being approximately 3 to 1 (NCES, 2010). With the significant rise of technology in schools, the field of CAT as a formative assessment in the K-12 educational system is poised to grow. Research has shown that when formative assessments are used effectively learning outcomes improve, which subsequently is reflected in accountability testing (Black & Williams, 1998; Fisher, Lapp, Flood, & Moore, 2006). Therefore, further research exploring whether the use of the MAP as an assessment tool in planning interventions is recommended. This research would involve comparing the outcomes of students who take only the PASS test versus students who first take MAP, are provided remediation based on identified weaknesses from MAP, and then take the PASS test.

The focus of this research was to investigate the predictive validity of the MAP Reading in relation to performance on the PASS ELA. The current study does not support the use of the MAP in predicting performance on the PASS ELA. However, since this study was conducted, NWEA completed an alignment study to connect the scale of the PASS with NWEA's RIT scale (NWEA, 2010). Thus, additional research may be warranted to determine if the findings of this study are replicated. The ability of the MAP to predict performance on future administrations of PASS may increase based on changes made to the MAP from the alignment study. In addition, further research is also needed to compare student outcomes in other subject matter areas, such as math, writing, science, and social studies using both computer based tests and paper and pencil tests to determine what subject areas can be assessed accurately through computer based assessments.

Finally, the school accountability system operating currently in the nation requires schools to disaggregate data to look at subgroups so determinations regarding growth are tracked for all students. This study included three out of the four subgroups required by NCLB (i.e., Special Education, Free and Reduced Lunch, and Race/Ethnicity). Further research needs to be conducted looking at the fourth subgroup, English Language Learner. In addition, the needs and abilities of special education students are vast and wide. Investigating the use of CAT in relation to various disabilities may provide support in using CAT in the assessment of present levels of performance as required in a student's individualized education program (IEP).

Summary

This study investigated the predictive validity of demographic variables and the MAP in relation to student performance on the South Carolina's PASS ELA test. Establishing further support for this relationship could improve teachers' acceptability of the MAP for the purpose of clear goal setting with regards to student progress. Various predictive factors of student performance on the PASS ELA test were analyzed. Demographic factors included sex, race/ethnicity, SES, special education, and degree of exposure to CAT. The specific predictive factors of the MAP that were analyzed included the MAP Reading RIT score, the MAP Understanding and Using Literary Texts Goal Performance area, the MAP Understanding and Using Informational Texts Goal Performance area, and the MAP Building Vocabulary Goal Performance area. From among the demographic variables and MAP scores analyzed, the aim was to identify the best predictors of student performance on the PASS ELA.

Archival test data and demographic information were obtained from five elementary and three middle schools located in the target school district. The sample was comprised of 3,861 students in grades 3-8 who had scores on the Spring 2009 MAP Reading and Spring 2009 PASS ELA assessments. The variables were coded in the SPSS data editor and test data were analyzed using associational measures including Cross-tabulation, ANOVA, Pearson correlation, and MLR leading to the construction of a hypothetical path model. From among the demographic variables, it was concluded that there were no relationships of practical significance between the PASS ELA scores and the sex, race/ethnicity, SES, special education, or grades of the students. Significant correlations between the MAP Understanding and Using Literary Texts Goal Performance, the MAP Understanding and Using Informational Texts Goal Performance, and the MAP Building

Vocabulary Goal Performance were found. However, there was no direct relationship of practical significance between the MAP scores and the PASS ELA scores.

Regular state alignment studies have been conducted by the manufacturers of the MAP to examine the correspondence between the MAP and state standardized tests used to measure student achievement. Four prior studies investigating the alignment between the spring administrations of the PACT (South Carolina's previous high-stakes test) and the MAP tests have suggested that the MAP tests are reliable and valid predictors of student performance on the PACT (Hauser, 2002; Cronin & Hauser, 2003; Cronin, 2004; Cronin, 2007). When updates or changes occur to state tests such as with the introduction of PASS during the 2008-2009 school year, a new MAP/State test alignment study must be completed. Such updates cannot be attempted, however, until after the state test has been administered at least once (NWEA, n.d.b). With SCDOE introducing the PASS in Spring 2009, NWEA could not attempt an alignment study until after the Spring 2009 PASS test data were published. The Spring 2009 PASS test data were released in December 2009. Therefore, speculation can derive that a discrepancy exists between previous research findings, which suggests that the MAP is a reliable and valid indicator of student performance on state tests, and the current study because NWEA was unable to conduct the recommended alignment study on the MAP and PASS test data on which this study was based. However, since this study was conducted, NWEA completed an alignment study to connect the scale of the PASS with NWEA's RIT scale (NWEA, 2010). Thus, additional research may be warranted to determine if the findings of this study are replicated.

REFERENCES

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college.* Washington, DC: U.S. Department of Education. Retrieved from http://www2.ed.gov/rschstat/research/pubs/toolboxrevisit/index.html

Advancement Project. (2010). *Test, punish, and push out: How 'Zero Tolerance' and high-stakes testing funnel youth into the school-to-prison pipeline*. Retrieved from http://www.advancementproject.org/sites/default/files/publications/01-EducationReport-2009v8-HiRes.pdf

Airasian, P., & Abrams, L. (2002). Classroom student evaluation. In D. Nevo and D. Stufflebeam (Eds.), *International handbook of educational evaluation*. Norwell, MA: Kluwer Academic Press.

American College Testing (ACT, 2006). *Reading between the lines: What the ACT reveals about college readiness in reading.* Iowa City, IA: Author. Retrieved from http://www.act.org/research/policymakers/pdf/reading_report.pdf

Ash, K., & Sawchuk, S. (2008, November 19). Adjusting to test takers. *Education Week, 28*(13), 19-21.

Barger, J. (2003). *Comparing the DIBELS oral reading fluency indicator and the North Carolina end of grade reading assessment* (Technical Report). Asheville, NC: North Carolina Teacher Academy.

Barksdale-Ladd, M. A. & Thomas, K. F. (2000). What's at stake in high stakes testing: Teachers and parents speak out. *Journal of Teacher Education, 51*, 384-397.

Barton, P. (2000). *What jobs require: Literacy, education, and training, 1940-2006*. Princeton, NJ: Educational Testing Services Policy Information Center. (ERIC Document Reproduction Service No. ED439136)

Bausell, C.V., & Klemick, E. (2007, March 29). Tracking US trends. *Education Week, 26*(30),42–44.

Bennett, R. E. (2001). How the internet will help large-scale assessment reinvent itself [Electronic version]. *Education Policy Analysis Archive, 9*(5), Retrieved from http://epaa.asu.edu/epaa/v9n5.html

Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment, 1*(1), 1-24. Retrieved from http://www.jtla.org

Biancarosa, G., & Snow, C. E. (2004). *Reading next: A vision for action and research in middle and high school literacy.* Washington, DC: Alliance for Excellent Education. Retrieved from http://www.all4ed.org/files/ReadingNext.pdf

Bielinski, J., & Ysseldyke, J. (2000*). Interpreting trends in the performance of special education* *students* (Technical Report 27). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

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

Bolt, D., & Crawford, R. (2000). *Digital divide: Computers and our children's future*. New York: TV Books.

Braden, J. P. (2002). Best practices for school psychologists in educational accountability: High stakes testing and education reform. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 301-320). Washington, DC: National Association of School Psychologists.

Bridgeman, B., Bejar, I. I., & Friedman, D. (1999). Fairness issues in a computer-based architectural licensure examination. *Computers in Human Behavior, 15*, 419-440.

Bridgeman, B., Lennon, M. L., & Jackenthal, A. (2001). *Effects of screen size, screen resolution, and display rate on computer-based test performance* (ETS RR-01-23). Princeton, NJ: Educational Testing Services.

Brown-Chidsey, R., Boscardin, M. L., & Sireci, S. G. (2001). Computer attitudes and opinions of students with and without disabilities. *Journal of Educational Computing Research, 24*, 183-204.

Buck, J., & Torgesen, J. K. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* (Technical Report 1) Tallahassee, FL: Florida Center for Reading Research, Retrieved from http://www.fcrr.org/science/scienceTechnicalReports.htm

Carlson, R. D. (1993-94). Computer adaptive testing: A shift in the evaluation paradigm. *Journal of Educational Technology Systems, 22*(3), 213-24.

Center on Education Policy. (2007). *Answering the question that matters most: Has student achievement increased since No Child Left Behind?* Washington, DC: Author. Retrieved from http://www.cep-dc.org/index.cfm?fuseaction=document.showDocumentByID&nodeID=1&DocumentID=200

Center on Education Policy. (2008). *Has student achievement increased since 2002? State test score trends through 2006–07*. Washington, DC: Author. Retrieved from http://www.cep-dc.org/index.cfm?fuseaction=document_ext.showDocumentByID&nodeID=1&DocumentID=241

Center on Education Policy. (2009). *Summary: Lessons from the classroom level about federal and state accountability in Rhode Island and Illinois*. Washington, DC: Author. Retrieved from http://www.cep-dc.org/index.cfm?fuseaction=Page.viewPage&pageId=500&parentID=481

Cerillo, T. L., & Davis, J.A. (2004). *Comparison of paper-based and computer based administrations of high-stakes, high-school graduation tests.* Paper presented at the Annual Meeting of the American Education Research Association, San Diego, CA.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.

Chaney, E., & Gilman, D. A. (2005). Filing in the blanks: Using computers to test and teacher. *Computers in the Schools, 22*, 157-168.

Choi, S. W., & Tinkler, T. (2002). *Evaluating comparability of paper and computer-based assessment in a K-12 setting*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

Chudowsky, N., & Chudowsky, V. (2005). *States test limits of federal AYP flexibility.* Washington, DC: Center on Education Policy.

Coalition for Juvenile Justice (CJJ, 2001). *Abandoned in the back row: New lessons in education and delinquency prevention.* Retrieved from http://juvjustice.org/resource_122.html

Cohen, J. (1988). *Statistical power analysis for the behavior sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

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

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

Cronin, J. (2004, August). *A study of the ongoing alignment of the NWEA RIT Scale with the South Carolina Palmetto Achievement Challenge Tests (PACT).* Retrieved from http://www.nwea.org/research/state.asp

Cronin, J. (2007, March). *A study of the alignment of the NWEA RIT Scale with the South Carolina Assessment System.* Retrieved from http://www.nwea.org/research/getreport.asp?ReportID=162

Cronin, J., Dahlin, M., Adkins, D., & Kingsbury, G. G. (2007). *The proficiency illusion*. Washington, DC: Thomas Fordham Institute. Retrieved from http://www.fordhaminstitute.org/publications/index.cfm?topic=1

Cronin, J., Dahlin, M. Xiang, Y., & McCahon, D. (2009). *The accountability illusion.* Washington, DC: Thomas Fordham Institute. Retrieved from http://www.edexcellence.net/index.cfm/news_the-accountability-illusion

Cronin, J., & Hauser, C. (2003, March). *Aligning the NWEA RIT Scales with the South Carolina Palmetto Achievement Challenge Tests*. Retrieved from http://www.nwea.org/research/state.asp

Davis, M. (2006). NCLB panel gathers views on testing and data collection. *Education Week, 25*(37), 25-29.

Donahue, P. L., Daane, M. C., & Jin, Y. (2003). *The nation's report card: Reading 2003*. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2003/2005453.pdf

Duke, D. L., & Reck, B. L. (2003). The evolution of educational accountability in the Old Dominion. In D. L. Duke, M. Grogan, P. D. Tucker & W. F. Heinecke (Eds.), *Educational leadership in an age of accountability: The Virginia experience* (pp. 36-68). Albany, NY: SUNY Press.

Edley, C., Jr., & Wald, J. (2002, December 16). The grade retention fallacy. *The Boston Globe*.

