

USING BEGINNING-OF-YEAR DIAGNOSTIC READING MEASURES TO PREDICT
THIRD GRADE COMPREHENSION SCORES IN VIRGINIA

by

Todd Allen Johnson

Liberty University

A Dissertation Presented in Partial Fulfillment

Of the Requirements for the Degree

Doctor of Education

Liberty University

July 2014

USING BEGINNING-OF-YEAR DIAGNOSTIC READING MEASURES TO PREDICT
THIRD GRADE COMPREHENSION SCORES IN VIRGINIA

by Todd Allen Johnson

A Dissertation Presented in Partial Fulfillment

Of the Requirements for the Degree

Doctor of Education

Liberty University, Lynchburg, VA

July 2014

APPROVED BY:

Karen L. Parker, Ed.D., Committee Chair

Meredith Park Furrow, Ed.D., Committee Member

Edward J. Brantmeier, Ph.D., Committee Member

Scott B. Watson Ph.D., Associate Dean, Advanced Programs

ABSTRACT

USING BEGINNING-OF-YEAR DIAGNOSTIC READING MEASURES TO PREDICT THIRD GRADE COMPREHENSION SCORES IN VIRGINIA

The purpose of this correlational investigation is to examine the magnitude of the relationships between three diagnostic reading measures and scores obtained on the third grade Virginia SOL reading test. Specifically, this study investigated the relative utility of predictor variables against one another while also evaluating the strength of the model utilizing all three variables simultaneously. Grounded in the Simple View of Reading as outlined by Gough and Tunmer (1986), this study incorporated the Phonological Awareness Literacy Screening (or PALS), the STAR Reading Test, and the QRI-5 as independent variables. While each instrument has been the subject of previous investigations, no prior study has compared the value of these commonly used assessments in predicting the same outcome measure. This study incorporated a multiple regression analysis to investigate two research questions designed to evaluate the ability of diagnostic reading measures to predict outcome scores on the third grade Virginia Standards of Learning reading assessment. This investigation utilized a data set from a sample of third grade students attending a semi-rural school division in Virginia. Pearson correlation coefficients revealed strong relationships between each of the independent variables and scores obtained on the third grade SOL test. Further, regression analysis revealed a significant relationship between a linear combination of predictor variables and SOL achievement scores. These findings suggest that students can be effectively screened for inclusion in remedial and enrichment programs prior to the administration of high stakes end-of-course reading tests.

Keywords: Reading Comprehension, Assessment, Simple View of Reading, Phonological Awareness, Decoding, Vocabulary, PALS, STAR Reading, QRI, Standards of Learning (SOL)

© Copyright 2014 by Todd Allen Johnson

All Rights Reserved

Acknowledgements

This researcher would like to acknowledge the many individuals who made significant contributions toward the completion of this dissertation. My perseverance was entirely contingent upon the support, encouragement and love of family, friends and mentors.

I wish to thank my former assistant principal Pam Brown who cared enough to give me the straight talk I needed at a pivotal point in my life. In 1987, you offered an opportunity for reflection that redirected the course of my entire life. Your words had a profound impact on me, and you continue to be an inspiration in my professional career.

I also wish to thank Kenny Boyers. As my boss, you have been very patient. As my friend, you have always been there with a listening ear. As a mentor, you are an endless source of wisdom. Your humble disposition and ethical decision making set you apart as a model for all who aspire to administration in education.

I am grateful for the guidance and direction offered by Dr. Karen Parker, Dean of the College of Education. You have been such a blessing. Your knowledge of the dissertation process has bolstered my confidence and diffused my anxiety. I also wish to thank Dr. Meredith Park Furrow for your constant reassurance and insight. In addition, I am indebted to Dr. Ed Brantmeier for your mentorship, encouragement and friendship every step of the way.

I wish to thank my wife, Amy, for understanding this endeavor. This has consumed valuable time as a father and husband. Thank you for recognizing that this was a sacrifice I made for our family. I hope that this accomplishment will in some way enrich our future.

Lastly, I would be remiss without thanking my mother, Mary Surmacz Johnson for your unwavering faith in me. Throughout my entire life, I knew that I could always count on you. More than anyone, you have shared in this. Thank you for always being there for me.

Dedication

This study is dedicated to my two children, Leo and Harlan. I began this journey when Leo was four years old, and Harlan was just two. From the time you were born, you have been the driving force behind my personal ambitions, and there is nothing in this world that is more gratifying to me than being the father of two such amazing children. You have been my source of inspiration and encouragement. My motivation in this venture was never one of personal gain. Rather, I envisioned this process as an opportunity that would enable me to be a better provider as a father and a husband. I am thankful that I had the foresight to begin this degree when I did. Time is fleeting, and I did not want to miss out on our time together during critical formative years. While writing this dissertation, our time has been very special. I have been blessed in being able to bring you to work with me each day. Over the last several years, you have both learned how to read, and that has made the topic of this study so much more rewarding on a personal level. How many fathers are ever fortunate enough to experience what I have? Although it has been frustrating at times, I know that I will always look back on this period in our lives with so much pride. I am grateful for you both, and I love you more than you will ever know.

Table of Contents

	Page
ABSTRACT	3
Acknowledgements.....	5
Dedication	6
List of Tables	9
List of Figures.....	10
CHAPTER 1: INTRODUCTION.....	11
Background of the Study	13
Statement of the Problem.....	15
Purpose Statement.....	16
Significance of the Study.....	17
Research Question and Hypotheses	18
Identification of Variables	19
Definitions of Key Terms	20
CHAPTER 2: REVIEW OF THE LITERATURE.....	24
Introduction.....	24
Conceptual Framework.....	25
Related Literature.....	25
Summary.....	61
CHAPTER 3: METHODOLOGY	65
Design	66
Questions and Hypotheses	67

Participants.....	68
Setting	70
Instrumentation	70
Procedures	77
Data Analysis	79
CHAPTER 4: FINDINGS	81
Purpose of the Study	81
Overview.....	82
Data Analysis and Methodology.....	82
Results	88
Summary.....	93
CHAPTER 5: DISCUSSION.....	95
Summary of Findings	96
Discussion of Findings.....	98
Limitations of the Study.....	105
Implications	106
Recommendations for Future Research	109
REFERENCES.....	112
APPENDIX A: PERMISSION LETTER FROM SCHOOL DIVISION	125
APPENDIX B: INSTITUTIONAL REVIEW BOARD APPROVAL.....	126

List of Tables

	Page
Table 1: Virginia Fourth Grade NAEP and SOL Reading Proficiency.....	14
Table 2: Decoding Objectives Assessed on the PALS Test.....	48
Table 3: Demographic Information.....	68
Table 4: Intercorrelation Matrix	87
Table 5: Collinearity Statistics for Predictor Variables	87
Table 6: Descriptive Statistics.....	88
Table 7: Correlation Coefficients of Reading Measures with SOL.....	89
Table 8: Multiple Regression Coefficients and Significance Values.....	91

List of Figures

	Page
Figure 1: Scatterplot of 2 nd grade PALS and 3 rd grade SOL	83
Figure 2: Scatterplot of QRI-5 and 3 rd grade SOL	84
Figure 3: Scatterplot of STAR test and 3 rd grade SOL	84
Figure 4: Scatterplot of residuals versus predicted values	85
Figure 5: Normal probability plot of residuals	86

CHAPTER ONE: INTRODUCTION

For elementary school teachers, the word accountability has become synonymous with performance standards on assessments such as the Virginia Standards of Learning. Achievement scores on standardized tests have been used to evaluate students, teachers and schools for more than a decade. As a reauthorization of the Elementary and Secondary Education Act of 1965, the No Child Left Behind legislation of 2001 has achieved considerable notoriety through a series of controversial penalties and restrictions for schools that fail to achieve established benchmarks in math and reading. In the decade that followed, more than 50,000 research articles were published on the topic of reading instruction (Knuth, 2011). Today, federal legislation has established provisions requiring all public schools to assess students in reading and math each year in grades three through eight (Ravitch, 2009; Siegrest & Van Patten, 2007).

In order to hold states more accountable for closing the achievement gap between various demographic subgroups, each state is required to adopt rigorous academic standards and utilize appropriate achievement assessments aligned with those objectives (Hewitt, 2011; Ravitch, 2009). In exchange, each state is afforded the flexibility of adopting testing instruments that are specifically designed to meet their needs while remaining compliant with federal directives (U.S. Department of Education, 2002). In the Commonwealth of Virginia, students are required to take a battery of tests known as the Standards of Learning (Virginia Department of Education, 2011).

Although the federal government has imposed sweeping restrictions on states' rights to monitor educational progress through increased bureaucracy, funding to enforce the legislation has not materialized. School leaders are threatened with sanctions, and teachers are made to feel shamed when test scores fall short (Siegrest & Van Patten, 2007). It has also been argued that the

rhetoric aimed at closing the achievement gap has only served to further alienate marginalized populations and students with disabilities (Alvarez & Corn, 2008; Hewitt, 2011). Perhaps the most disturbing effect of NCLB is the fact that children are subjected to “intense pressure, anxiety, and tension created by a barrage of standardized tests” (Siegest & Van Patten, 2007, p. 146).

As a result of increased accountability measures, schools throughout the United States are expected to demonstrate continued growth in the area of reading, and the federal government expects educators to incorporate research-based approaches in an effort to raise achievement (Invernizzi, Landrum, Howell & Warley, 2005; U.S. Department of Education, 2002). The emphasis on research oriented pedagogy in reading is largely the result of the findings reported in the two-year study conducted by the National Reading Panel (2000). Shanahan (2003) notes that even the U.S. Department of Education’s Reading First Program was strongly influenced by the NRP (U.S. Department of Education, 2002, April). Little more than a decade later, many states have been afforded greater latitude and autonomy through a reprieve from some NCLB legislation. Nevertheless, continued progress in the areas of reading and math remains nonnegotiable.