Education Accountability Act, South Carolina Code Ann. § 59-18-310 (1998 & Supp. 2008)

Education Trust. (2002). *New school finance data analysis shows deep inequities, but gaps in some states decreasing.* Retrieved from http://www.edtrust.org/dc/press-room/press-release/new-school-finance-data-analysis-shows-deep-inequities-but-gaps-in-some-

Ehren, B. J., Lenz, B. K., & Deshler, D. D. (2004). Enhancing literacy proficiency with adolescents and young adults. In C. A. Stone, E. R. Silliman, B. J. Ehren, & K. Apel, *Handbook of language and literacy: Development and disorders* (pp. 681–702). New York, NY: Guilford Press.

Field, A. P. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: Sage Publications.

Fisher, D., Lapp, D., Flood, J., Frey, N., & Moore, K. (2001). Why class size matters: An investigation of teacher, administrator, and parent perceptions. *The Yearbook of the National Reading Conference, 50,* 189-199.

Fichten, C. S., Asuncion, J. V., Barile, M., Ferraro, V., & Wolforth, J. (2009). Accessibility of e-learning and computer and information technologies for students with visual impairments in postsecondary education. *Journal of Visual Impairment & Blindness, 103* (9), 543-557.

Fletcher, J. M., & Lyon, G. R. (1998). Reading: A research-based approach. In W. M. Evers (Ed.), *What's gone wrong in America's classroom.* (pp. 49-90). Stanford, CA: Hoover Institute Press.

Flugum, K. R., & Reschley, D. J. (1994). Preferral interventions: Quality indices and outcomes. *Journal of School Psychology, 32,* 1-14.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education, 41*, 121-139.

Folk, V. G., & Smith, R. (1998, September). *Model for delivery of computer-based tests*. Paper presented at the ETS-sponsored colloquium on Computer-Based Tests: Building the

Foundation for Future Assessments, Philadelphia, PA.

Fryer, R., & Levitt, S. (2006). The Black-White test score gap through the third grade. *American Law and Economic Review*, *8*(2), 249-281.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*, 45–58.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review*, *22*, 27-48.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, *9*(2), 20-28.

Gallagher, A., Bridgeman, B., & Cahalan, C. (2000). *The effect of computer-based tests on racial/ethnic, gender, and language groups* (ETS Research Rep. No. 00-8). Princeton, NJ: Educational Testing Services. Retrieved from http://www.ets.org/Media/Research/pdf/RR-00-08-Gallagher.pdf.

Gates, A. (1961). Sex differences in reading ability. *Elementary School Journal*, *61*(8), 431–34.

Glassnapp, D. R., Poggio, J., Poggio, A., & Yang, X. (2005). *Student attitudes and perceptions regarding computerized testing and the relationship to performance in large scale assessment programs*. Paper presented at the annual meeting on the National Council on Measurement in Education, Montreal.

Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61-88). New York: Guilford Press.

Good, R. H., Simmons, D. C., & Kameenui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*, 257-288.

Grigg, W. S., Daane, M. C., Jin, Y., & Campbell, J. R. (2003). *The nation's report card: Reading 2002*. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2002/2003521.pdf

Hauser, R. (2002, January). *Alignment of the NWEA RIT scale to PACT.* Retrieved from http://www.nwea.org/research/state.asp

Hawke, J. L., Wadsworth, S. J., Olson, R. K., & DeFries, J. C. (2007). Etiology of reading difficulties as a function of gender and severity. *Reading and Writing, 20*, 13–25.

Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart and Winston.

Heinecke, W. F., Moon, T., & Curry-Corcoran, D. (2003). U.S. schools and the new standards and accountability initiative. In D. L. Duke, M. Grogan, P. D. Tucker, & W. F. Heinecke (Eds.), *Educational leadership in an age of accountability.* Albany: SUNY Press.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve. Intelligence and class structure in American life*. New York: Free Press Paperback.

Hetter, R. D., Segall, D. O., & Bloxom, B. M. (1997). Evaluating item calibration medium in computerized adaptive testing. In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 161—167). Washington, DC: American Psychological Association.

Higgins, J., Russell, M., & Hoffman, T. (2005). Examining the effect of computer-based passage presentation on reading test performance. *Journal of Technology, Learning, and Assessment, 3*(4). Retrieved from http://www.jtla.org

Hintze, J. M., Owen, S. V., Shapiro, E. S., & Daly, E. J. (2000). Generalizability of oral reading fluency measures: Application of G-theory to curriculum-based measurement. *School Psychology Quarterly, 15*, 52-68.

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of CBM-R and high-stakes testing. *School Psychology Review, 34*, 372-386.

Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher, 54*(5), 482-494.

Hogrebe, M., Nist, S., & Newman, I. (1985). Are there gender differences in reading achievement? An investigation using the high school and beyond data. *Journal of Educational Psychology, 77*(6), 716–24.

Howe, K. B., Scierka, B. J., Gibbons, K. A., & Silberglitt, B. (2003). A school-wide organization system for raising reading achievement using general outcome measures and evidence-based instruction: One education district's experience. *Assessment for Effective Intervention, 28,* 59-72.

Hunter, J. E. (1997). Needed: A ban on the significance test. *American Psychological Society, 8*, 3-7.

Ito, K. & Sykes, R. C. (2004). *Comparability of scores from norm-referenced paper-and-pencil and web-based linear tests for grades 4-12*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA. Retrieved from http://www.ctb.com/media/articles/pdfs/ResearchArticles/ComparabilityofScores.pdf?FOLDER%3C%3Efolder_id=2534374302134983

Jarrell, D. (2005). Five commonsense fixes for NCLB. *American School Board Journal, 192(*9), 66-67.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children, 59*, 421-432.

Johnson, M. & Green, S. (2006). On-line mathematics assessment: The impact of mode on performance and question answering strategies. *The Journal of Technology, Learning, and Assessment, 4*(5), 4-34. Retrieved from http://www.jtla.org

Kamil, M. L. (2003, November). *Adolescents and literacy: Reading for the 21st century*. Washington, DC: Alliance for Excellent Education.

Kaufman, P., Alt, N. M., & Chapman, C. (2001). *Dropout rates in the United States: 2000.* Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002114

Keller-Margulis, M. A., Shapiro, E. S., & Hintze, J. M. (2008). Long-term diagnostic accuracy of curriculum based measures in reading and mathematics. *School Psychology Review, 37,* 374-390.

Kingsbury, G. G., & Hauser, C. (2004). *Computer adaptive testing and No Child Left Behind*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA. Retrieved from http://www.kingsburycenter.org/sites/default/files/Computerized_Adaptive_Testing_and_NCLB.pdf

Kirsch, I. S., Jungeblut, A., Jenkins, L. & Kolstad, A. (1993). *Adult literacy in America*. Washington, D.C.: Government Printing Office

Klein, S. P., & Hamilton, L. (1999). *Large-scale testing: Current practices and new directions* (Research Report IP-182). Santa Monica, CA: Rand Education. (ERIC Document Reproduction Service No. ED460965)

Kleinfeld, J. (2006). *Five powerful strategies for connecting boys to schools.* Paper presented at the White House Conference on Helping America's Youth, Indianapolis, IN.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research.* Washington DC: American Psychological Association

Kobrin, J. L., & Young, J. W. (2003). The cognitive equivalence of reading comprehension test items via computerized and paper-and-pencil administration. *Applied Measurement in Education, 16*(2), 115-140.

Lee, J. (2006). Input-guarantee versus performance-guarantee approaches to school accountability: Cross-state comparisons of policies, resources, and outcomes. *Peabody Journal of Education, 81*(4), 43-64.

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, DC: Economic Policy Institute.

Lee, J., Grigg, W., and Donahue, P. (2007). *The nation's report card: Reading 2007*. Washington, DC: National Center for Education Statistics.

Linn, R. L. (2001). A century of standardized testing: Controversies and pendulum swings. *Educational Assessment, 7*(1), 29-38.

Marchant, G. J., & Paulson, S. E. (2005). The relationship of high school graduation exams to graduation rates and SAT scores. *Education Policy Analysis Archives, 13*(6)

Marshall, M. (2009). Literacy today comment. *Literacy Today, 60,* 6-6.

McCann, R. (2006). The equivalence of online and traditional testing for different subpopulations and item types. *British Journal of Educational Testing, 37*(1), 79-81.

McCall, M. S., Hauser, C., Cronin, J., Kingsbury, G. G., & Hauser, R. (2006). *Achievement gaps: An examination of differences in student achievement and growth*. Retrieved from www.nwea.org/assets/research/national.asp

McGlinchey, M. T., & Hixson, M.D. (2004). Using curriculum-based measurement to predict validity of CBM-R and high-stakes testing. *School Psychology Review, 34,* 372-386.

McMaster, K., & Espin, C. A. (2007). Technical features of curriculum-based measurement in writing: A literature review. *The Journal of Special Education, 41*, 68-84.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114*(3), 449-458.

Mead, S. (2006). *The truth about boys and girls.* Washington, DC: Education Sector.

Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement*, *23*(3), 187-194.

Meredith, K. S., Steele, J. L., & Dawson, M. M. (1997). Reading. In G. G. Bear, K. M. Minke, & A. Thomas (Eds.), *Children's needs II: Development, problems, and alternatives.* (pp. 397-405). Bethesda, MD: The National Association of School Psychologists.

Mittag, K. C., & Thomson, B. (2000). A national survey of AERA members' perceptions of statistical significance tests and other statistical issues. *Educational Researcher*, *29*, 14-19.

Moats, L. C. (1999). *Teaching reading is rocket science: What expert teachers of reading should know and be able to do.* Washington, DC: American Federation of Teachers. (ERIC Document Reproduction Service No. ED445323)

Mullis, I. V. S, Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools in 35 countries*. Chestnut Hill, MA: Boston College.

Murphy, P. K., Long, J., Holleran, T., & Esterly, E. (2000). Persuasion online or on paper: A new take on an old issue. *Learning & Instruction, 13*(5), Vol. 13 Issue 5, 511-532.

National Commission on Excellence in Education. (NCEE, 1983). *A nation at risk*. Washington, DC: U.S. Department of Education. Retrieved from: http://www.ed.gov/pubs/NatAtRisk/index.html

National Center for Education Statistics. (NCES, 2002). *Statistical program standards*. Retrieved from http://nces.ed.gov/StatProg/2002/std1_5.asp

NCES. (2006). *National assessment of adult literacy: A first look at the literacy of America's adults in the 21st century*. Retrieved from http://nces.ed.gov/NAAL/PDF/2006470%5f1.PDF

NCES. (2009a). *The nation's report card: Reading 2009* . Washington, DC: U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2009/2010458.pdf

NCES. (2009b). *Technical report and data file user's manual. For the 2003 National Assessment of Adult Literacy*. Retrieved from http://nces.ed.gov/naal/datafiles.asp

NCES. (2010). *Educational technology in U.S. public schools: Fall 2008.* Retrieved from http://nces.ed.gov/pubs2010/2010034.pdf

Neter, J., Kutner, M. H., Wasserman, W., & Nachsteim, C. (1996). *Applied linear regression models*. Chicago, IL: Irwin.

Nichols, P., & Kirkpatrick, R. (2005). *Comparability of the computer-administered tests with existing paper-and-pencil tests in reading and mathematics tests*. Paper presented at the AERA Division D Graduate Student Seminar, Montreal, Canada.

Nichols, S., & Berliner, D. (2007). *Collateral damage: How high-stakes testing corrupts America' schools.* Boston: Harvard Education Press.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2001). Washington, DC: U.S. Department of Education.

Northwest Evaluation Association. (NWEA, n.d.a). Assessments. In *Assessment Systems*. Retrieved from http://www.nwea.org/assessments/

NWEA. (n.d.b). *Linking MAP to state tests: Proficiency cut score estimation procedures.* Retrieved from http://www.nwea.org/sites/www.nwea.org/files/NWEA%20State%20Standards%20Alignment%20Study%20Methods%20_3_.pdf

NWEA. (2003). *Technical manual: For use with Measures of Academic Progress and Achievement Level Tests.* Portland, OR: Northwest Evaluation Association.

NWEA. (2004). *Reliability and validity estimates: NWEA Achievement Level Tests and Measures of Academic Progress*. Retrieved from http://www.nwea.org/assets/research/NWEA%20Reliability%20&%20Validity.pdf

NWEA. (2007a). *NWEA state standards alignment study methods*. Retrieved from http://www.nwea.org/research/state.asp

NWEA. (2007b). Using SC 2007 academic standards for English language arts. *Measures of Academic Progress (MAP) South Carolina State – Aligned Version 5*. Retrieved from www.nwea.org/research/state.asp?state=SC

NWEA. (2010). *South Carolina alignment report*. Retrieved from http://www.nwea.org/our-research/state-information

Olejnik, S., & Algina, S. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*, 241-286.

Olson, R. K. (2002). Dyslexia: Nature and nurture. *Dyslexia, 8*, 143–159.

O'Malley, K. J., Kirkpatrick, R., Sherwood, W., Burdick, H. J., Hsieh, M. C., & Sanford, E. E. (2005). *Comparability of a paper based and computer based reading test in early elementary grades*. Paper presented at the AERA Division D Graduate Student Seminar, Montreal, Canada.

Ommerborn, R., & Schuemer, R. (2001). *Using computers in distance study: Results of a survey amongst disabled students*. Retrieved from: http://www.fernuni-hagen.de/StabEQ/qbehind.htm

Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer-Verlag.

Pearson Educational Measurement. (2002). *Final report on the comparability of computer-delivered and paper tests for Algebra I, Earth Science, and English.* Austin, TX: Author.

Pearson Educational Measurement. (2003). *Virginia standards of learning web-based assessments comparability study report – Spring 2002 administration: Online & paper tests*. Austin, TX: Author.

Pearson Educational Measurement. (2006). *Comparability analysis of online and paper*-and-pencil MEAP. Iowa City, IA: Author.

Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers.* Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Perie, M., Grigg, W., & Donahue, P. (2005). *The nation's report card: Reading 2005*. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2005/2006451.pdf

Perie, M., Moran, R., & Lutkus, A. D. (2005). *NAEP 2001, trends in academic progress: Three decades of student performance in reading and mathematics*. Washington, DC: National Center for Education Statistics.

Peterson, P. E., & West, M. R. (2003). *No child left behind: The politics and practice of school accountability*. Washington, DC: Brookings Institute Press.

Phillips, M., & Chin, T. (2004). School inequality: What do we know? In Kathryn Neckerman (Ed.). *Social Inequality* (pp. 467-519). New York: Russell Sage Foundation.

Poggio, J., Glasnapp, D. R., Yang, X., & Poggio, A. J. (2005). A comparative evaluation of score results from computerized and paper & pencil mathematics testing in a large scale state assessment program. *The Journal of Technology, Learning, and Assessment, 3*(6), 1-30.

Retrieved from http://www.jtla.org

Pogorzelski S., & Wheldall, K. (2002). Do differences in phonological processing performance predict gains made by older low-progress readers following intensive literacy intervention? *Educational Psychology, 22,* 413-427.

Pommerich, M. (2004). Developing computerized versions of paper tests: Mode effects for passage-based tests. *The Journal of Technology, Learning, and Assessment, 2*(6), 1-44. Retrieved from http://www.jtla.org

Pomplun, M., Frey, S., & Becker, D. F. (2002). Score equivalence of paper and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement, 62*(2), 337–354.

Popham, W. J. (2004, November). Curriculum matters. *American School Board Journal*, 30–33.

Porter, A. C., Linn, R. L., & Trimble, S. (2005). The effects of state decisions about NCLB Adequate Yearly Progress targets. *Educational Measurement: Issues and Practice, 24*(4), 32-39.

Proctor, B. D., & Dalaker, J. (2002). *Poverty in the United States: 2001. Current population reports*. Washington, DC: Bureau of the Census. (ERIC Document Reproduction Service No. ED469530)

Reardon, S. F., Atteberry, A., Arshan, N & Kurlaender, M. (2009). *Effects of the California high school exit exam on student persistence, achievement and graduation*. Stanford, CA: Stanford University: Institute for Research on Policy & Practice.

Reckase, M. D. (1989). Adaptive testing: The evolution of a good idea. *Educational Measurement: Issues and Practice, 8*(3), 11-15.

Riddle, W. (2004). *Adequate yearly progress: Implementation of the No Child Left Behind act*. Washington, DC: Congressional Research Service.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F., (2005). Teachers, schools, and academic achievement, *Econometrica, 73*(2), 417-458.

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesson, J. K. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third-grade reading comprehension outcomes. *Journal of School Psychology, 46,* 343-366.

Roman, S. R. (2004). Illiteracy and older adults: Individual and societal implications. *Educational Gerontology, 30*(2), 79-93.

Rose, L. C., & Gallup, A. M. (2001). The 33rd annual Phi Delta Kappan/Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan, 83*(1), 41-58.

Rose, L. C., & Gallup, A. M. (2003). The 35th annual Phi Delta Kappan/Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan, 85*(1), 41-52.

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper [Electronic Version]. *Education Policy Analysis Archives (EPPA), 7(20).* Retrieved from http://epaa.asu.edu/epaa/v7n20

Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper [Electronic Version]. *Education Policy Analysis Archives (EPPA), 5(3).* Retrieved from http://epaa.asu.edu/epaa/v5n3

Russell, M., & Haney, W. (2000). Bridging the gap between testing and technology in schools [Electronic Version]. *Education Policy Analysis Archives (EPPA), 8(19).* Retrieved from http://epaa.asu.edu/epaa/v5n3

Russell, M., & Plati, T. (2001). Effects of computer versus paper administration of a state-mandated writing assessment. *Teachers College Record*. Retrieved from http://www.tcrecords.org

Rutter, M., Caspi, A., Fergusson, D., Horwood, L. J., Goodman, R., Maughan, B., …Carroll, J. (2004). Sex differences in developmental reading disability: New findings from 4 epidemiological studies. *The Journal of The American Medical Association, 291*, 2007–2012.

Sandene, B., Horkay, N., Bennett, R., Allen, N., Kaplan, B., & Oranje, A. (2005). *Online assessment in mathematics and writing*: *Reports from NAEP technology-based assessment project, research and development series* (NCES 2005-457). Washington DC: National Center for Educational Statistics. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/studies/2005457_1.pdf

Sands, W. A., Waters, B. K., & McBride, J. R. (Eds.) (1997). *Computerized adaptive testing: From inquiry to operation.* Washington DC: American Psychological Association.

Schaeffer, G. A., Bridgeman, B., Golub-Smith, M. L., Lewis, C., Potenza, M. T., & Steffen, M. (1998). *Comparability of pencil-and-pencil and computer adaptive test scores on the GRE General Test (RR-98-38).* Princeton, NJ: Educational Testing Service.

School Administration Student Information. (SASI*xp*™, 2003). [Computer Software]. Upper Saddle River, NJ: Pearson Education.

Scruggs, T. E. and Mastropieri, M. A. (2002) On babies and bathwater: Addressing the problems of identification of learning disabilities. *Learning Disability Quarterly, 25*, 155-68.

Shapiro, E. S., Keller, M. A., Edwards, L., Lutz, G., & Hintze, J. M. (2006). General outcome measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 42,* 19-35.

Shaw, R., & Shaw, D. (2002). *DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP).* (Technical Report).

Eugene: University of Oregon.

Sheridan, M. (1976). *Sex differences and reading: An annotated bibliography*. Newark, DE: International Reading Association.

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in Schools, 43*(5), 527-535.

Silberglitt, B., & Hintze, J. M. (2005). Formative assessment using CBM-R cut scores to track progress toward success on state-mandated achievement tests: A comparison of methods. *Journal of Psychoeducational Assessment, 23,* 304-325.

Sim, G., & Horton, M. (2005). *Performance and attitude of children in computer based versus paper based testing*. World Conference on Educational Multimedia, Hypermedia and Telecommunications, Montreal. Retrieved from http://www.chici.org/references/performance_and_attitude.pdf

Simpson, M. A., Gong, B., & Marion, S. (June 2006). *Effect of minimum cell sizes and confidence interval sizes for special education subgroups on school-level AYP determinations*. Synthesis Report 61. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://cehd.umn.edu/nceo/OnlinePubs/synthesisreports.html

Smith, M. L., & Fey, P. (2000). Validity and accountability in high-stakes testing. *Journal of Teacher Education, 51*(5), 334-344.

South Carolina Code Ann. § 59-18-310 (E) (Supp. 2007)

South Carolina Department of Education. (SCDOE, n.d.a). *Palmetto Assessment of State Standards (PASS) Grades 3-8*. Retrieved from http://ed.sc.gov/agency/Accountability/Assessment/NewAssessmentProgramGr3-8.html

SCDOE. (n.d.b). *Adoption List of Formative Assessments*. Retrieved from http://ed.sc.gov/agency/Accountability/Assessment/old/assessment/FormativeAssessment.html

SCDOE. (n.d.c). *State scores by grade level and demographic category - 2009 - PASS*. Retrieved from http://www.ed.sc.gov/topics/assessment/scores/pass/2009/statescoresdemo.cfm

SCDOE. (n.d.d.) *2008 State Report Card - No Child Left Behind -Adequate Yearly Progress.* Retrieved from http://ed.sc.gov/topics/assessment/scores/ayp/2008/fullratings.cfm

SCDOE. (2009). *York 4 2009 state report card - No Child Left Behind - Adequate yearly progress*. Retrieved 3/20/2010 from http://ed.sc.gov/topics/assessment/scores/ayp/2009/show_district_ayp_scores.cfm?ID=4604

SCDOE. (2008). *South Carolina academic standards for English language arts*. Retrieved from http://ed.sc.gov/agency/Standards-and-Learning/Academic Standards/old/cso/standards/ela/documents/StateBoardApprovedFinalMay14.doc.

SCDOE. (2007, May 17). *Rex proposes new testing system, recommends unifying state and federal school accountability*. Retrieved from http://www.ed.sc.gov/news/more.cfm?articleID=771

South Carolina Education Oversight Committee (n.d.a). *Increasing the impact of the accountability system: Following the progress of amendments to the Education Accountability Act (Act 282 of 2008*). Retrieved from http://eoc.sc.gov/EAAof1998.htm

Spray, J. A., Ackerman, T. A., Reckase, M. D., & Carlson, J. E. (1989). Effect of the medium of item presentation on examinee performance and item characteristics. *Journal of Educational Measurement, 26*, 261-271.

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state mandated performance-based assessment using oral reading fluency. *School Psychology Review, 30,* 407-419.

Stanchfield, J. (1973). *Sex differences in learning to read*. Bloomington, IN: Phi Delta Kappa Educational Foundations.

Stephenson, E. (2006). Evading the No Child Left Behind act: State strategies and federal complicity. *Brigham Young University Education and Law Journal, 1*, 157-188.

Supovitz, J. (2009). Can high stakes testing leverage educational improvement? prospects from the last decade of testing and accountability reform. *Journal of Educational Change, 10*(2-3), 211-227.

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction, 8*, 351-362.

Sweller, J., & Chandler, P. (1991). Why is some material difficult to learn? *Cognition and Instruction, 12*, 185-233.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston: Allyn & Bacon.

Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I. (1998). *The relationship between computer familiarity and performance on computer*. Retrieved from http://www.ets.org/Media/Research/pdf/RR-98-08.pdf

TestView. (2004). [Computer software]. Columbia, SC: VC3, Inc.

Thompson, B. (2002). “Statistical,” “Practical,” and “Clinical”: How many kinds of significance do counselors need to consider? *Journal of Counseling & Development, 80,* 64-71.

Triplett, C. F., & Barksdale, M. A. (2005). Third through sixth graders' perceptions of high-stakes testing. *Journal of Literacy Research, 37,* 237-260.

Trotter, A. (2001). Testing computerized exams. *Education Week, 20* (37), 30-35. Retrieved from http://www.edweek.org.ew.westory.cfm?slug=37online.h20

U.S. Census Bureau. (2002). *The big payoff: Educational attainment and synthetic estimates of work-life earnings*. Retrieved from http://www.census.gov/prod/2002pubs/p23-210.pdf

U.S. Department of Agriculture. (2008). *Eligibility manual for school meals.* Washington, DC: Author. Retrieved from http://www.fns.usda.gov/cnd/Governance/notices/iegs/EligibilityManual.pdf

U.S. Department of Education. (1983). *A nation at risk: The imperative for school reform*. Washington, DC: Author, Commission on Excellence in Education.

U.S. Department of Education Office of Elementary and Secondary Education. (2002). *Guidance for the reading first program.* Retrieved from http://www.ed.gov/programs/readingfirst/guidance.doc.

Valencia, S. W., & Wixson, K. K. (April 1, 1999). *Policy-oriented research on literacy standards and assessment*. Ann Arbor, MI: Center for the Improvement of Early Reading Achievement. (CIERA Report No. 3-004)

Vander Meer, C. D., Lentz, F. E., & Stollar, S. (2005). *The relationship between oral reading fluency and statewide achievement testing in Ohio* (Technical Report). Eugene, OR: University of Oregon.

Van Haren, B., Russ, S., Chiang, B., & Fiedler, C. (2006). Impacted of revised eligibility criteria on LD incidence rates: A Wisconsin case study. *Proceedings of the 14th World Congress on Learning Disabilities. A Multidisciplinary Approach to Learning Disabilities*, 26-39.

Viadero, D. (2000). Lags in minority achievement: Defy traditional explanations. *Education Week*. Retrieved from http://www.edweek.org/ew/ewstory.cfm?slug=28causes.h19

Villano, M. (2006). Assessing formative assessment. *Technology and Learning, 26*(6), 8-11.

Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates Publishing.

Wang, S. (2004). *Online or paper: Does delivery affect results? Administration mode comparability study for Stanford Diagnostic Reading and Mathematics Tests*. San Antonio, Texas: Harcourt. Retrieved from http://pearsonassess.com/NR/rdonlyres/D381C2EA-18A6-4B52-A5DC-DD0CEC3B0D40/0/OnlineorPaper.pdf

Warren, J. R., Jenkins, K. N., & Kulick, R. B. (2006). High school exit examinations and state-level completion and GED rates, 1975 through 2002. *Educational Evaluation and Policy Analysis*, *28*(2), 131-152.

Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, *41*, 85-120.

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied*

Psychological Measurement, 6, 473-492.

Westat, Inc. (2001). *The longitudinal evaluation of school change and performance (LESCP) in Title I schools. Final report. Volume 1*. Rockville, MD: Author. (ERIC Document Reproduction Service No. ED457305).

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594-604.

Williams, S. (n.d.) *Online Testing*. Retrieved from www.indiana-etc.org/pdfs%5COnline-Testing.pdf

Wise, L. L., Hauser, R. M., Mitchell, K. J., & Feuer, M. J. (1999). *Evaluation of the voluntary national tests*. Washington, DC: National Academy Press.

Yell, M. L., Katsiyannas, A., & Shiner, J. G. (2006). The no child left behind act, adequate yearly progress, and students with disabilities. *Teaching Exceptional Children, 38*(4), 32-29.

Zabala, D., Minnici, A., McMurrer, J., & Briggs, L. (2008). *State high school exit exams: Moving toward end-of-course exams.* Washington, DC: Center on Education Policy. Retrieved from http://www.cep-dc.org/document/docWindow.cfm?fuseaction=document.viewDocument& documentid=244&documentFormatId=3803

Zandvliet, D. B., & Farragher, P. (1997). A comparison of the computer and written forms of multiple choice testing. *Journal of Research on Computing in Education, 29*(4), 423-438.

Zara, A. R. (1992, April). *A comparison of computerized adaptive and paper-and-pencil versions of the national registered nurse licensure examination*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

APPENDIX A

Fort Mill School District #4
Office of Special Services
120 E. Elliott Street
Fort Mill, SC 29715
Phone: 803-548-2527 Fax: 803-548-8273

May 11, 2009

Dr. Keith Callicutt
Superintendent of Schools
Fort Mill School District
120 East Elliott St.
Fort Mill, South Carolina 29715

Dear Dr. Callicutt:

I am currently a doctoral candidate in the Educational and School Psychology program at Indiana University of Pennsylvania. The focus of my dissertation is the use of computer adaptive testing to predict performance on state-mandated reading assessments in third through eighth grade. Establishing further support for the predictive relationship between the Measures of Academic Progress (MAP) to the Palmetto Assessment of State Standards will provide the district with useful information in determining the effectiveness of the MAP as a formative assessment. I am requesting permission to access test data from any student record that has both a 2009 MAP spring reading score and a 2009 PASS ELA score. All test data will be held in strict confidence with the data being encoded by number rather than student name. In addition, to ensure that no personally identifiable information about an individual student will be revealed, all individual test scores will be combined and only group results will be analyzed and reported. In accordance with federal regulations, data will be maintained confidentially for 3 years from completion of the project. At your request, results of the study will be shared with you and other designated school officials. Group results may also be published in Dissertation Abstracts.

Your signature below reflects permission to use the requested data. Thank you in advance for your attention and support of this matter.

Sincerely,

Amy L. Maziarz
Amy L. Maziarz
Coordinator of Special Services

My signature below reflects permission of data collection for the purposes of this study.

V. Keith Callicutt 5-2-09

V. Keith Callicutt Date

APPENDIX B

Cross-tabulated Frequencies and Percentages of Sample Demographics

Sex	SES	Special education	Race/Ethnicity	Grade 3	4	5	6	7	8	Total
Female	Free lunch	No special education	Black	9 (0.2%)	13 (0.3%)	9 (0.2%)	17 (0.4%)	8 (0.2%)	14 (0.4%)	70 (1.8%)
			White	11 (0.3%)	13 (0.3%)	18 (0.5%)	15 (0.4%)	10 (0.3%)	14 (0.4%)	81 (2.1%)
			Other	5 (0.1%)	8 (0.2%)	13 (0.3%)	7 (0.2%)	7 (0.2%)	2 (0.1%)	42 (1.1%)
			Total	25 (0.7%)	34 (0.9%)	40 (1.0%)	39 (1.0%)	25 (0.7%)	30 (0.8%)	193 (5.0%)
		Special education	Black	3 (0.1%)	3 (0.1%)	2 (0.1%)	5 (0.1%)	4 (0.1%)	3 (0.1%)	20 (0.5%)
			White	2 (0.1%)	5 (0.1%)	7 (0.2%)	5 (0.1%)	4 (0.1%)	1 (<0.1%)	24 (0.6%)
			Other	2 (0.1%)	2 (0.1%)	1 (<0.1%)	0 (0.0%)	1 (<0.1%)	1 (<0.1%)	7 (0.2%)
			Total	7 (0.2%)	10 (0.3%)	10 (0.3%)	10 (0.3%)	9 (0.2%)	5 (0.1%)	51 (1.3%)
	Reduced lunch	No special education	Black	3 (0.1%)	3 (0.1%)	1 (<0.1%)	2 (0.1%)	4 (0.1%)	3 (0.1%)	16 (0.4%)
			White	8 (0.2%)	9 (0.2%)	5 (0.1%)	2 (0.1%)	8 (0.2%)	4 (0.1%)	36 (0.9%)
			Other	3 (0.1%)	2 (0.1%)	2 (0.1%)	0 (0.0%)	2 (0.1%)	2 (0.1%)	11 (0.3%)
			Total	14 (0.4%)	14 (0.4%)	88 (0.2%)	4 (0.1%)	14 (0.4%)	9 (0.2%)	63 (1.6%)
		Special education	Black	0 (0.0%)	1 (<0.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (<0.1%)
			White	0 (0.0%)	1 (<0.1%)	2 (0.1%)	2 (0.1%)	0 (0.0%)	1 (<0.1%)	6 (0.2%)
			Other	0 (0.0%)	1 (<0.1%)	0 (0.0%)	1 (<0.1%)	1 (<0.1%)	0 (0.0%)	3 (0.1%)
			Total	0 (0.0%)	3 (0.1%)	2 (0.1%)	3 (0.1%)	1 (<0.1%)	1 (<0.1%)	10 (0.3%)

Sex	SES	Special education	Race/Ethnicity	Grade						
				3	4	5	6	7	8	Total
Female	No free/reduced lunch	No special education	Black	13 (0.3%)	12 (0.3%)	14 (0.4%)	16 (0.4%)	14 (0.5%)	12 (0.3%)	81 (2.1%)
			White	224 (5.8%)	243 (6.3%)	203 (5.3%)	207 (5.4%)	222 (5.8%)	218 (5.7%)	1317 (34.1%)
			Other	23 (0.6%)	16 (0.4%)	14 (0.4%)	9 (0.2%)	16 (0.4%)	15 (0.4%)	93 (2.4%)
			Total	260 (6.7%)	271 (7.0%)	231 (6.0%)	232 (6.0%)	252 (6.5%)	245 (6.4%)	1491 (38.6%)
		Special education	Black	3 (0.1%)	2 (0.1%)	0 (0.0%)	1 (<0.1%)	1 (<0.1%)	1 (<0.1%)	8 (0.2%)
			White	20 (0.5%)	21 (0.5%)	15 (0.4%)	10 (0.3%)	13 (0.3%)	10 (0.3%)	89 (2.3%)
			Other	0 (0.0%)	0 (0.0%)	2 (0.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (0.1%)
			Total	23 (0.6%)	23 (0.6%)	17 (0.4%)	11 (0.3%)	14 (0.3%)	11 (0.3%)	99 (2.6%)
			Female Total	329 (8.5%)	355 (9.2%)	308 (8.0%)	299 (7.7%)	315 (8.2%)	301 (7.8%)	1907 (49.4%)
Male	Free lunch	No special education	Black	15 (0.4%)	15 (0.4%)	15 (0.4%)	9 (0.2%)	10 (0.3%)	8 (0.2%)	72 (1.87%)
			White	29 (0.8%)	24 (0.6%)	17 (0.4%)	13 (0.3%)	19 (0.5%)	14 (0.4%)	116 (3.0%)
			Other	14 (0.4%)	5 (0.1%)	8 (0.2%)	4 (0.1%)	6 (0.2%)	2 (0.1%)	39 (1.0%)
			Total	58 (1.5%)	44 (1.1%)	40 (1.0%)	26 (0.7%)	35 (0.9%)	24 (0.6%)	227 (5.9%)
		Special education	Black	3 (0.1%)	8 (0.2%)	4 (0.1%)	4 (0.1%)	5 (0.1%)	3 (0.1%)	27 (0.7%)
			White	3 (0.1%)	7 (0.2%)	6 (0.2%)	10 (0.3%)	11 (0.3%)	6 (0.2%)	43 (1.1%)
			Other	2 (0.1%)	4 (0.1%)	0 (0.0%)	3 (0.1%)	1 (<0.1%)	1 (<0.1%)	11 (0.3%)
			Total	8 (0.2%)	19 (0.5%)	10 (0.3%)	17 (0.4%)	17 (0.4%)	10 (0.3%)	81 (2.1%)