While many teachers and administrators are quick to condemn the federal government for reducing educational quality to sterile reports and quantifiable outcomes, it is the reality of 21st century pedagogy. Although reading is a complex human behavior, norm-referenced standardized tests provide an objective measure of each child’s ability to comprehend text. In compliance with federal directives, the first of these tests must be administered no later than the end of the third grade year (Ravitch, 2009).

In the present atmosphere of high-stakes testing, there remains considerable disagreement regarding the constituent skills of reading comprehension. This dissertation incorporates a correlational research design in an effort to evaluate the relationship between several research-

based assessments and the third grade Virginia SOL test in reading. This investigation also consolidates much of the current research literature pertaining to the prerequisites of comprehension, providing elementary teachers with an overview of reading skill progression. With an understanding of the foundational elements of comprehension and substantive assessment data, teachers can predict student success or failure with some accuracy. Furthermore, the early detection of reading difficulty enables teachers to provide remedial assistance in an effort to bolster comprehension strategies prior to the end-of-course test administration (Bailey & Drummond, 2006).

This chapter presents background information relevant to the study together with a statement of the problem and purpose of the current investigation. This chapter will also address the professional significance of the study as well as the research questions and hypotheses, concluding with a survey of the key terms used throughout this document.

Background of the Study

Learning to read is arguably the most complex skill that children are expected to develop in elementary school (Braze, Tabor, Shankweiler, & Mencl, 2007). By the time they reach the fourth grade, more than one in three students is already behind in the ability to understand written text on grade level (Katzir *et al.*, 2006). Even more unsettling, owing to the relationship between illiteracy, substance abuse, poverty and crime, states such as Arizona and California are reportedly using fourth grade reading assessment data to forecast the need for future prison cells (Riccards, 2012; Young, 2013). This information is corroborated by Shippen, Houchins, Crites, Derzis and Patterson (2010) who found that regardless of their demographic profile, prison inmates tend to score one to two standard deviations below their non-incarcerated peers on standardized reading tests.

In spite of federal directives aimed at closing the achievement gap in reading, there remains a significant divide between major ethnic subgroups (Booker, Invernizzi, & McCormick, 2007; Joshi *et al.*, 2009b; Ravitch, 2009; Siegrest & Van Patten, 2007). Evidence for this continued disparity is included in Table 1 with Virginia’s data extracted from the Nation’s Report Card for Reading (The National Center for Education Statistics, 2011). While the NAEP Test is a congressionally mandated assessment that provides a representative cross section of student reading performance in each state, the SOL test is administered to all fourth grade students throughout the Commonwealth of Virginia. Regardless of the assessment, there remains a considerable gap in the achievement of demographic subgroups.

Table 1
Virginia Fourth Grade NAEP and SOL Reading Proficiency

	Virginia’s 4 th Grade Students (NAEP)	Virginia’s 4 th Grade Students (SOL)
Total Proficiency %	39	87
White Proficiency %	49	92
Hispanic Proficiency %	21	81
Black Proficiency %	19	77

Today, teachers and administrators grapple with increased accountability in the form of standardized assessments. Beginning in third grade, reading comprehension is assessed annually through grade eight. As students transition to an instructional focus in which comprehension is paramount, third grade teachers must evaluate the predictive strength of constituent developmental reading skills. Fortunately, these skills develop predictably and sequentially. Moreover, several of these requisite skills are known to be highly correlated with emerging comprehension (Fuchs *et al.*, 2012; Hulslander, Olson, Willcutt, & Wadsworth, 2010; Verhoeven & Van Leeuwe, 2008).

In 1986, Philip Gough and William Tunmer published a short article pertaining to common reading disabilities. Their influential study presents a model of comprehension known as the Simple View of Reading. According to Gough and Tunmer (1986), comprehension is the product of decoding and linguistic comprehension. In the absence of either skill, comprehension cannot take place. Although their model has undergone revision and scrutiny for decades, it continues to be an influential theoretical construct to this day (Gough, Hoover, & Peterson, 1996; Høien-Tengesdal, 2010; Hoover & Gough, 1990; Kirby & Savage, 2008; Tunmer & Chapman, 2012).

Like the Simple View of Reading, the Lexical Quality Hypothesis put forth by Perfetti and Hart (2002) implicates the role of decoding through orthographic and phonological skill. Likewise, semantic word knowledge is an integral component of comprehension. This model asserts that comprehension is largely contingent upon the strength of lexical representations and the nexus of interrelationships between basic reading skills.

As third grade teachers attempt to forecast student performance on tests of reading comprehension, they can ill afford to rely on their own subjective opinions which have been shown to be remarkably inconsistent (Hinnant, O'Brien, & Ghazarian, 2009). In contrast, assessments which evaluate the known subcomponents of comprehension would be expected to have a strong relationship with performance scores. Therefore, this study incorporates predictor variables implicated in the Simple View of Reading and the Lexical Quality Hypothesis as being robust predictors of achievement on tests of reading comprehension.

Statement of the Problem

In compliance with federal mandates, states must assess student reading annually in grades three through eight (United States Department of Education, 2002). Although each state is afforded the benefit of utilizing its own assessment instrument (Ravitch, 2009), reading tests are

remarkably similar in that they focus on evaluating the pupil's ability to use selected reading strategies to demonstrate comprehension (Afflerbach, Pearson, & Paris, 2008; Ness, 2011; Pilonieta, 2010). For third grade students, this focus on comprehension represents a dramatic shift from fluency based instruction which places a priority on phonological awareness and decoding (Adlof, Catts & Little, 2006; Cartwright, 2006; Dooley, 2010). The problem is that third grade reading teachers frequently lack the baseline comprehension data that would enable them to predict which students are at risk of failure. In order to fill that void, assessment data which evaluates the constituent skills of comprehension is likely to be useful in predicting achievement scores on such tests.

Purpose Statement

The purpose of this correlational research investigation is to incorporate regression analysis in an effort to examine the magnitude of the relationships between beginning-of-year assessment data and reading scores on the third grade Standards of Learning assessment in Virginia. This study utilizes three predictor variables and one dependent variable. Predictor variables were selected based on their ability to assess the common sub-skills of comprehension identified in the Simple View of Reading (Gough & Tunmer, 1986) and the Lexical Quality Hypothesis (Perfetti & Hart, 2002), together with components of reading instruction recommended by the National Reading Panel (2000). As a measure of reading comprehension, the Virginia Standards of Learning test has been included as the outcome, or dependent variable. In selecting variables that have a significant relationship with achievement scores in reading, elementary teachers will be able to accurately identify students who would benefit from remedial assistance before end-of-course assessments are administered. Likewise, the present study will also help determine the unique contribution and predictive value for each of the comprehension sub-skills.

Significance of the Study

Within an educational climate defined by norm-referenced tests and increased rigor, there remains a curious educational paradox in the area of elementary reading. Research indicates that many popular textbook publishers fail to address the critical areas as outlined by the NRP, and many of our nation's teacher educators are themselves misinformed about core literacy concepts (Joshi *et al.*, 2009a). As a result, many classroom teachers begin their careers lacking the necessary foundational knowledge regarding reading instruction (Joshi *et al.*, 2009b), and a full third of our nation's students begin their fourth grade year reading below grade level (Katzir *et al.*, 2006).

The present study includes a thorough analysis of the research literature in an effort to identify predictor variables of reading comprehension in the middle elementary grades. While this investigation incorporates commonly used assessment instruments in measuring these predictors, none of the extant research literature is known to compare these three specific tests against one another. Therefore, this study provides valuable insight regarding the predictive quality of these instruments, allowing instructional leaders to make precise data-driven decisions relevant to reading remediation and enrichment. Moreover, an efficient and effective diagnostic approach will likely save valuable teaching time and division resources that could be better spent elsewhere.

The outcome of this research expands upon the professional knowledge base pertaining to elementary reading by combining elements of theoretical models put forth by Gough and Tunmer (1986) and Perfetti and Hart (2002). By integrating measures of decoding and vocabulary, this investigation represents an extension of the Simple View of Reading and the Lexical Quality Hypothesis, thus offering further clarity to more recent research conducted by Protopapas, Mouzaki, Sideridis, Kotsolakou, and Simos (2013) as well as Protopapas, Simos, Sideridis, and

Mouzaki (2012).

As school leaders search for strategies to meet state accreditation and federal accountability measures, a research-based approach to reading instruction is essential. This study provides an empirical context regarding the predictable sequence of developmental skills children acquire as they work toward reading for meaning. Therefore, the implications of this investigation are applicable to elementary reading teachers throughout the Commonwealth of Virginia and other states that utilize assessments similar to the Standards of Learning. In addition, the results will furnish educators with the critical tools necessary to provide differentiated instruction to all students. In analyzing information regarding deficits in prerequisite literacy skills such as orthography and phonological awareness, teachers can provide early targeted assistance. Therefore, with appropriate intervention, our nation's reading teachers can reduce the number of students reading below grade level prior to the end of their elementary school experience.

Research Questions and Hypotheses

The present study was designed to evaluate the following research questions and corresponding null hypotheses:

Research Question 1: What is the relationship between diagnostic reading measures obtained at the beginning of the year with achievement scores on the third grade Standards of Learning assessment in reading?

Null Hypothesis 1: There will be no significant correlation between scores on the Phonological Awareness Literacy Screening and scores on the Virginia Standards of Learning reading assessment.

Null Hypothesis 2: There will be no significant correlation between scores on the STAR Reading Test and scores on the Virginia Standards of Learning reading assessment.

Null Hypothesis 3: There will be no significant correlation between scores on the Qualitative Reading Inventory and scores on the Virginia Standards of Learning reading assessment.