Sex	SES	Special education	Race/Ethnicity	Grade						
				3	4	5	6	7	8	Total
Male	Reduced lunch	No special education	Black	3 (0.1%)	1 (<0.1%)	2 (0.1%)	4 (0.1%)	1 (<0.1%)	2 (0.1%)	13 (0.3%)
			White	6 (0.2%)	5 (0.1%)	6 (0.2%)	13 (0.3%)	6 (0.2%)	4 (0.1%)	40 (1.0%)
			Other	0 (0.0%)	2 (0.1%)	2 (0.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (0.1%)
			Total	9 (0.2%)	8 (0.2%)	10 (0.3%)	17 (0.4%)	7 (0.2%)	6 (0.2%)	57 (1.5%)
		Special education	Black	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (0.1%)	1 (<0.1%)	1 (<0.1%)	5 (0.1%)
			White	4 (0.1%)	2 (0.1%)	2 (0.1%)	4 (0.1%)	3 (0.1%)	3 (0.1%)	18 (0.5%)
			Other	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
			Total	4 (0.1%)	2 (0.1%)	2 (0.1%)	7 (0.2%)	4 (0.1%)	4 (0.1%)	23 (0.6%)
	No free/reduced lunch	No special education	Black	6 (0.2%)	12 (0.3%)	11 (0.3%)	15 (0.4%)	10 (0.3%)	16 (0.4%)	70 (1.8%)
			White	201 (5.2%)	203 (5.3%)	191 (5.0%)	229 (5.9%)	223 (5.8%)	212 (5.5%)	1259 (32.6%)
			Other	18 (0.5%)	12 (0.3%)	9 (0.2%)	13 (0.3%)	8 (0.2%)	12 (0.3%)	72 (1.9%)
			Total	225 (5.8%)	227 (5.9%)	211 (5.5%)	257 (6.7%)	241 (6.2%)	240 (6.2%)	1401 (36.3%)
		Special education	Black	2 (0.1%)	0 (0.0%)	0 (0.0%)	3 (0.1%)	0 (0.0%)	0 (0.0%)	5 (0.1%)
			White	29 (0.8%)	31 (0.8%)	24 (0.6%)	23 (0.6%)	27 (0.7%)	21 (0.5%)	155 (4.0%)
			Other	1 (<0.1%)	1 (<0.1%)	0 (0.0%)	3 (0.1%)	0 (0.0%)	0 (0.0%)	5 (0.1%)
			Total	32 (0.8%)	32 (0.8%)	24 (0.6%)	29 (0.8%)	27 (0.7%)	21 (0.5%)	165 (4.3%)
			Male Total	336 (8.7%)	332 (8.6%)	297 (7.79%)	353 (9.1%)	331 (8.6%)	305 (7.9%)	1954 (50.6%)
			Grand Total	665 (17.2%)	687 (17.8%)	605 (15.7%)	652 (16.9%)	646 (16.7%)	606 (15.7%)	3861 (100.0%)

APPENDIX C

Descriptive Statistics for MAP and PASS ELA Scores

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
3	Female	Black	Free lunch	No	N	9	9	9	9	9
					Minimum	184.3	189	183	178	605
					Maximum	207.3	218	213	201	698
					Mean	196.4	198.4	198.4	192.3	656.6
					SD	7.9	10.3	9.8	7.3	31.9
				Yes	N	3	3	3	3	3
					Minimum	155	166	156	143	489
					Maximum	188.7	185	193	188	564
					Mean	172.7	176	173	169	533.7
					SD	16.9	9.5	18.7	23.3	39.5
			Reduced Lunch	No	N	3	3	3	3	3
					Minimum	203.3	200	202	192	521
					Maximum	206.3	219	215	204	661
					Mean	204.7	207	208.3	198.7	609.7
					SD	1.5	10.4	6.5	6.1	77.1
			No Free/Reduced Lunch	No	N	13	13	13	13	13
					Minimum	193.7	202	189	190	573
					Maximum	217	226	224	217	741
					Mean	207.5	209.2	206.5	206.8	658.5
					SD	7.2	7.3	10.8	7.3	45
				Yes	N	3	3	3	3	3
					Minimum	165	157	174	164	669
					Maximum	229.3	233	227	228	689
					Mean	187.9	185.7	192	186	677.7
					SD	36	41.3	30.3	36.4	10.3

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
3	Female	White	Free Lunch	No	N	11	11	11	11	11
					Minimum	195.3	192	195	184	600
					Maximum	208	217	211	206	764
					Mean	200.2	203.9	202.6	194.1	652.8
					SD	4.5	8.7	4.8	6.2	43.5
				Yes	N	2	2	2	2	2
					Minimum	172.7	177	173	163	632
					Maximum	175.7	191	174	167	694
					Mean	174.2	184	173.5	165	663
					SD	2.1	9.9	0.7	2.8	43.8
			Reduced Lunch	No	N	8	8	8	8	8
					Minimum	187	187	183	172	605
					Maximum	213.3	222	213	205	741
					Mean	200.5	204.6	200	196.8	658.1
					SD	8.7	9.6	11.5	11.1	50.3
			No Free/Reduced Lunch	No	N	224	224	224	224	224
					Minimum	185.3	184	176	184	513
					Maximum	232.3	236	241	235	832
					Mean	211.1	214	210.1	209	662.3
					SD	9	10.7	10.4	10	48
				Yes	N	20	20	20	20	20
					Minimum	165.7	168	163	166	538
					Maximum	221	227	224	216	754
					Mean	194.7	195.6	196.2	192.4	654.9
					SD	16.9	17.9	18.5	15.3	63.3
		Hispanic	Free Lunch	No	N	5	5	5	5	5
					Minimum	175.3	167	173	181	595
					Maximum	194.7	200	204	193	692
					Mean	189.2	187.2	193.2	187.2	650
					SD	7.9	12.9	12.2	4.4	41.9

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
3	Female	Hispanic	Free Lunch	Yes	N	2	2	2	2	2
					Minimum	160.3	161	158	162	653
					Maximum	192.3	192	192	193	669
					Mean	176.3	176.5	175	177.5	661
					SD	22.6	21.9	24	21.9	11.3
			Reduced Lunch	No	N	2	2	2	2	2
					Minimum	177	172	186	173	643
					Maximum	195	186	203	196	651
					Mean	186	179	194.5	184.5	647
					SD	12.7	9.9	12	16.3	5.7
			No Free/Reduced Lunch	No	N	8	8	8	8	8
					Minimum	189	179	187	189	585
					Maximum	225.7	227	229	227	715
					Mean	207.9	209.3	207.3	207.3	662.5
					SD	12.9	17.8	13.1	11.7	44.8
		Asian or Other Pacific Islander	No Free/Reduced Lunch	No	N	14	14	13	14	14
					Minimum	198.7	205	191	192	573
					Maximum	224.3	228	228	225	712
					Mean	212.7	217.1	209.7	211.4	652.9
					SD	7.7	8.1	9.3	8.8	48.9
	Male	Black	Free Lunch	No	N	15	15	15	15	15
					Minimum	175.3	176	173	177	571
					Maximum	206.3	202	207	211	731
					Mean	192.6	191.1	193.9	192.9	652.1
					SD	8	8.6	9.6	8.9	42.1
				Yes	N	3	3	3	3	3
					Minimum	149.7	155	147	147	582
					Maximum	184.7	184	190	180	656
					Mean	170.2	173.3	173.3	164	625
					SD	18.3	15.9	23.1	16.5	38.4

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
3	Male	Black	Reduced Lunch	No	N	3	3	3	3	3
					Minimum	197	201	196	190	605
					Maximum	208	217	208	208	712
					Mean	202.8	207.7	201.7	199	652.7
					SD	5.5	8.3	6	9	54.4
			No Free/Reduced Lunch	No	N	6	6	6	6	6
					Minimum	192.7	193	198	183	610
					Maximum	217	220	218	215	692
					Mean	203.3	204.3	206.8	198.7	651
					SD	10.5	12	8.2	12.7	34.4
				Yes	N	2	2	2	2	2
					Minimum	181.3	178	184	182	637
					Maximum	193	193	194	192	698
					Mean	187.2	185.5	189	187	667.5
					SD	8.2	10.6	7.1	7.1	43.1
		White	Free Lunch	No	N	29	29	29	29	29
					Minimum	166	164	165	169	556
					Maximum	217	226	220	218	741
					Mean	194.3	196.2	192.2	194.4	648.5
					SD	12.2	12.6	14.3	12.2	44.4
				Yes	N	3	3	3	3	3
					Minimum	159.3	163	157	158	564
					Maximum	197.3	193	209	193	689
					Mean	182.2	181.7	184.7	180.3	643.7
					SD	20.2	16.3	26.2	19.4	69.2
			Reduced Lunch	No	N	6	6	6	6	6
					Minimum	195	196	189	187	598
					Maximum	210	216	206	213	694
					Mean	200.2	204.8	196.8	198.8	652.2
					SD	6.4	8	6.6	8.9	37

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
3	Male	White	Reduced Lunch	Yes	N	4	4	4	4	4
					Minimum	162.3	164	154	169	551
					Maximum	203	212	200	197	689
					Mean	180.3	181.3	181.8	177.8	640.3
					SD	16.8	21.3	19.7	13.1	61.1
			No Free/Reduced Lunch	No	N	201	201	201	201	201
					Minimum	181.3	180	175	177	517
					Maximum	233	237	237	242	832
					Mean	207.2	208.7	207.3	205.5	657.6
					SD	10.2	11.6	11.1	11.8	46.6
				Yes	N	29	29	29	29	29
					Minimum	169.7	168	170	171	544
					Maximum	221	227	229	223	754
					Mean	198.2	198.6	199.6	196.4	653.1
					SD	12.8	15.4	13.3	13.2	53.2
		Hispanic	Free Lunch	No	N	12	12	12	12	12
					Minimum	183.7	179	185	178	487
					Maximum	207	211	210	209	741
					Mean	197	197.8	197.5	195.8	631.3
					SD	7.9	10.1	7.5	9	62.4
			No Free/Reduced Lunch	No	N	9	9	9	9	9
					Minimum	202	202	195	202	580
					Maximum	230	234	234	222	725
					Mean	210.9	216.2	207.2	209.1	667.6
					SD	8.4	9.5	11.6	6.1	50
				Yes	N	1	1	1	1	1
					Minimum	183	192	172	185	689
					Maximum	183	192	172	185	689
					Mean	183	192	172	185	689
					SD					

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
3	Male	Asian or Other Pacific Islander	Free Lunch	No	N	2	2	2	2	2
					Minimum	191.7	198	187	190	590
					Maximum	210.7	209	218	205	616
					Mean	201.2	203.5	202.5	197.5	603
					SD	13.4	7.8	21.9	10.6	18.4
			Reduced Lunch	No	N	7	7	7	7	7
					Minimum	203	203	197	203	513
					Maximum	220.7	221	218	225	712
					Mean	211.7	213	209.6	212.4	636.7
					SD	7.3	8	8.3	8.8	70.5
4	Female	Black	Free Lunch	No	N	13	13	13	13	13
					Minimum	183.7	185	181	173	566
					Maximum	221.3	236	222	216	701
					Mean	204.1	209.3	204.5	198.5	638.6
					SD	9.8	11.8	11.4	11.2	41
				Yes	N	3	3	3	3	3
					Minimum	166.3	166	169	164	523
					Maximum	210.7	214	210	208	589
					Mean	190.6	189.7	193	189	560.7
					SD	22.5	24	21.4	22.6	34
			Reduced Lunch	No	N	3	3	3	3	3
					Minimum	202	197	204	201	566
					Maximum	208.3	214	208	208	584
					Mean	206.2	207	206.3	205.3	578
					SD	3.7	8.9	2.1	3.8	10.4
			No Free/Reduced Lunch	No	N	12	12	12	12	12
					Minimum	194.3	196	187	196	575
					Maximum	218.3	228	219	221	724
					Mean	210.8	214.8	207.4	210.3	663.6
					SD	7.5	9.2	8.9	7.8	38.2