Research Question 2: What is the relationship between a linear combination of three predictor variables (the joint model comprised of PALS, STAR and QRI) and achievement scores on the third grade Standards of Learning assessment in reading?

Null Hypothesis 4: There is no significant relationship between a linear combination of the three predictor variables and achievement scores on the third grade Virginia SOL test in reading.

Null Hypothesis 5: There will be no difference in the amount of SOL reading variance captured by the PALS, STAR and QRI reading measures.

Null Hypothesis 6: No linear combination of variables predicts achievement scores on the third grade Virginia SOL reading test better than the linear combination of all three predictor variables utilized simultaneously.

Identification of Variables

This study incorporates a non-experimental multiple regression design in testing each of the two research questions and corresponding hypotheses. Further, the study utilizes three independent variables as predictors of the criterion, or dependent variable. These variables are operationally defined as follows:

Phonological Awareness Literacy Screening: Otherwise known by the acronym PALS, this instrument serves as one of the three independent variables used to predict the dependent variable. The PALS assessment is utilized as a measure of student decoding. This study includes student scores achieved on the assessment administered at the end of the second

grade year. PALS is a criterion referenced test which yields individual task scores as well as a total summed score (Invernizzi, Meier, & Juel, 2013). The test is administered one-on-one with students during a 23-45 minute testing session (National Center on Response to Intervention, 2013).

STAR Reading Test: This instrument serves as one of the independent variables in the present study. The STAR Reading Test has been included as a measure of vocabulary. STAR Reading generates scores reported as grade level equivalencies. Although the computer generated test is usually administered to a whole group of students simultaneously, each student takes the test independently. The entire testing session lasts approximately ten minutes (Renaissance Learning, 2011).

QRI-5: Like the PALS test, the Qualitative Reading Inventory – 5 is administered to students one-on-one. This test serves as an independent variable based on its ability to evaluate student word recognition (or linguistic comprehension), one of the subcomponents of comprehension as outlined in the Simple View of Reading (Leslie & Caldwell, 2011). The test takes approximately 45 minutes to administer, and scores are reported as independent reading levels.

Third Grade Reading SOL: As a measure of reading comprehension ability, the third grade Standards of Learning assessment has been included as the criterion, or dependent variable. A criterion-referenced assessment, this test incorporates fictional and nonfictional material in assessing student comprehension skills (Virginia Department of Education, 2011). This is a multiple-choice, untimed computer administered assessment.

Definitions of Key Terms

The following terms will be used throughout this investigation:

Alphabetic Principle: “The concept that letters stand for speech sounds” (Ganske, 2000, p. 8). Recognition of the individual phonemes in spoken language is also a necessary component of the alphabetic principle.

Assessment: "A broad repertoire of behaviors involved in noticing, documenting, recording, and interpreting children's behaviors and performance" (Casbergue, 2011, p.16). Formative assessments tend to be less formal or anecdotal observations regarding a child's learning while summative assessment involves more formal means of judging concept or skill mastery. Summative assessments include tests and quizzes as well as norm referenced standardized examinations.

Comprehension: Constructing meaning of what is read as the reader interacts with text. The student incorporates reading strategies combined with unique background knowledge to understand print materials within a socio-cultural context (Dooley, 2010). Comprehension also involves “Recalling information from text, extracting themes, engaging in higher order thinking skills, constructing a mental picture of text, and understanding text structure” (Ness, 2011, p. 98).

Decoding: Utilizing letter-sound recognition in an effort to identify individual words in print. It is the process by which graphemes (written letters) are converted to phonemes, or speech sounds. Decoding is commonly viewed as a prerequisite skill to building fluency and reading comprehension (Cartwright, 2006). This skill is occasionally referred to as phonological recoding (Ehri, 2005).

Fluency: A skill that has a strong relationship with sight word reading ability, it is “the ability to read single words quickly and accurately in and out of context” (Barth, Catts, & Anthony, 2009, p. 568). According to Kim, Wagner and Foster (2011), fluent reading helps to free working memory for higher order thinking and the construction of meaning.

Orthography: “The writing system of a language, specifically, the correct sequence of letters, characters, or symbols” (Bear, Invernizzi, Templeton, & Johnston, 2008, p. 384).

According to Wright and Ehri (2007), English is characterized by orthographic depth due to inconsistent mapping between graphemes and phonemes together with complex syllabic structure. These orthographic qualities are thought to hinder the rate at which beginners learn to read in English.

Phoneme: The smallest unit of sound that has the potential to affect the meaning of words. Approximately 41 phonemes are utilized in the English language, combining to form syllables and words. (National Reading Panel, 2000).

Phonics: “The system by which symbols represent sounds in an alphabetic writing system” (Adams, Foorman, Lundberg, & Beeler, 1998, p. 3).

Phonemic Awareness: A sub-skill of phonological awareness. “It refers to the ability to identify and reflect on the smallest units of sound: individual phonemes” (Bear *et al.*, 2008, p. 96). An example of phonemic awareness is the segmentation of spoken words into the individual constituent phonemes, or sounds.

Phonological Awareness: Recognizing that spoken language is comprised of a sequence of phonemes, or sounds, which form words when they are combined (Johnson 2004). According to Bear *et al.*, (2008), phonological awareness also involves understanding alliterative sounds, recognizing rhyming words, and identifying word syllables. Nithart *et al.* (2011) describe the skill as the “ability to perceive, segment and explicitly manipulate the sounds of spoken words” (p. 346).

Reading readiness: “The idea that until children have reached the optimum age and capacity for learning to read, instruction in literacy skills will be useless - or worse - damaging,

to children's development" (Casbergue, 2011, p. 14).

Sight word: "Securing spellings of words to their pronunciations and meanings in memory so that when the words are seen, they are read automatically from memory rather than by applying a decoding strategy" (Wright & Ehri, 2007, p. 116). Many students are able to identify common words with some automaticity and without significant mental processing by the age of seven or eight (Andrews & Bond, 2009).

Summary

This correlational dissertation incorporated a sample of third grade students from a mid-sized school division in the Commonwealth of Virginia. The present study utilized multiple regression analysis to compare the practical utility of three diagnostic measures in predicting student achievement on the third grade Standards of Learning reading assessment. This chapter has presented the background of the study together the statement of the problem, purpose statement and the significance of the study. Research questions and hypotheses have also been presented with an identification of variables and the definitions of key terms. Chapter two presents a review of the research literature pertaining to the development and assessment of reading comprehension.

CHAPTER TWO: REVIEW OF THE LITERATURE

Introduction

Third grade is a critical year for elementary reading teachers and school administrators alike as it marks the first time students are administered standardized reading assessments in compliance with federal mandates. Curiously, research on the subject of elementary literacy indicates that one third of our nation's students are already reading below grade level by the culmination of their third grade year (Joshi *et al.*, 2009b; Katzir *et al.*, 2006). Furthermore, lack of appropriate intervention and remediation contributes to low self-esteem, poor motivation, disciplinary concerns and general academic underachievement (Sloat, Beswick, & Willms, 2007). Fortunately, it has also been shown that the early detection of skill deficits helps prevent later problems in reading development (Bailey & Drummond, 2006).

In the state of Virginia, students take the grade three Standards of Learning (SOL) reading test as a measure of basic reading competency. Underlying proficient reading are constituent skills thought to be foundational to comprehension. As those abilities preclude reading for meaning, instruments which evaluate the known components of comprehension are expected to be predictive of success on standardized tests such as the SOL.

This chapter begins with an examination of the conceptual models that have advanced our understanding of emergent comprehension. A cursory search of the professional literature reveals that the Simple View of Reading (Gough & Tunmer, 1986) and the Lexical Quality Hypothesis (Perfetti & Hart, 2002) are complementary theoretical orientations with similar foundational components. This information is further supplemented with a survey of the related empirical research pertaining to developmental reading skills. This includes an examination of the continuum bridging the decoding process and the evolution of fluency to reading for

understanding. This survey continues with a focused presentation of comprehension predictors including phonological awareness and vocabulary. The literature review concludes with a presentation of various instruments thought to be useful in predicting achievement on standardized tests. These include the Phonological Awareness Literacy Screening (PALS), the STAR Reading Test and the Qualitative Reading Inventory (QRI). Following a review of previous studies, critical elements are summarized and gaps in the current research are presented, establishing the significance of this research investigation.

Conceptual Framework

As educators grapple with the complexities of teaching reading, it is clear that our ability to identify struggling readers has changed tremendously over the last several decades (McKenna & Walpole, 2008). According to the guidelines of Reading First, a federally endorsed reading program, there are five essential components of basic reading instruction. These include a knowledge of phonetics, word decoding, vocabulary development, fluency and comprehension (U.S. Department of Education, 2002, April). As students transition into the upper grades in elementary school, comprehension becomes the primary goal, and a host of conceptual models have attempted to explain this complex skill. Two of the most influential models include the Simple View of Reading and the Lexical Quality Hypothesis.

The Simple View of Reading. In 1986, Philip Gough and William Tunmer published a brief article pertaining to the impact of decoding on students with reading disabilities. Published at a time when whole language and phonics advocates were in opposing camps, this unifying theoretical construct offered affirmation to both sides. This simple conceptualization of balanced literacy would become immensely popular over the next several decades (Kirby & Savage, 2008). In Britain, the theory has even been adopted by the Department for Education and

Employment as an integral component of the National Literacy Strategy Framework for Teaching (Stuart, Stainthorp, & Snowling, 2008).

At the time their work was published, Gough and Tunmer (1986) expressed concern that there was considerable disagreement regarding the role of the prerequisite skills that impact comprehension. Offering clarity on the topic, the authors propose that decoding is an essential skill that is foundational to comprehension. A child's ability to decode is contingent upon the rules of letter-sound correspondence. More specifically, beginning readers acquire an alphabetic orthography, or code, which allows them to recognize familiar words in print. With experience, the reader decodes words more rapidly and with greater precision. In building a better understanding of letter-sound correspondence rules, students also become adept at pronouncing pseudowords which follow regular pronunciation patterns. Although decoding is an essential prerequisite of comprehension, it is not sufficient by itself, and many skillful decoders continue to struggle with comprehension.

In addition to decoding, the reader must also recognize and understand the words they encounter as part of their personal vocabulary. Although this linguistic comprehension is a central element of the Simple View of Reading, Gough and Tunmer (1986) are surprisingly vague regarding the role of this skill and its impact on reading comprehension. Rather, the authors include the simple formula $R = D \times C$ as a means of representing the contribution of decoding (D) and linguistic comprehension (C) as sub-skills of reading comprehension (R). This model implies that (D) and (C) make an equal contribution to (R), and a value of zero for either component would yield an overall comprehension score of zero. In other words, without the ability to decode, linguistic comprehension is useless in its contribution to reading comprehension and vice versa.

Although Gough and Tunmer make reference to some of the research studies supporting their construct, they neglect to include any significant data to validate their model. However, they note that the Simple View of Reading “offers considerable meat for debate, for it has a number of testable implications” (p. 7). Recognizing that their construct rests mainly on hypotheticals, they encourage other researchers to investigate the matter further.

Expanding upon the original conception of the Simple View of Reading, Hoover and Gough (1990) published a revised theory incorporating a more thorough analysis of the individual subcomponents. Suggestions are also included for assessing both decoding and linguistic comprehension skills. While still lacking in original research, the authors refer to numerous studies thought to support the new model.

The revised theory asserts that decoding is of primary importance in the early elementary grades as students build a foundation in reading. Later, however, linguistic comprehension becomes the more dominant predictor of comprehension. This conflicts with the original theory which implied that decoding and linguistic comprehension shared an equal contribution. Although Hoover and Gough (1990) describe a shift in importance as linguistic comprehension becomes the stronger component, the authors continue to reaffirm that “the simple view holds that these two parts are of equal importance” (p. 128).

While Gough and Tunmer encouraged other researchers to test the implications of the Simple View of Reading, Hoover and Gough (1990) cite several studies that provide support for the theory. In one investigation, vocabulary knowledge was used in measuring the impact of linguistic comprehension. Likewise, the authors concede that tests of phonological awareness have been used for the purpose of measuring student decoding skills. Regardless of the instrumentation used in measuring these subcomponents, the revised model holds that

comprehension is the product of student skills in decoding and linguistic comprehension.

Gough, Hoover and Peterson (1996) published another follow-up study pertaining to the Simple View of Reading. Their work helped to illuminate many of the ambiguous details inherent in the original model. In reaffirming the basic model, they speculated that decoding and linguistic comprehension were likely to be highly correlated with one another. As evidence, they note that students who struggle in one area tend to struggle with the other skill as well. In this study, the terms linguistic comprehension, word comprehension and listening comprehension are used synonymously. Regardless of the preferred terminology, one of the major implications was that decoding and linguistic comprehension should be taught differently and assessed separately. Gough *et al.*, (1996) reaffirm the multiplicative quality of reading as the result of decoding coupled with linguistic comprehension. Comprehension does not take place in the absence of either skill. Expanding upon the work of Hoover and Gough (1990), their work further reinforced the idea that student reliance on decoding gradually declines through the years, while vocabulary becomes a critical component of linguistic comprehension.

More than two decades after Gough and Tunmer published their Simple View of Reading, Tunmer and Chapman (2012) published the most recent incarnation of this theory. They propose that decoding should be assessed with an instrument measuring automated sight word recognition. However, they also offer a cautionary word in discriminating between decoding and the more advanced skill, fluency. Although Gough and Tunmer originally referred to (C) as linguistic comprehension, Tunmer and Chapman (2012) prefer the term oral language comprehension, explaining that it is typically measured with instruments used to gauge vocabulary knowledge. In evaluating the continued utility of the original conception of the Simple View of Reading, however, they conclude that “the fundamental two-component

structure of the model should remain intact” (p. 462).

Influence of the Simple View of Reading. Although the Simple View of Reading has generally been used to explain the emergence of comprehension in elementary students, some studies found it to be useful in predicting reading with middle and high school students (Tilstra, McMaster, Van den Broek, Kendeou, & Rapp, 2009), and in accounting for the two primary reading deficits present in unskilled adult readers (Braze, Tabor, Shankweiler, & Mencl, 2007; Savage & Wolforth, 2007). When utilized with students in fourth, seventh and ninth grades, the Simple View of Reading continues to explain a significant portion of the variance in comprehension. In evaluating the strength of the model, Joshi and Aaron (2000) noted that the Simple View accounts for approximately 48% of the variance in comprehension. Nevertheless, as students develop into competent readers, their reliance on decoding gradually diminishes.

Although the Simple View of Reading continues to hold wide appeal, there are those who question its validity. For instance, some suggest that it fails to adequately describe the complex processes involved in reading for meaning (Høien-Tengesdal, 2010). Others have criticized the model for its failure to isolate the unique contribution of each subcomponent (Ouellette & Beers, 2010). Additional studies recommend inclusion of a third component beyond decoding and language comprehension (Adlof *et al.*, 2006; Protopapas, Mouzaki, Sideridis, Kotsolakou, & Simos, 2013). Finally, an additive model of decoding and word comprehension has been offered as a more appropriate explanation of reading comprehension (Kirby & Savage, 2008; Savage & Wolforth, 2007).

In working with Norwegian and Swedish elementary students, Høien-Tengesdal (2010) evaluated the model’s ability to account for comprehension in speakers of languages other than English. In a correlational study of nearly 500 students in Norway and more than 200 students in

Sweden, it was determined that decoding and word comprehension each have an influential, yet unequal contribution in understanding what is read. While other factors such as rapid digit naming and phonemic awareness were also found to be related to comprehension, it is likely that the impact is indirect. Although they propose an additive model as being more precise, they still conclude that decoding and language comprehension are the two most important sub-skills among successful readers.

Studies conducted by Georgiou, Das and Hayward (2009) and Joshi and Aaron (2000) also attempted to compare the traditional multiplicative model of the Simple View against an additive model. In both cases, results failed to demonstrate that the additive model accounts for comprehension beyond the traditional model. Although Georgiou *et al.*, (2009) hypothesized that other cognitive processes had an impact beyond decoding and language comprehension, those skills neglected to account for unexplained variance in the original theory.

Evaluating the differing impact of the two subcomponents of comprehension, Ouellette and Beers (2009) studied students in grade one and in grade six. Incorporating multiple measures of both decoding and language comprehension, they concluded that the sub-skills proposed by Gough and Tunmer play an unequal role in comprehension. Over time, the impact of language comprehension becomes the dominant skill as vocabulary explains more of the variance in assessments of reading comprehension. Conversely, the contribution of decoding is gradually reduced, reaffirming the conclusions of Gough *et al.*, (1996) and Hoover and Gough (1990).

Although Adlof *et al.*, (2006) concede that the original model of the Simple View accounts for as much as 45-85% of the variance in comprehension, they investigated the possibility that fluency could account for some of the remaining variance. In their study of more

than 600 students in grades two, four and eight, they evaluated student fluency after controlling for the effects of decoding and language comprehension. Results show that fluency accounted for very little independent variance in reading comprehension. Moreover, fluency had such a strong relationship with decoding that their independent contributions were virtually indistinguishable.

A recent model of the Simple View proposes that comprehension is the result of a print-specific factor and a language comprehension element that is print-independent (Protopapas *et al.*, 2013; Protopapas *et al.*, 2012). While decoding and sight word reading each constitute print-dependent sub-skills, vocabulary measures are offered as the metalinguistic print-independent component. In a study of 436 Greek elementary school students, it was found that measures of vocabulary account for additional variance in reading comprehension beyond decoding and language comprehension. Given these results, they proposed that a lexical skill component should be included in the Simple View of Reading in order to more accurately predict comprehension (Protopapas *et al.*, 2013).

Although the bulk of the research literature pertaining to the Simple View of Reading has been conducted with samples of students who speak English as their first language, numerous studies have established the strength of the model in predicting comprehension in other languages as well (Høien-Tengesdal, 2010; Protopapas *et al.*, 2013; Protopapas *et al.*, 2012; Verhoeven & Van Leeuwe, 2008). Additional research has also demonstrated the construct's effectiveness in explaining emerging comprehension in English Language Learners (Gottardo & Mueller, 2009; Leider, Proctor, Silverman, & Harring, 2013; Lervag & Aukrust, 2010; Proctor, Carlo, August, & Snow, 2005). In fact, the Simple View has even been used to account for comprehension variance among Native American First Nations children in Canada (Georgiou,

Das, & Hayward, 2009).

In evaluating the utility of the Simple View with a large sample of sixth grade students in Norway, Høien-Tengesdal (2010) found that, while other sub-skills such as orthography explain slight amounts of unexplained variance, the original model sufficiently addresses the two main skills underlying comprehension. Reaffirming earlier findings reported by Hoover and Gough (1990) and Gough *et al.*, (1996), Protopapas *et al.*, (2013) conducted a study with Greek students, reporting that the importance of decoding is gradually diminished as the role of vocabulary intensifies beginning in the middle elementary grades. Further, the differential impact of word knowledge in older children was also found to be significant as Dutch children progress from first grade through sixth grade (Verhoeven and Van Leeuwe, 2008).