NOTE: Groups with $N \leq 1$ were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
4	Female	Black	No Free/Reduced Lunch	Yes	N	2	2	2	2	2
					Minimum	195	193	194	198	626
					Maximum	206.3	213	206	200	661
					Mean	200.7	203	200	199	643.5
					SD	8	14.1	8.5	1.4	24.7
		White	Free Lunch	No	N	13	13	13	13	13
					Minimum	194.7	195	193	192	535
					Maximum	215	220	212	219	686
					Mean	206.5	209.8	204.2	205.7	631.3
					SD	6.5	7.6	6.2	8.6	44.8
				Yes	N	5	5	5	5	5
					Minimum	179.3	182	179	177	562
					Maximum	201.7	205	199	201	733
					Mean	189.1	192.6	188.8	185.8	628.2
					SD	8.1	8.4	7.2	10	65.3
			Reduced Lunch	No	N	9	9	9	9	9
					Minimum	191.3	190	192	187	556
					Maximum	217	219	218	223	724
					Mean	203.1	203.2	203.2	203	629.3
					SD	9.3	11.3	10.4	12.7	57.9
			No Free/Reduced Lunch	No	N	243	243	243	243	243
					Minimum	174	180	170	172	531
					Maximum	243.3	251	242	242	801
					Mean	217.7	220	217.4	215.6	660
					SD	9.1	10.3	9.9	10.9	51.4
				Yes	N	21	21	21	21	21
					Minimum	164.7	168	154	172	556
					Maximum	217.3	224	219	220	712
					Mean	197.3	200.6	194.7	196.7	639.4
					SD	13.4	13.6	15.4	14.5	41.4

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
4	Female	Hispanic	Free Lunch	No	N	6	6	6	6	6
					Minimum	169.7	177	168	164	610
					Maximum	211.7	217	209	210	767
					Mean	198.9	203.5	198.2	195.2	669.7
					SD	15.4	14.7	15.7	16.7	58.6
			No Free/Reduced Lunch	No	N	6	6	6	6	6
					Minimum	207	212	203	206	594
					Maximum	234.7	246	227	231	692
					Mean	220.9	225.7	216.5	220.5	638.8
					SD	9.4	12.5	8.4	10.3	35.7
		Asian or Other Pacific Islander	Free Lunch	No	N	2	2	2	2	2
					Minimum	199.7	202	193	204	622
					Maximum	217	211	218	222	637
					Mean	208.3	206.5	205.5	213	629.5
					SD	12.3	6.4	17.7	12.7	10.6
			No Free/Reduced Lunch	No	N	10	10	10	10	10
					Minimum	182	180	192	174	575
					Maximum	228.3	225	228	232	710
					Mean	210	211.9	210.7	207.4	656.9
					SD	13.9	13.4	11.6	17.8	42.1
	Male	Black	Free Lunch	No	N	15	15	15	15	15
					Minimum	190.7	182	196	187	531
					Maximum	214.3	216	217	214	712
					Mean	202.3	202	202.8	202.1	624.9
					SD	7.2	10.6	6.1	8.7	41.9
				Yes	N	8	8	8	8	8
					Minimum	154	152	155	152	581
					Maximum	197	192	204	202	651
					Mean	177.6	174.3	180.3	178.3	610.1
					SD	16.7	17	17.3	17.3	28.8

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
4	Male	Black	No Free/Reduced Lunch	No	N	12	12	12	12	12
					Minimum	194.7	198	192	189	616
					Maximum	221	224	223	221	741
					Mean	212	213.6	211.6	210.8	659.8
					SD	8.5	8.6	9	10	35.7
		White	Free Lunch	No	N	24	24	24	24	24
					Minimum	178.7	175	180	181	548
					Maximum	218.3	225	229	223	725
					Mean	203.8	204.4	205.2	201.8	635.1
					SD	10.2	11.6	11.9	10.6	44.4
				Yes	N	7	7	7	7	7
					Minimum	174.3	179	171	167	531
					Maximum	214.7	212	211	223	832
					Mean	191.7	193.1	191	190.9	669.6
					SD	14.9	13.9	17.2	17.7	93
			Reduced Lunch	No	N	5	5	5	5	5
					Minimum	205.3	203	206	200	631
					Maximum	215.3	225	212	215	739
					Mean	209.7	212.2	209.4	207.6	682.4
					SD	4	8.6	2.8	6.7	40.5
				Yes	N	2	2	2	2	2
					Minimum	203.7	198	209	203	528
					Maximum	211	209	221	204	605
					Mean	207.3	203.5	215	203.5	566.5
					SD	5.2	7.8	8.5	0.7	54.4
			No Free/Reduced Lunch	No	N	203	203	203	203	203
					Minimum	190.3	189	189	186	531
					Maximum	242.3	254	239	238	832
					Mean	216.2	218.8	215.7	214.1	660.4
					SD	8.8	11	10	9.8	47.6

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
4	Male	White	No Free/Reduced Lunch	Yes	N	31	31	31	31	31
					Minimum	162	160	159	167	575
					Maximum	227.7	231	226	226	715
					Mean	199.6	200.7	199.5	198.5	651.3
					SD	18.6	20.4	19.1	18.7	39.3
		Hispanic	Free Lunch	No	N	5	5	5	5	5
					Minimum	199.7	197	200	198	535
					Maximum	210.7	214	216	209	701
					Mean	205.1	206.4	205.6	203.2	619.6
					SD	4.1	6.9	6.4	4.3	65.3
				Yes	N	4	4	4	4	4
					Minimum	181	186	182	170	612
					Maximum	212.7	224	209	214	658
					Mean	197.3	204	196.3	191.5	636.8
					SD	17.8	18.4	14.8	21.3	18.9
			Reduced Lunch	No	N	2	2	2	2	2
					Minimum	200.7	200	209	193	605
					Maximum	210	215	212	203	651
					Mean	205.3	207.5	210.5	198	628
					SD	6.6	10.6	2.1	7.1	32.5
			No Free/Reduced Lunch	No	N	4	4	4	4	4
					Minimum	210	215	210	203	651
					Maximum	220.7	227	223	223	741
					Mean	216.3	221.5	214.3	213	690.3
					SD	4.8	5	5.9	8.2	39.7
		Asian or Other Pacific Islander	No Free/Reduced Lunch	No	N	6	6	6	6	6
					Minimum	209.7	207	211	205	541
					Maximum	229	233	227	229	701
					Mean	220	221.7	220.7	217.7	635.7
					SD	7	10.1	5.9	8.5	65.8

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
5	Female	Black	Free Lunch	No	N	9	9	9	9	9
					Minimum	179.7	181	172	186	593
					Maximum	221.3	228	223	226	689
					Mean	204.4	205.8	202.8	204.8	637.4
					SD	13.9	15.6	15.6	13.5	28.3
				Yes	N	2	2	2	2	2
					Minimum	191.3	195	196	183	675
					Maximum	206	211	213	194	688
					Mean	198.7	203	204.5	188.5	681.5
					SD	10.4	11.3	12	7.8	9.2
			No Free/Reduced Lunch	No	N	14	14	14	14	14
					Minimum	208	201	202	201	605
					Maximum	237.7	244	237	234	801
					Mean	219.7	224.6	218.1	216.4	670.7
					SD	9.2	12.8	10.1	9.5	53
		White	Free Lunch	No	N	18	18	18	18	18
					Minimum	178.3	178	179	178	555
					Maximum	221.3	224	228	221	790
					Mean	207.2	209.1	207.4	205.1	641.7
					SD	10	12.6	10.2	10.8	57.4
				Yes	N	7	7	7	7	7
					Minimum	174.7	182	172	170	591
					Maximum	205.7	210	205	206	789
					Mean	195.6	198.4	193	195.3	673.3
					SD	11.6	11.2	12.5	12.2	70.2
			Reduced Lunch	No	N	5	5	5	5	5
					Minimum	211	213	204	207	497
					Maximum	219.3	221	219	224	741
					Mean	214.1	216.8	212.4	213.2	616
					SD	3.4	3	6.5	7.2	91.6

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
5	Female	White	Reduced Lunch	Yes	N	2	2	2	2	2
					Minimum	186.7	182	193	185	701
					Maximum	203.3	204	206	200	764
					Mean	195	193	199.5	192.5	732.5
					SD	11.8	15.6	9.2	10.6	44.5
			No Free/Reduced Lunch	No	N	203	203	203	203	203
					Minimum	198.3	197	193	196	525
					Maximum	254	259	251	265	849
					Mean	222.3	225.2	221.2	220.5	655.7
					SD	8.8	10.6	9.7	10.4	49.8
				Yes	N	15	15	15	15	15
					Minimum	163.7	161	156	160	562
					Maximum	231.7	238	234	230	741
					Mean	204	207.3	204.3	200.5	655.1
					SD	21	23.2	23.1	19.7	57.6
		Hispanic	Free Lunch	No	N	12	12	12	12	12
					Minimum	191.7	186	199	190	551
					Maximum	218.7	225	225	217	702
					Mean	208.2	210.9	209.3	204.3	637.5
					SD	7.5	10.6	8.4	8	44.1
			Reduced Lunch	No	N	2	2	2	2	2
					Minimum	205.7	220	199	198	551
					Maximum	216.7	224	212	214	639
					Mean	211.2	222	205.5	206	595
					SD	7.8	2.8	9.2	11.3	62.2
			No Free/Reduced Lunch	No	N	7	7	7	7	7
					Minimum	189.3	196	186	186	610
					Maximum	231.3	234	231	229	725
					Mean	215.6	219.9	214	212.9	675.1
					SD	13.3	12.2	14.9	13.7	41.7

NOTE: Groups with $N \leq 1$ were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
5	Female	Asian or Other Pacific Islander	Reduced Lunch	No	N	7	7	7	7	7
					Minimum	217.3	218	211	217	556
					Maximum	233.3	237	237	235	741
					Mean	226.9	227.1	226.7	226.9	657.6
					SD	6.9	7	9.6	7.5	67.3
	Male	Black	Free Lunch	No	N	15	15	15	15	15
					Minimum	178.7	182	179	175	598
					Maximum	222	226	228	225	741
					Mean	209.2	211.5	209.8	206.5	649.1
					SD	11.3	11.7	12.5	12.9	41.9
				Yes	N	4	4	4	4	4
					Minimum	147	140	150	151	584
					Maximum	200.7	205	200	197	701
					Mean	174	174.3	175.3	172.5	633
					SD	25.6	28.4	26.4	22.9	56.5
			Reduced Lunch	No	N	2	2	2	2	2
					Minimum	207.3	220	204	198	618
					Maximum	222	223	229	214	639
					Mean	214.7	221.5	216.5	206	628.5
					SD	10.4	2.1	17.7	11.3	14.8
			No Free/Reduced Lunch	No	N	11	11	11	11	11
					Minimum	189.3	192	187	186	524
					Maximum	224.7	226	225	223	731
					Mean	207.5	208.6	206.8	207.1	643.6
					SD	9.4	10.1	9.8	9.9	61.7
		White	Free Lunch	No	N	17	17	17	17	17
					Minimum	196.7	198	194	198	546
					Maximum	224	233	229	227	741
					Mean	215.1	217.2	215.4	212.7	644.2
					SD	6.8	9	8.5	9.1	41.9