In their critical analysis of the Simple View, Kirby and Savage (2008) express concern about the model's ability to address reading development in second language learners. In navigating the complexities of predicting comprehension in American English Language Learners, Gottardo and Mueller (2009) evaluated the appropriateness of the Simple View with a sample of first and second grade students from Spanish-speaking homes. Emphasizing the critical impact of English vocabulary on comprehension, "the results support the validity of the simple view of reading as a model for the development of reading comprehension in young ELs" (p. 330). Recognizing that ELL students frequently perform below average on standardized assessments of reading comprehension in spite of adequate decoding skills, Leider *et al.*, (2013) emphasized the importance of moving beyond phonological awareness and simple word reading tasks. Incorporating cloze exercises as a measure of linguistic comprehension within the Simple View of Reading, English vocabulary measures proved to be the more significant skill in predicting comprehension among bilingual Latino students in the elementary setting.

Nevertheless, comprehension in Spanish-speaking ELL students continues to be the product of both decoding and vocabulary just as it is with their monolingual counterparts (Proctor *et al.*, 2005). Therefore, although second language learners frequently lag behind their first language peers in their understanding of text, the Simple View of Reading continues to be an appropriate model for gauging comprehension in ELL students (Lervag & Aukrust, 2010).

Although Kirby and Savage (2008) describe the Simple View of Reading as an incomplete theory, they add that it was never intended to be all encompassing. Rather, its greatest utility is its simplicity and versatility. While imperfect, it continues to provide educators with a general model of the interaction between the two most critical factors of comprehension. In their closing comments, they call for the appropriate assessment of reading sub-skills as a means of providing differentiated instruction commensurate with each child's unique needs.

The Lexical Quality Hypothesis. Although the Simple View of Reading asserts that comprehension is comprised of two distinctive and independent components, Perfetti and Hart (2002) emphasize the interrelationships of constituent reading skills. Their Lexical Quality Hypothesis views comprehension as being largely contingent upon word reading skill. This model stipulates that verbal efficiency results in the automatic recall of words, enabling the reader to focus attentional resources to understanding. This element of verbal efficiency is largely the result of the quality lexical representations which facilitate meaningful word retrieval from memory.

The retrieval of word meanings is of central importance in the Lexical Quality Hypothesis. According to Perfetti and Hart (2002), readers encounter words on multiple occasions allowing them to store orthographic, phonological and semantic information in memory. Therefore, repeated exposure to words facilitates memory of important elements

pertaining to spelling, pronunciation and meaning. Retrieval of word knowledge is thus dependent upon lexical quality. As opposed to viewing the constituent literacy skills as being separate and unrelated, the Lexical Quality Hypothesis assumes a fair amount of overlap or redundancy amongst foundational comprehension skills.

Within the context of this model, observed differences in student comprehension are thought to result from variance in quality lexical representations (Perfetti, 2007). Like the Simple View of Reading, this model places considerable importance on decoding and basic word identification. Perfetti (2007) notes that “children with inefficient word-level processes would have problems with comprehension” (p. 358). Conversely, students with efficient word representations are likely to comprehend well. Nevertheless, all readers will possess a mix of both low and high quality representations (Perfetti & Hart, 2002). Regardless of a student’s skill level, it is said that lexical quality can be refined with practice. As students continue to read, they bind together connections of the three constituent word level skills, reinforcing a word’s representations (Andrews & Bond, 2009; Kucan, 2012; Perfetti, 2007).

Just as decoding is a central component of the Simple View of Reading, it is also a critical element of the Lexical Quality Hypothesis. In fact, the latter stipulates that word identification is the most important contributor to comprehension (Perfetti, 2007). Students clearly attend to a word’s orthography and phonology during the decoding process. However, vocabulary knowledge also plays a key role in the Lexical Quality Hypothesis (Kucan, 2012). Understanding written text requires students to be familiar with a word’s semantic representation, or meaning. Again, there are obvious parallels between the semantic element of the Lexical Quality Hypothesis and the linguistic comprehension component in the Simple View of Reading (Protopapas *et al.*, 2012).

In comparing the Simple View of Reading and the Lexical Quality Hypothesis, there are several conspicuous similarities (Braze *et al.*, 2007). Both models effectively reduce the complex behaviors involved in reading to basic skills. There is also agreement that comprehension requires proficiency in decoding together with understanding of word knowledge. However, the primary difference between these two influential models is outlined by Protopapas *et al.*, (2012) who note that the Lexical Quality Hypothesis provides an “emphasis on ties among, rather than distinctions between, constructs” (p. 235).

Related Literature

Throughout the United States, third grade generally marks a transitional period as students move from fluency based instruction to a curriculum with an emphasis on comprehension. Reading for meaning requires the student to develop a complex set of skills that is usually acquired in a predictable sequence. For elementary teachers, third grade marks a pivotal time as federal guidelines require states to assess student achievement in reading each year in grades three through eight. While reading teachers begin to focus instruction on comprehension, predicting performance on high-stakes assessments can be frustrating due to a lack of appropriate baseline data. Therefore, as children negotiate this crucial instructional paradigm shift, teachers must be knowledgeable of the prerequisite skills and assessments that help define and measure the emergence of comprehension.

Prior to the authorization of No Child Left Behind in 2001, the subject of elementary reading instruction captured the attention of the U.S. Congress. Shortly afterwards, the U.S. Department of Education, in conjunction with the National Institute of Child Health and Human Development, formed the National Reading Panel (NRP). After meeting for more than two years, the 14 members of the NRP published their research findings in a meta-analysis of more

than 500 pages (National Reading Panel, 2000). Their report synthesized the work of hundreds of quality investigations on the topic of reading instruction, identifying a core set of reading sub-skills that contribute to comprehension. In summary, the report implicates phonemic awareness and phonological awareness as contributors to decoding and fluency. In addition, considerable attention is devoted to the contribution of vocabulary and its impact on comprehension (National Reading Panel, 2000). Over the last two decades, the report of the NRP has been instrumental in advancing the direction of research-based instruction in reading comprehension (Shanahan, 2003).

Early reading skills. Before students can read for meaning, they must first master a progression of basic skills. In demonstrating a basic reading readiness, children internalize simple orthography and alphabetic knowledge (Bear *et al.*, 2008; Casbergue, 2011). With fundamental phonological awareness, children incorporate letter-sound recognition to identify the individual words encountered in print (Cartwright, 2006; Nithart *et al.*, 2011; Savage, Carless & Ferraro, 2007). When students are able to identify words with immediacy, they begin the journey to becoming a fluent reader. However, moving beyond decoding and fluency based instruction requires the student to process multiple sources of information simultaneously. They must combine phonological and semantic information together with other contextual cues and unique background experiences in order to make sense of printed materials. Finally, comprehension is also contingent upon one's socio-cultural experience (Dooley, 2010)

The decoding process. Decoding is the strategy that readers employ when utilizing letter-sound recognition to identify individual words. Specifically, it is the process by which graphemes (written letters) are converted to phonemes, or speech sounds (Cartwright, 2006). Findings reported by the National Reading Panel (2000) indicate that decoding is enhanced by

skill development in phonological awareness, and the constituent skill, phonemic awareness. As students encounter printed words for the first time, they must transform graphemes into appropriate phonemes. Then, students employ the phonological skill of blending those phonemes to form recognizable words. Tests such as the Phonological Awareness Literacy Screening specifically target the student's ability to blend phonemes and decode text (Invernizzi *et al.*, 2013).

As a sub-skill of phonological awareness, systematic instruction in phonics is thought to develop the ability to become a skillful decoder by expediting acquisition of the alphabetic principle (Stuebing, Barth, Cirino, Francis, & Fletcher, 2008). Decoding begins with individual words, progressing to the phrase and sentence levels. During the first three years of elementary school, decoding is highly predictive of comprehension, and this skill greatly facilitates reading speed over time (Joshi and Aaron, 2000). One popular theory suggests that the decoding process can be broken down into the following four stages: the pre-alphabetic, the partial alphabetic, the full alphabetic and the consolidated alphabetic (Ehri, 2005).

Children frequently begin to decode by identifying isolated words without attending to letter sound correspondence. Gradually, children come to recognize the individual phonemes at the beginning and end of spoken words. In doing so, they can apply newly acquired knowledge to selected words in a way that is no longer arbitrary. Next, students discover that new words can be systematically decoded, converting each letter into the appropriate sound. However, students often continue to struggle with words that do not follow a regular pattern of spelling. At that point, students often acquire a rich foundation of sight words that are recognized immediately (Ehri, 2005; Stuart *et al.*, 2008). According to Kirby *et al.*, (2003), the majority of students move into the final stage of decoding by third grade, even retaining irregularly spelled

words.

It has been said that "beginning readers take a necessary step toward fluent reading when they 'unglue' from print sometime between the second and fourth grades and negotiate a crucial transition from decoding-focused reading to more fluent, meaning-focused reading" (Cartwright, 2006, p. 628). The professional literature abounds with articles pertaining to the importance of decoding, and it is generally viewed as a prerequisite skill to fluency and comprehension (Keenan, Betjemann, & Olson, 2008). Some studies have attempted to quantify the connection between the skills, and estimates suggest that decoding contributes as much as 80% of the variance in comprehension (Høien-Tengesdal, 2010).

The move from decoding (or phonological recoding) to fluent reading is greatly facilitated through the process of building a repertoire of sight words. In this way, the reader moves away from viewing individual words as the sum of its parts to recognizing the word as a single unit. According to Ehri (2005), any word can become a sight word when it is encountered repeatedly. Contrary to popular belief, sight words are not merely high-frequency words. When readers become skillful decoders, their phonological awareness helps to establish a link between spelling and the word's pronunciation and meaning in long-term memory. Curiously, irregularly spelled words are often learned as sight words as the grapheme-phoneme correspondence is of little help in recognizing the word's pronunciation.

While there is little doubt that decoding is a critical skill that contributes to comprehension, some have proposed that it develops alongside the ability to read for meaning (Dooley, 2010). The connection between the two variables is difficult to analyze, and it remains unclear if the relationship between decoding and comprehension is causal. Comprehension happens only when the student can identify words and understand their meaning (Cartwright,

2006). Student reliance on decoding is gradually minimized as students are able to recognize words with immediacy. Thus, decoding begins a steady decline sometime between the second and fourth grade years (Joshi & Aaron, 2000; Tilstra *et al.*, 2009). Verhoeven and Van Leeuwe (2008) explain: “as children develop better word-decoding skills, their reading comprehension becomes more constrained by their vocabulary and listening comprehension skills” (p. 419).