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
5	Male	White	Free Lunch	Yes	N	6	6	6	6	6
					Minimum	189	193	194	180	582
					Maximum	223.3	233	229	225	715
					Mean	211.3	216.5	211	206.5	632.3
					SD	11.9	13.3	11.5	14.7	51.1
			Reduced Lunch	No	N	6	6	6	6	6
					Minimum	207.3	199	209	213	605
					Maximum	224.3	239	220	229	741
					Mean	216.2	214.7	213	221	656
					SD	6.2	13.8	4.5	5.1	53.8
				Yes	N	2	2	2	2	2
					Minimum	199.7	198	196	205	675
					Maximum	206	205	201	212	692
					Mean	202.8	201.5	198.5	208.5	683.5
					SD	4.5	4.9	3.5	4.9	12
			No Free/Reduced Lunch	No	N	191	191	191	191	191
					Minimum	196	193	188	198	551
					Maximum	246.7	251	250	248	795
					Mean	221.6	223.6	221.3	220	658.8
					SD	9.1	11.6	10.3	9.6	42.4
				Yes	N	24	24	24	24	24
					Minimum	177.7	174	179	177	568
					Maximum	227.7	231	230	228	741
					Mean	204.2	203.5	205.2	204	660.1
					SD	13	13.8	14.5	13.3	48.7
		Hispanic	Free Lunch	No	N	8	8	8	8	8
					Minimum	199.7	201	205	191	584
					Maximum	216	220	218	214	731
					Mean	208.3	211.3	210.6	203.1	633.5
					SD	5.6	8.3	4.9	9.1	46.3

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
5	Male	Hispanic	Reduced Lunch	No	N	2	2	2	2	2
					Minimum	204.7	202	203	201	622
					Maximum	209.7	222	211	204	665
					Mean	207.2	212	207	202.5	643.5
					SD	3.5	14.1	5.7	2.1	30.4
			No Free/Reduced Lunch	No	N	7	7	7	7	7
					Minimum	202.7	195	203	208	616
					Maximum	238.7	233	242	243	718
					Mean	219.3	219.1	218.1	220.6	645.6
					SD	12.6	15	13.2	13	35.8
6	Female	Black	Free Lunch	No	N	17	17	17	17	17
					Minimum	197	205	194	189	559
					Maximum	224	228	225	226	678
					Mean	212.8	214.6	210.9	212.8	631.9
					SD	7.7	6.8	8.5	10.7	35.8
				Yes	N	5	5	5	5	5
					Minimum	184.3	193	187	171	535
					Maximum	210.3	208	217	206	715
					Mean	197.2	200.2	198.8	192.6	607.8
					SD	9.8	6.4	11.1	14.4	77.1
			Reduced Lunch	No	N	2	2	2	2	2
					Minimum	208.7	213	201	212	668
					Maximum	222	220	222	224	683
					Mean	215.3	216.5	211.5	218	675.5
					SD	9.4	4.9	14.8	8.5	10.6
			Free/Reduced Lunch	No	N	16	16	16	16	16
					Minimum	198	195	200	194	579
					Maximum	248	259	245	243	774
					Mean	222.1	224.4	221.3	220.6	663.1
					SD	14.1	16.5	13.9	15	47.9

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
6	Female	White	Free Lunch	No	N	15	15	15	15	15
					Minimum	199	202	187	201	528
					Maximum	226.3	235	228	236	731
					Mean	215.2	218.3	213.5	213.9	631
					SD	8.3	9.9	10.3	10.6	49
				Yes	N	5	5	5	5	5
					Minimum	196.3	194	197	196	605
					Maximum	215	217	214	224	688
					Mean	205.9	205.2	205.6	207	641.4
					SD	7.5	8.9	6.1	11.4	34.6
			Reduced Lunch	No	N	2	2	2	2	2
					Minimum	218.7	220	224	211	600
					Maximum	224.3	221	228	225	712
					Mean	221.5	220.5	226	218	656
					SD	4	0.7	2.8	9.9	79.2
				Yes	N	2	2	2	2	2
					Minimum	202.7	204	202	202	628
					Maximum	211	212	212	209	643
					Mean	206.8	208	207	205.5	635.5
					SD	5.9	5.7	7.1	4.9	10.6
			No Free/Reduced Lunch	No	N	207	207	207	207	207
					Minimum	195.7	200	189	198	489
					Maximum	249	256	252	252	832
					Mean	226.9	228.1	226	226.7	656.8
					SD	8.9	10.3	11	9.7	54.6
				Yes	N	10	10	10	10	10
					Minimum	190.3	190	192	186	591
					Maximum	229.7	239	229	231	715
					Mean	212.1	212.6	210.7	213	652
					SD	13.7	16.4	11.9	15.5	44.2

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
6	Female	Hispanic	Free Lunch	No	N	7	7	7	7	7
					Minimum	199.7	199	194	206	561
					Maximum	218.3	222	222	218	698
					Mean	211.2	211.3	209.6	212.9	631.9
					SD	7.2	9.3	9.1	4.8	52.1
			No Free/Reduced Lunch	No	N	3	3	3	3	3
					Minimum	219.7	217	214	222	605
					Maximum	231	233	233	227	670
					Mean	223.6	224.7	221.7	224.3	643.7
					SD	6.4	8	10	2.5	34.2
		Asian or Other Pacific Islander	No Free/Reduced Lunch	No	N	5	5	5	5	5
					Minimum	222	221	207	220	620
					Maximum	236	234	239	238	712
					Mean	227.7	229.6	224.2	229.4	680.2
					SD	5.8	5	12.4	6.5	38.2
	Male	Black	Free Lunch	No	N	9	9	9	9	9
					Minimum	202.3	199	198	196	578
					Maximum	223.7	222	228	226	741
					Mean	213.5	213.9	212.7	213.9	646.8
					SD	7	7.1	8.2	9.1	57
				Yes	N	4	4	4	4	4
					Minimum	181	186	183	174	556
					Maximum	205.7	208	199	211	627
					Mean	196.3	198.8	193.5	196.5	586
					SD	10.8	9.2	7.3	16.1	30.7
			Reduced Lunch	No	N	4	4	4	4	4
					Minimum	199.3	201	193	204	595
					Maximum	220.3	221	218	224	681
					Mean	209.9	210.8	204.3	214.8	651.8
					SD	8.6	8.4	10.6	9.8	40.1

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
6	Male	Black	Reduced Lunch	Yes	N	3	3	3	3	3
					Minimum	189	187	193	187	600
					Maximum	206.7	212	210	206	632
					Mean	199.7	201	200	198	614
					SD	9.4	12.8	8.9	9.8	16.4
			No Free/Reduced Lunch	No	N	15	15	15	15	15
					Minimum	194	196	190	193	618
					Maximum	234.3	247	241	232	713
					Mean	216	216.7	214.4	216.8	654.7
					SD	12	15.2	13.2	12	25.5
				Yes	N	3	3	3	3	3
					Minimum	190.7	197	193	182	675
					Maximum	221.3	226	217	221	702
					Mean	208.1	212.3	207.7	204.3	689.7
					SD	15.8	14.6	12.9	20.1	13.7
		White	Free Lunch	No	N	13	13	13	13	13
					Minimum	189.7	191	190	188	546
					Maximum	225	230	226	230	724
					Mean	215.7	218	212.5	216.5	636.6
					SD	9.7	11.5	10.4	11.3	52.8
				Yes	N	10	10	10	10	10
					Minimum	176	173	173	178	580
					Maximum	234.3	241	233	229	661
					Mean	201.4	200.9	203.1	200.3	625.1
					SD	17.8	19	19.4	17.6	32.1
			Reduced Lunch	No	N	13	13	13	13	13
					Minimum	191.3	188	190	196	571
					Maximum	223.3	236	228	226	761
					Mean	212.8	210.9	212	215.4	654.8
					SD	8.6	12.1	10.8	7.7	59.1

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
6	Male	White	Reduced Lunch	Yes	N	4	4	4	4	4
					Minimum	197	192	199	200	577
					Maximum	217.7	214	220	221	710
					Mean	209.8	205.8	210	213.8	641.3
					SD	9.5	9.9	9.7	9.5	57.3
			No Free/Reduced Lunch	No	N	229	229	229	229	229
					Minimum	192	189	186	187	533
					Maximum	246.3	254	249	249	795
					Mean	224.1	224.6	223.6	224.2	653.4
					SD	10	11	11.6	10.9	47.6
				Yes	N	23	23	23	23	23
					Minimum	167	161	170	170	546
					Maximum	225.3	229	234	231	741
					Mean	205	204.4	204.7	205.9	640.2
					SD	18.3	20.9	17.4	18.3	56.1
		Hispanic	Free Lunch	No	N	3	3	3	3	3
					Minimum	204	199	202	206	630
					Maximum	217	225	212	214	683
					Mean	210.2	212.3	207	211.3	650
					SD	6.5	13	5	4.6	28.8
				Yes	N	3	3	3	3	3
					Minimum	181.7	182	183	180	559
					Maximum	210.3	217	212	215	767
					Mean	199.9	202.3	198	199.3	643.7
					SD	15.8	18.2	14.5	17.8	109.3
			No Free/Reduced Lunch	No	N	9	9	9	9	9
					Minimum	174.7	184	167	173	535
					Maximum	230.3	238	227	233	698
					Mean	215.2	220.9	209.6	215.2	647.2
					SD	16	16.1	17.5	17.1	47.9

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
6	Male	Asian or Other Pacific Islander	No Free/Reduced Lunch	No	N	3	3	3	3	3
					Minimum	173.7	177	172	172	633
					Maximum	231.7	228	222	245	675
					Mean	207.4	206	203	213.3	647.7
					SD	30.2	26.2	27.1	37.4	23.7
				Yes	N	2	2	2	2	2
					Minimum	203	202	198	209	588
					Maximum	206.7	210	199	211	694
					Mean	204.8	206	198.5	210	641
					SD	2.6	5.7	0.7	1.4	75
7	Female	Black	Free Lunch	No	N	8	8	8	8	8
					Minimum	208	206	207	201	535
					Maximum	227	235	226	232	741
					Mean	218.1	220.4	217.1	216.8	630.5
					SD	7.5	9.5	6.7	10.5	62.5
				Yes	N	4	4	4	4	4
					Minimum	191.3	190	190	194	556
					Maximum	202.7	202	207	201	741
					Mean	198.1	196.5	199.5	198.3	630.5
					SD	4.9	4.9	8.3	3	78.4
			Reduced Lunch	No	N	4	4	4	4	4
					Minimum	215	212	215	213	610
					Maximum	226.7	230	223	227	741
					Mean	221	222.8	219.5	220.8	673.8
					SD	4.8	7.6	3.3	5.9	62.7
			No Free/Reduced Lunch	No	N	14	14	14	14	14
					Minimum	213.7	207	210	215	548
					Maximum	242	251	242	241	702
					Mean	230	230.2	230.4	229.5	645.2
					SD	7.9	11.6	8.6	7.4	48.5