Building fluency. Once students are able to decode with ease, they make a transition toward becoming fluent readers, an essential step toward reading for meaning (Cartwright, 2006; Kim, Petscher, Schatschneider, & Foorman, 2010; Kim, Wagner, & Foster, 2011; National Reading Panel, 2000; Rasinski, Homan & Biggs, 2009; Schwanenflugel *et al.*, 2006). In describing the importance of fluency, the U.S. Department of Education (2002, April) concludes that it “provides a bridge between word recognition and comprehension” (p. 3). However, becoming a fluent reader is often a difficult adjustment, and research indicates that there are more than eight million dysfluent readers in grades four through twelve (Joshi *et al.*, 2009b).

Although elementary teachers have long relied on oral reading fluency in gauging reading progress, the skill has often been overlooked by those who were more concerned with advancing students toward silent reading (Clark, Morrison, & Wilcox, 2009; Rasinski, 2009). Conversely, fluency has also been overshadowed by emphasis placed on fundamental decoding skills (Katzir *et al.*, 2006). Fluent readers do not stumble over individual words as they read aloud. Rather, they are able to process them more quickly because they are not constrained by the need to decode one letter and one sound at a time (Adlof *et al.*, 2006). One quality of fluent reading that emerges in second and third grade students is the presence of inflection and reading with expression (Schwanenflugel *et al.*, 2006). Student practice results in reading with increased automaticity and the freeing of cognitive resources, thus permitting working memory to focus on

building understanding (Kim *et al.*, 2011). Likewise, findings of the National Reading Panel indicate that explicit fluency based tutoring and instruction significantly impact ability to read for meaning (Stuebing *et al.*, 2008).

Confirming what many practitioners believe, Kim *et al.*, (2011) report that the correlation between oral reading fluency and reading comprehension is especially pronounced for students in the first grade ($r = .73$ to $.76$). Similar results were found for students in third grade with correlation coefficients of $.67$ and $.70$. Gradually, as students become fluent readers, phonological awareness and decoding make a smaller contribution to comprehension allowing advanced skills to play a more significant role (Kim *et al.*, 2010; Kim *et al.*, 2011).

In a study of more than 500 students in grade eight, Barth *et al.*, (2008) were able to isolate four component skills that contribute to reading fluency. Their work demonstrates the critical importance of accessing phonological information in decoding as a means to develop sight word vocabulary. Similar results were found by Katzir *et al.* (2006). Although word naming speed, oral language comprehension and working memory also contribute to reading fluency, they had little impact. Rather, efficient sight word reading was found to account for most of the variance in fluent reading. The implications of the study further suggest that continued practice in reading helps to strengthen phonological representations resulting in more efficient recall from working memory.

Recognizing fluency as a critical attribute of comprehension, considerable attention has been given to techniques aimed at improving this skill. In planning for effective remediation for struggling readers, Rasinski *et al.*, (2009) offer several strategies to use in building fluency skills. For instance, practice through repeated readings helps to facilitate increased reading speed and word recall. However, they recommend that students practice in the presence of a teacher or

parent who can provide adequate support and direction. Stronger readers such as parents and teachers can also model fluent reading with expression and proper intonation for emergent readers. Other effective interventions include assisted reading and readers' theater performances. While elementary students consider timed passages and rereading to be tedious and uninspiring, many students appreciate the change of pace provided by readers' theater performances. With an emphasis on inflection, expression, tone and pacing, students experience substantial gains through repeated readings and hearing fluent reading as it is modeled by peers and classroom teachers (Clark *et al.*, 2009).

Reading for understanding. Reading comprehension is a complex mental activity. While children progress through a predictable sequence of skills, each child grows in their own unique way. Surprisingly, there is widespread confusion regarding how to best define comprehension. Cain, Oakhill, and Bryant (2004) explain that much of the professional literature on the subject is focused on foundational skills of comprehension. For instance, May (2011) provides a list of teacher actions including modeling of strategies, discussion techniques and monitoring student engagement. Although the absence of an objective definition is commonplace, Dooley (2010) describes comprehension as a process whereby "meanings are made as readers (with background knowledge and strategic approaches to texts) enter into situations (or activities with particular implicit or explicit purposes) and transact with a text to create meanings" (pp. 120-121).

Most experts agree that comprehension is the ultimate goal of reading. However, there remains considerable disagreement regarding the number and sequence of prerequisite skills that children are expected to master beforehand. The idea that there is a natural evolution and progression of skills is appealing, and it remains very popular to this day. This approach

suggests that students who build skills in decoding will become fluent readers who can then begin to read for meaning (Howell, Partridge, Landrum, & Invernizzi, 2004; Ryder, Tunmer, & Greaney, 2007). This point is also summarized by Kesler (2010) who notes that "as less attention is required for decoding, more attention becomes available for comprehension" (p. 274).

While this developmental perspective continues to have its supporters, others take exception with a sequential evolution of reading skills. Some suggest that comprehension begins to emerge simultaneously with decoding and fluency. As opposed to a traditional view which sees children attending to text one word at a time, other models insist that children synthesize pictures, text layout, and prior knowledge much earlier than was previously thought. For instance, Dooley (2010) writes that young children interact with text in unique ways that are different from older students. This view of developmental comprehension represents a departure from the conventional approach to teaching reading, and it takes into account the critical impact of each child's background knowledge and experience.

Predictors of achievement in reading comprehension. Given the increased attention to high stakes testing, a large volume of professional research has been devoted to identifying the early predictor variables associated with success on reading tests. The same predictors can also help teachers identify struggling students, ensuring an accurate diagnosis of skill deficits. School administrators can therefore offer targeted assistance and remedial intervention to students who need help the most (Fuchs *et al.*, 2012; Hulslander, Olson, Willcutt, & Wadsworth, 2010; Kim *et al.*, 2010; Savage *et al.*, 2007; Verhoeven & Van Leeuwe, 2008).

According to Adlof *et al.*, (2010), there are numerous demographic and environmental factors which have a long association with comprehension. One highly predictive example

includes mother's level of educational attainment. Likewise, general intelligence has also been found to be a strong predictor (Hulslander *et al.*, 2010). In contrast to unique demographic variables, a survey of the professional literature reveals a pattern of frequently repeated predictors.

As an important component of decoding, numerous studies have implicated phonological awareness as being highly predictive of comprehension (Adlof *et al.*, 2010; Hulslander *et al.*, 2010; Kirby, Parrila & Pfeiffer, 2003; Knuth, 2011; Savage, Carless, & Ferraro, 2007). Likewise, vocabulary is frequently cited as having a strong association with ability to read for meaning (Adlof *et al.*, 2010; Hulslander *et al.*, 2010; Knuth, 2011). Nevertheless, the impact of these variables fluctuates with the child's developmental stage in reading (Adlof *et al.*, 2010; Knuth, 2011; Verhoeven & Van Leeuwe, 2008). More importantly, each factor has been utilized in evaluating the two subcomponents of comprehension outlined by Gough and Tunmer (1986) whose Simple View of Reading has been shown to be a stable predictor of comprehension throughout the elementary years (Savage & Wolforth, 2007).

Phonological awareness. The professional literature pertaining to phonological awareness can be confusing. Puffpaff (2009) illustrates this point, noting the complicated terminology that is mistakenly perceived as being synonymous. For instance, the terms phonological awareness, phonemic awareness and phonological sensitivity are frequently used interchangeably. Regardless of the confusing terminology, phonological awareness has been shown to have a strong relationship with later reading ability. As noted by Adlof *et al.*, (2010), "because of their relationship to word decoding skills, alphabet knowledge and phonological awareness are good predictors of early reading outcomes" (p. 333).

At the most basic level, phonological awareness involves the mental process of

recognizing the spoken sounds that combine to create words. Therefore, phonological awareness is an integral component of decoding as students convert graphemes into phonemes. This involves translating and manipulating written symbols (such as letters) into spoken sounds. However, the correspondence between graphemes and phonemes is not always one-to-one. For instance, there are numerous examples in which a combination of graphemes (such as /sch/ or /th/) is translated into a single phoneme (Ganske, 2000). Bear, Invernizzi, Templeton and Johnston (2008) further note that phonological awareness also involves the recognition of alliterative sounds and rhyming words together with the ability to identify the individual syllables in words.

Many research studies indicate that phonological awareness precludes a child's ability to comprehend. However, it cannot be determined whether the relationship is causal, correlational or reciprocal (Pufpaff, 2009). Nevertheless, Nithart *et al.* (2011) suspect a causal relationship between the two skills.

Much of the professional literature pertaining to reading disabilities in elementary students implicates the connection between phonological awareness and comprehension, and it is noted that understanding is often hindered by phonological weaknesses (Adlof *et al.*, 2010; Fuchs *et al.*, 2012; Kirby *et al.*, 2003). Summarizing the effects of phonological awareness on later ability, Adams *et al.*, (1998) found that deficits in this skill are already pronounced in disadvantaged preschool children. Moreover, the longitudinal impact is illustrated by the fact that "measures of schoolchildren's ability to attend to and manipulate phonemes strongly correlate with their reading success through the twelfth grade" (p. 2).

The work of Ryder *et al.*, (2008) also makes a strong case for the relationship between phonological awareness and comprehension. Their investigation took place in New Zealand

where literacy instruction typically takes the form of a constructivist, whole language approach. In a sample of 24 first and second grade students, it was found that structured phonological interventions helped students build decoding skills which in turn contributed to demonstrable gains in comprehension.