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
7	Female	White	Free Lunch	No	N	10	10	10	10	10
					Minimum	200	190	204	202	571
					Maximum	228	236	236	227	701
					Mean	218.8	218.4	219.3	218.6	639.8
					SD	10.6	15.3	11.1	8.7	38.8
				Yes	N	4	4	4	4	4
					Minimum	203	200	198	206	616
					Maximum	233.3	235	217	248	661
					Mean	215.6	219.5	206.8	220.5	638.8
					SD	14.2	17.3	8.2	19.3	19.2
			Reduced Lunch	No	N	8	8	8	8	8
					Minimum	200.7	198	201	203	575
					Maximum	231.7	235	236	235	731
					Mean	219.6	217.6	219.9	221.4	647.6
					SD	10.2	12.3	11.9	11.5	55.4
			No Free/Reduced Lunch	No	N	222	222	222	222	222
					Minimum	175.7	174	170	183	499
					Maximum	250.7	258	254	255	849
					Mean	229.6	230.6	229.5	228.7	657.7
					SD	9.8	11.3	11.1	10.8	51.3
				Yes	N	13	13	13	13	13
					Minimum	199.3	203	198	195	566
					Maximum	239.3	238	239	248	754
					Mean	217.6	219.8	216.7	216.3	652.3
					SD	13	12.2	12.6	16.9	53.6
		Hispanic	Free Lunch	No	N	5	5	5	5	5
					Minimum	204	203	205	201	584
					Maximum	229.3	221	235	232	724
					Mean	213.3	213.2	216.4	210.4	653
					SD	10	7.6	11.3	13	59.4

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
7	Female	Hispanic	Reduced Lunch	No	N	2	2	2	2	2
					Minimum	212.3	213	209	208	585
					Maximum	212.7	216	217	212	689
					Mean	212.5	214.5	213	210	637
					SD	0.2	2.1	5.7	2.8	73.5
			No Free/Reduced Lunch	No	N	12	12	12	12	12
					Minimum	216	207	216	216	580
					Maximum	236.3	242	241	233	741
					Mean	227.1	227.4	227.8	226	660.8
					SD	6.3	9.7	8	5.7	49
		Asian or Other Pacific Islander	Free Lunch	No	N	2	2	2	2	2
					Minimum	216.3	217	217	215	600
					Maximum	227	236	223	222	679
					Mean	221.7	226.5	220	218.5	639.5
					SD	7.5	13.4	4.2	4.9	55.9
			No Free/Reduced Lunch	No	N	4	4	4	4	4
					Minimum	218.7	221	226	209	643
					Maximum	233.7	242	238	231	741
					Mean	228.1	231	230.8	222.5	694.5
					SD	7.1	8.7	5.9	9.4	49.7
	Male	Black	Free Lunch	No	N	10	10	10	10	10
					Minimum	201.7	203	193	208	492
					Maximum	227.3	231	232	233	718
					Mean	214.3	214.6	212.1	216.2	616.5
					SD	7.1	7.4	11.4	8.3	65.1
				Yes	N	5	5	5	5	5
					Minimum	181	180	183	180	568
					Maximum	211	214	208	214	651
					Mean	200.2	200.6	199.2	200.8	620.4
					SD	11.9	13.8	10.1	12.9	33

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
7	Male	Black	No Free/Reduced Lunch	No	N	10	10	10	10	10
					Minimum	198.3	200	197	198	547
					Maximum	242.3	239	244	244	675
					Mean	221.8	224.3	220.2	221	615.9
					SD	12.8	11.2	13.2	16.1	41.7
		White	Free Lunch	No	N	19	19	19	19	19
					Minimum	208.7	204	202	205	517
					Maximum	231.3	246	232	230	681
					Mean	218.3	217.9	218.8	218.1	627.3
					SD	6.6	10.3	8.5	7.1	44
				Yes	N	11	11	11	11	11
					Minimum	180.7	180	176	179	530
					Maximum	225.7	222	219	236	754
					Mean	200.8	198.7	202.5	201.3	651
					SD	13	14.3	13	15.9	53.5
			Reduced Lunch	No	N	6	6	6	6	6
					Minimum	209.3	215	205	189	551
					Maximum	228	234	230	232	661
					Mean	220.4	225	218.5	217.8	602
					SD	6.3	7.4	8.5	16.2	39.6
				Yes	N	3	3	3	3	3
					Minimum	214	222	213	207	605
					Maximum	224.3	228	220	225	694
					Mean	218.8	225.7	217.7	213	643
					SD	5.2	3.2	4	10.4	45.9
			No Free/Reduced Lunch	No	N	223	223	223	223	223
					Minimum	200.7	195	194	205	518
					Maximum	252.7	263	252	262	832
					Mean	228.8	229.6	228	228.6	659.6
					SD	9.4	11.3	10.9	9.8	53.5

NOTE: Groups with $N \leq 1$ were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
7	Male	White	No Free/Reduced Lunch	Yes	N	27	27	27	27	27
					Minimum	182.3	185	180	182	551
					Maximum	237.3	238	242	235	731
					Mean	213.7	214.3	212.5	214.1	629.2
					SD	11.9	12.8	14.4	11.3	50
		Hispanic	Free Lunch	No	N	6	6	6	6	6
					Minimum	182	178	188	180	585
					Maximum	220.7	219	232	226	678
					Mean	209.2	206.3	211.3	209.8	633.7
					SD	14.7	15.5	14.7	16.4	38.9
			No Free/Reduced Lunch	No	N	7	7	7	7	7
					Minimum	212	208	207	207	566
					Maximum	241.7	243	249	237	689
					Mean	227.3	229.1	228.3	224.4	635.3
					SD	11.4	13	13.8	9.7	47.4
8	Female	Black	Free Lunch	No	N	14	14	14	14	14
					Minimum	198.3	198	195	201	578
					Maximum	224	230	229	233	698
					Mean	214.2	215.7	212.1	214.9	649.1
					SD	7.7	10.4	8.3	8.7	35
				Yes	N	3	3	3	3	3
					Minimum	210.7	207	209	204	600
					Maximum	213.7	217	213	225	702
					Mean	212.6	212.7	211	214	666.7
					SD	1.6	5.1	2	10.5	57.8
			Reduced Lunch	No	N	3	3	3	3	3
					Minimum	222.7	226	216	222	612
					Maximum	230	238	229	228	663
					Mean	227.4	232.3	224.7	225.3	635.3
					SD	4.1	6	7.5	3.1	25.8

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
8	Female	Black	No Free/Reduced Lunch	No	N	12	12	12	12	12
					Minimum	225	225	222	223	575
					Maximum	245.7	251	246	247	701
					Mean	235.1	235.4	234.4	235.3	643.8
					SD	5.7	9.1	6.4	7.1	39.9
		White	Free Lunch	No	N	14	14	14	14	14
					Minimum	215	213	203	210	600
					Maximum	230	235	232	229	741
					Mean	222.9	225.9	222.3	220.4	659.7
					SD	5.9	6.3	9.2	5.8	36.3
			Reduced Lunch	No	N	4	4	4	4	4
					Minimum	222.3	210	215	228	570
					Maximum	230.7	228	229	237	774
					Mean	226	222.8	223.3	232	690.8
					SD	3.8	8.5	6.9	4.7	87
			No Free/Reduced Lunch	No	N	218	218	218	218	218
					Minimum	205	196	201	202	511
					Maximum	254.7	261	255	263	801
					Mean	234.1	234.8	233.7	233.9	657
					SD	9.4	10.9	10.6	10.6	46.3
				Yes	N	10	10	10	10	10
					Minimum	206.3	204	201	207	561
					Maximum	236	253	239	230	701
					Mean	218.9	221.4	218.2	217.2	641.6
					SD	10.1	16.2	10.9	7.1	49.1
		Hispanic	Free Lunch	No	N	2	2	2	2	2
					Minimum	208.7	201	213	212	600
					Maximum	223.3	218	222	230	761
					Mean	216	209.5	217.5	221	680.5
					SD	10.4	12	6.4	12.7	113.8

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
8	Female	Hispanic	Reduced Lunch	No	N	2	2	2	2	2
					Minimum	226	225	224	229	568
					Maximum	229.3	225	227	236	610
					Mean	227.7	225	225.5	232.5	589
					SD	2.4	0	2.1	4.9	29.7
			No Free/Reduced Lunch	No	N	7	7	7	7	7
					Minimum	219.3	216	220	222	561
					Maximum	248	252	244	249	712
					Mean	238.4	241.4	235.1	238.6	647.7
					SD	9	12.2	7.8	8.9	49.7
		Asian or Other Pacific Islander	No Free/Reduced Lunch	No	N	8	8	8	8	8
					Minimum	226	228	217	226	610
					Maximum	252.3	258	252	247	733
					Mean	238	242	234.9	237.3	653.9
					SD	9.9	10.3	13.9	7	42.3
	Male	Black	Free Lunch	No	N	8	8	8	8	8
					Minimum	201	203	199	201	551
					Maximum	231.3	236	236	235	731
					Mean	218.2	218.5	217.3	218.8	646.1
					SD	11	11.5	13.3	11.3	74.1
				Yes	N	3	3	3	3	3
					Minimum	197.3	203	193	196	518
					Maximum	207.7	211	206	214	653
					Mean	203.9	205.7	201	205	595.3
					SD	5.7	4.6	7	9	69.6
			Reduced Lunch	No	N	2	2	2	2	2
					Minimum	212	211	210	215	561
					Maximum	221.3	221	223	220	653
					Mean	216.7	216	216.5	217.5	607
					SD	6.6	7.1	9.2	3.5	65.1

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
8	Male	Black	No Free/Reduced Lunch	No	N	16	16	16	16	16
					Minimum	196	189	195	204	541
					Maximum	242	250	242	239	683
					Mean	225	226.1	225	224	622.9
					SD	11.2	14.3	12.6	10.6	46.5
		White	Free Lunch	No	N	14	14	14	14	14
					Minimum	212.3	207	204	210	575
					Maximum	233.7	237	235	242	702
					Mean	223.2	225.9	220.2	223.6	641.1
					SD	5.9	8.4	7	8.2	36
				Yes	N	6	6	6	6	6
					Minimum	178	181	174	179	616
					Maximum	227.3	232	220	240	671
					Mean	209	207.8	203.3	215.8	651
					SD	20.9	21.7	18.1	24.1	22
			Reduced Lunch	No	N	4	4	4	4	4
					Minimum	219.7	221	217	217	610
					Maximum	229.3	229	232	229	661
					Mean	224.6	226	224	223.8	626.8
					SD	4.2	3.5	6.3	5.7	23.8
				Yes	N	3	3	3	3	3
					Minimum	212.3	216	210	206	578
					Maximum	224	227	225	220	646
					Mean	217	220.3	216.7	214	622.7
					SD	6.2	5.9	7.6	7.2	38.7
			No Free/Reduced Lunch	No	N	212	212	212	212	212
					Minimum	186	179	184	195	535
					Maximum	253.7	263	258	258	851
					Mean	232.3	232.6	232.5	231.9	654.6
					SD	10.3	12.3	11.4	11.1	50.8

NOTE: Groups with N ≤ 1 were not included in table.

Grade	Sex	Race/Ethnicity	SES	Special Education		MAP Reading	MAP Literary Texts	MAP Informational Texts	MAP Building Vocabulary	PASS ELA
8	Male	White	No Free/Reduced Lunch	Yes	N	21	21	21	21	21
					Minimum	198.3	190	197	194	489
					Maximum	240	238	244	241	692
					Mean	217	218.4	217.2	215.5	637.7
					SD	11.8	13	12.2	12.8	46.4
		Hispanic	No Free/Reduced Lunch	No	N	7	7	7	7	7
					Minimum	216.3	222	208	217	626
					Maximum	240.3	239	242	251	712
					Mean	226.8	228.7	222.6	229	670
					SD	9.1	6.9	12.8	12.2	34.6
		Asian or Other Pacific Islander	No Free/Reduced Lunch	No	N	5	5	5	5	5
					Minimum	219.7	224	218	208	553
					Maximum	249	254	246	251	741
					Mean	236.5	241	234.6	233.8	635.4
					SD	13.3	12.4	11.7	17.7	69.5

NOTE: Groups with N ≤ 1 were not included in table.