In their discussion of the impact of phonological processing on comprehension, Kim *et al.*, (2010) explain that “efficient word reading releases attentional resources to attend to meaning in text. Thus students who read dysfluently. . . expend their energy on identifying words rather than getting at meaning” (p. 653). Consequently, when students struggle with phonologic awareness, the result is usually a significant drop in a student’s level of comprehension (Hulslander *et al.*, 2010). While phonological awareness is an important precursor of decoding and comprehension in the early elementary years, Kirby *et al.*, (2003) concluded that the strength of the relationship begins a steady decline prior to the third grade year.

The relationship between phonological awareness and reading comprehension has widespread acceptance amongst practitioners in the classroom. However, some researchers (Hoiem-Tengesdal, 2010; Ouellette & Beers, 2010; Tilstra *et al.*, 2009) contend that the connection is not a simple one. Moreover, an emerging body of research suggests that other skills are stronger predictors of aptitude in reading as students enter the middle elementary grades (Fuchs *et al.*, 2012; Kim *et al.*, 2010; Nithart *et al.*, 2011).

Phonological Awareness Literacy Screening. The Phonological Awareness Literacy Screening was designed by researchers at the University of Virginia, and it is widely known by the acronym PALS. The test was developed with support from Virginia’s Early Intervention Reading Initiative and the Virginia Department of Education (Invernizzi *et al.*, 2013). The

assessment has also been endorsed by the National Center on Response to Intervention (2013). Information posted on its main web site indicates that PALS is used by 99% of the school divisions in the state of Virginia with more than 17,000 teachers administering the test each year (“PALS,” 2013). The PALS 1-3 assessment is a diagnostic tool that can help classroom teachers identify students who may be at risk for reading difficulties, and it can also offer insight as to what young readers should learn next in a progression of skills (Adlof *et al.*, 2010; Blackwell-Bullock, Invernizzi, Drake, & Howell, 2008-2009; Invernizzi, Landrum, Howell, & Warley, 2005).

Created in 1997, the PALS test is useful in the early identification of students who are reading below grade level. In providing feedback regarding mastery of fundamental literacy skills, PALS data can help teachers differentiate instruction in an effort to provide targeted skill remediation (Helman, 2005; Invernizzi, Landrum, Howell & Warley, 2005). Nevertheless, program developers strongly caution classroom teachers from an overreliance on PALS data in planning for instruction. For instance, Invernizzi *et al.*, (2013) note that “instructional decisions are best based on multiple sources of evidence: reading assessment data from other kinds of tests; reading group placement; lists of books read; and, most important, teacher judgment” (p. 5).

In 2000, the Virginia General Assembly approved funding to screen students in kindergarten through grade three with PALS. As a result, all students in kindergarten through second grade are required to be assessed annually. In grades K through 2, it is administered during an official testing window in the fall and again in the spring. In third grade, the test is optional and only occasionally administered to students who are new to Virginia schools.

According to the PALS 1-3 Technical Reference, the assessment is able to identify

several early predictors of later reading difficulties. These include measures of letter-sound recognition, alphabet knowledge and phonological awareness (Invernizzi *et al.*, 2013). The test itself consists of several tiers. The entry level is a general evaluation of the child's skill level in word recognition and spelling. Pupils who do not meet the established benchmark for their grade are administered the Level A assessments in order to gauge oral reading skills. This includes oral reading accuracy and fluency in addition to reading comprehension questions. Level B and Level C offer more in depth analysis as needed. PALS is administered to the student one-on-one, and there is no time limit for the performance items. The established benchmark for every grade should be thought of as a minimal proficiency level. At present, there are two main versions of the test administered in alternating years. These include form A and form B which are said to be comparable with one another.

The PALS assessment is utilized in nearly every school division in the state of Virginia, and it is frequently used in assessing objectives outlined in the state Standards of Learning (Invernizzi *et al.*, 2013). However, studies investigating the relationship between second grade PALS scores and the third grade reading SOL test have failed to find a significant correlation between the two (Gaither, 2008).

The research findings of the National Reading Panel (2000) indicate that decoding is contingent upon both phonological awareness and phonemic awareness. This includes skills such as blending. As a test of phonological awareness, the PALS test also evaluates decoding ability. The information found in Table 2 includes a description of the Virginia Standards of Learning decoding objectives that are assessed on the PALS test (Invernizzi *et al.*, 2013).

PALS is also a versatile instrument that has been used with many different student groups. For instance, the test has been utilized as a diagnostic instrument with English Language

Learners in Nevada (Helman, 2005). Likewise, it has been used in collecting baseline data in evaluating potential candidates for a reading remediation program in New York (Gattis *et al.*, 2010). Although it is widely used, in evaluating the instrument’s reliability and validity with diverse populations, the National Center on Response to Intervention (2013) only graded the PALS test as being “partially convincing.”

Table 2
Decoding Objectives Assessed on the PALS Test

PALS Task	Virginia SOL	Objective
Spelling	2.4a	Use knowledge of consonants, consonant blends, and consonant digraphs to decode and spell words.
Spelling	2.4a	Use knowledge of short, long, and r-controlled vowel patterns to decode and spell words.
Word Recognition	1.6e	Blend beginning, middle, and ending sounds to recognize and read words.
Sound to Letter	1.4	Orally identify and manipulate phonemes in syllables and multisyllabic words.
Letter Sounds	K.7b	Match consonant and short vowel sounds to appropriate letters.

Vocabulary. An emerging body of research pertaining to reading comprehension has focused on the predictive qualities of oral and receptive vocabulary. These variables were found to be strongly associated with a child’s ability to understand written text (Braze *et al.*, 2007; Kim *et al.*, 2010; Lervag & Aukrust, 2010; Protopapas *et al.*, 2013). Ouellette and Beers (2009) reported that oral vocabulary was a stronger predictor of comprehension than phonological awareness. Likewise, Nation and Snowling (2004) conducted a study of more than 70 students, ages 8-13, concluding that the predictive strength of a child’s oral language proficiency was similar to the contribution of phonological awareness.

By the time students first enter school, there is already considerable disparity in

vocabulary, and that divide only becomes more pronounced over the ensuing years (Kucan, 2012). Receptive vocabulary was also found to be highly correlated with comprehension regardless of student age (Ouellette, 2006). Additional research on the predictive strength of vocabulary suggests that the relationship between vocabulary and comprehension remains strong through adulthood. A broad vocabulary is said to aid comprehension in that it allows for increased scaffolding, the ability to generate inferences, mental recall of facts and summarizing ability (Verhoeven & Van Leeuwe, 2008).

Evidence of the predictive strength of vocabulary may also be found in research pertaining to comprehension in second language learners. Clearly, students who struggle to develop a command of the language will experience commensurate difficulty with reading comprehension. The work of Lervag and Aukrust (2010) demonstrates that native language speakers have a distinct advantage in comprehending written text when compared with peers who are second language learners. In such cases, the initial development and rapid growth in comprehension for native speakers is thought to be the result of preexisting differences in vocabulary.

In describing the predictive strength of vocabulary, DeVries (2010) provides several common sense explanations that contribute to its effect. The impact of environment on the development of a robust vocabulary is obvious. Students from educated households are said to have an average vocabulary of approximately 20,000 words by the time they finish kindergarten. In contrast, students from homes where formal education is lacking demonstrate a greatly reduced vocabulary of approximately 5,000 words. Parental modeling is thought to account for much of the difference between students. Parents who read to their children regularly do much to cultivate a strong command of the language as quality children's literature frequently

incorporates rich vocabulary to provide a sense of imagery for the reader. Similarly, parents who expose their children to educational television programming help develop strong oral and receptive language in their children.

The contribution of vocabulary to the Simple View of Reading is clear. According to Braze *et al.*, (2007), vocabulary is a central component of language comprehension. Whereas decoding involves the phonological skills that assist with word recognition, vocabulary is a critical component of language comprehension in the model put forth by Gough and Tunmer (1986). While listening comprehension has been offered as a measure of linguistic comprehension, vocabulary measures are stronger in their ability to capture additional variance in reading (Braze *et al.*, 2007; Protopapas *et al.*, 2013; Verhoeven & Van Leeuwe, 2008).

The value of using vocabulary as an indicator of reading comprehension is addressed within the *Reading Framework for the 2013 National Assessment of Educational Progress* (National Assessment Governing Board, 2013). As stipulated within the No Child Left Behind legislation, state reading assessments should not require students to identify the meaning of words in isolation, but rather as “the application of one’s understanding of word meanings to passage comprehension” (p. 33). Therefore, the NAEP recommends evaluating the understanding of vocabulary as a sub-skill of passage comprehension. Such items are included as a means of evaluating student ability to identify context specific word meanings which contribute to the central idea of the passage.

Recognizing vocabulary as a powerful predictor of comprehension, Kucan (2012) incorporates the work of Perfetti (2007) in recommending strategies to further develop this skill. Although vocabulary instruction is often superficial, Kucan advocates for a more holistic approach. Teacher modeling is a key element in creating a verbal environment where words are

celebrated and curiosity about language is fostered. Precision of language is cultivated as teachers incorporate words of a more sophisticated caliber across instructional settings. According to the Lexical Quality Hypothesis, students must be exposed to the orthographic, phonological and semantic representations of words (Perfetti & Hart, 2002; Perfetti, 2007). Therefore, reading practice reinforces quality lexical representations, and exposure to quality children's literature helps the student incorporate fresh vocabulary into their everyday language. Stories with rhyming words reinforce phonological understanding while spelling practice and word sorts facilitate the discovery of orthographic spelling patterns (Bear *et al.*, 2008; Ganske, 2000). Regardless of the specific strategy, a deliberate and intentional approach to vocabulary instruction is expected to yield sizable dividends on assessments of reading comprehension (Kucan, 2012).

STAR Reading Test. The STAR Reading Test is a commercial program designed by Renaissance Learning Inc. It is a nationally norm-referenced test that generates scores in a fifteen minute testing session. Program designers promote the product as an assessment that can be used to identify student reading levels, measure class growth and predict student outcomes on standardized reading assessments. On the cover page of a recent publication from Renaissance Learning (2011), it states that “the STAR assessments are highly rated for screening and progress monitoring by the U.S. Department of Education’s National Center on Response to Intervention!” It has also been promoted as a suitable assessment in the federal government’s Reading First and Title 1 intervention programs (Renaissance Learning, 2007).

The STAR Reading Test is used to calculate general reading comprehension scores for students in grades 1-12. The test can be administered with minimal effort, and it can be readministered as frequently as once per week as a means of documenting student growth. The

STAR test incorporates a simple design which requires students to read short passages and identify a single word deletion from among the four choices that are provided. There are two main types of test items. These include short comprehension items and extended comprehension items. Once completed, the program generates several reports providing student percentile ranks, reading levels, and grade equivalencies. It also provides a ZPD, or zone of proximal development, giving the student an optimal reading range to ensure continued growth.

Program designers note that the test is based on research implicating vocabulary as an accurate predictor of student comprehension (Renaissance Learning, 2011). Specifically, the use of the short comprehension items “is based on abundant and long-standing research verifying that vocabulary is closely tied to comprehension. . . short comprehension items contain one complete contextual sentence with a tightly controlled vocabulary level and a single-word deletion” (p. 15). The short comprehension items are also referred to as vocabulary-in-context items (Renaissance Learning, 2007). The individual items on the STAR test require students to evaluate the semantics and syntax of the sentence, gathering context clues that will prove beneficial in finding the answer that is the best fit. According to Iwata, Kojiri, Yamada and Watanabe (2011), this method, known as the cloze reading approach, is strongly influenced by depth of vocabulary knowledge and mastery of English grammar. In evaluating student vocabulary in context, the STAR assessment follows recommendations outlined by the National Assessment Governing Board (2013).

While the STAR Reading Test purports to be highly correlated with a host of standardized reading assessments, some information found in the professional literature appears to be contradictory (Boucher, 2005). In comparing percentile ranks obtained from the STAR Reading Test with corresponding percentiles from the sixth edition of the California

Achievement Test (or CAT 6), Boucher (2005) found a poor correlation. In that comparison, the STAR test repeatedly generated a much lower score. Nevertheless, information published by Renaissance Learning (2007) demonstrates that significant correlations were found between STAR and the California Achievement Test that included a sample of more than 300 elementary aged students.

Several studies illustrate STAR's reliability in predicting success on other standardized reading assessments (Adair, 2010; Churchwell, 2009; Renaissance Learning, 2007; Renaissance Learning, 2011). For instance, STAR test scores were used with middle school students to predict success on the Tennessee Comprehensive Assessment Program (Churchwell, 2009). Likewise, STAR proved to be a reliable predictor with sixth grade students who took the reading portion of the Texas Assessment of Knowledge and Skills, or TAKS (Adair, 2010).

In establishing concurrent validity for the assessment, Renaissance Learning (2007) includes a table of correlation values between the STAR Reading Test and other commonly used norm-referenced tests. For instance, significant relationships were found with the Iowa Test of Basic Skills, the Stanford 9 Achievement Test, the Missouri Master Achievement Test, and the Gates MacGinitie Reading Test. In their most recent publication pertaining to test reliability, Renaissance Learning (2011) noted that the instrument was correlated with standardized reading assessments from Alabama, Arizona, Colorado, Delaware, Florida, Georgia, Illinois, Kansas, Michigan, New York, Oklahoma, Utah and many other states. In total, program designers note that thirty studies have been conducted with more than 200,000 third grade students yielding a correlation coefficient of $r = .80$ for the instrument's predictive validity.

Qualitative Reading Inventory. A Qualitative Reading Inventory is one type of Informal Reading Inventory (or IRI) that is administered to individual students. According to Leslie and

Caldwell (2011), the QRI is designed to provide reading teachers with information regarding how students identify words and understand text. Likewise, the assessment also provides diagnostic information pertaining to the specific scenarios in which students fail to comprehend.

In utilizing a QRI, teachers are able to identify a child's reading level with precision. The assessment yields an independent reading level as well as instructional and frustrational levels. Comprised of word lists and short passages, the test is designed for students to read both orally and silently. Although it is a versatile instrument that can be administered to students in kindergarten through high school, it is also moderately time consuming. Administered one-on-one with each student, this form of informal reading inventory typically requires an allotment of 20 to 40 minutes to complete. In some cases, the process can take up to one hour.

Information obtained from the QRI is typically used for the purpose of placing students in heterogeneous reading groups or in selecting appropriately challenging literature for students to read independently. When it is retaken after a period of time, the QRI can also be used in gauging student growth in reading (Diehl, Armitage, Nettles, & Peterson, 2011). This assessment is a favorite among educators because of its flexibility in implementation. Leslie and Caldwell (2011) note that the testing proctor is afforded the latitude to make informed judgments based on the child's performance, and scores are "interpreted only in regard to the individual and not to any norm group" (p. 1).

The QRI is a preferred assessment among practitioners. In comparing several commonly used IRI tests, Nilsson (2008) describes the Qualitative Reading Inventory as a good choice based on available measures of reliability and construct validity. The QRI also incorporates examples of both narrative and expository text from social studies and science textbooks. It is recommended for use with all age groups from emerging readers through high school, and it

provides test administrators with useful information regarding the specific reading strategies that students employ as they read for meaning.

Owing to its ease of utility and the quality of formative data, the QRI is an instrument that many pre-service teachers are required to use in teacher education programs. Aspiring teachers are frequently taught to use the data in planning for differentiated reading instruction (Luttenegger, 2009). In assessing the unique abilities of each student, the QRI has been used in evaluating linguistic intelligence (Epelbaum, 2007).

In identifying candidates for remedial reading assistance, the QRI provides qualitative information regarding specific skill deficits (Mokhtari, Hutchison, & Edwards, 2010). This offers teachers the opportunity to plan for targeted assistance in an efficient manner. For instance, the Qualitative Reading Inventory can be used as a screening instrument to generate baseline data regarding comprehension levels. It can also be used to gauge student growth after the instructional intervention (Diehl *et al.*, 2011).

In spite of its simplicity and versatility, the Qualitative Reading Inventory has received some criticism regarding the passages that have been selected for inclusion. Wolpert and Vacca-Rizopoulos (2012) have demonstrated that the test design may yield skewed results, giving an elevated measure of the student's ability to comprehend text. The inclusion of expository passages is thought to engage prior knowledge of a topic thereby making the material easier to comprehend. Similarly, the predictable pattern inherent in narrative formats can also inadvertently impact comprehension by reducing miscues in the decoding process. Many of the concerns regarding inconsistent results with the QRI are shared by Epelbaum (2007) who suggests that student background knowledge and learning style preference may also adversely affect the outcome of the assessment.

Assessing reading comprehension. There is considerable disagreement regarding how to best assess a child's ability to read for understanding. At the root of the matter is a basic philosophical difference between legislators and educators. For instance, "there is often vehement disagreement about what constitutes appropriate evidence of achievement and equally passionate differences of opinion about how that evidence should be collected, analyzed, reported, and used to make instructional decisions" (Casbergue, 2011, p. 13). Unfortunately, many reading teachers perpetuate the ineffective practice of simply evaluating comprehension through impromptu questions in the form of book talks. Such an informal and superficial form of assessment is a better gauge of short-term memory than comprehension. According to Keene (2009), reading teachers do very little to actually teach students to comprehend more effectively, and many teachers rely on their own subjective opinion as a means of forecasting anticipated performance on standardized tests of comprehension. The value of this practice is inconsistent at best, and it is often times rather misleading. In predicting the reading ability of minority boys, Hinnant, O'Brien and Ghazarian (2009) found this subjective method to be remarkably unreliable.

Following the authorization of No Child Left Behind in 2001, educational assessment became a means by which to monitor the accountability of teachers and schools (U.S. Department of Education, 2002). Driven by the need to demonstrate Adequate Yearly Progress in the areas of math and reading, the demand for reliable and efficient assessment instruments increased exponentially. Moreover, a new market emerged for standardized tests which generated quantifiable data that would help measure growth while allowing for comparisons.

Over the last two decades, skills-based reading tests have become the norm. With increased pressure on student performance, assessment often drives instruction for classroom

teachers. The result is a tendency for reading teachers to plan for an instructional focus directed toward an isolated set of reading skills (Casbergue, 2011). For educators who subscribe to a developmental approach to assessment, standardized multiple-choice tests represent a stark contrast to qualitative assessments characteristic of a more constructivist approach to teaching. Stout (2009) describes this shift, noting "since the adoption of the No Child Left Behind Act in 2001, well-meaning authors of educational policy have changed teachers' curricula, pedagogy and schedules in ways that do not always support best practice" (p. 1).

According to Invernizzi, Landrum and Howell (2007), the push for research-based instruction is oddly reminiscent of post Sputnik, Cold War era initiatives. Curiously, the general citizenry has been remarkably supportive of this approach. With regard to reading assessment, there is often a disconnect between the practical needs of the teacher and the rigorous requirements for scientifically-based testing information. The resulting paradox is a tradeoff between reliability and validity. Therefore, the most reliable reading tests are often the least valid, and the most valid are frequently the least reliable.

Commonly assessed skills. As individual states have implemented standardized assessments in reading, textbook publishers have responded by offering instructional materials designed to help students refine their comprehension skills. However, teaching students comprehension strategies is nothing new. In cultivating specific skills, May (2011) notes that much of today's research has origins that go back to the work of F.B. Davis who had identified nine different comprehension skills as early as 1944. They included, among others, word meanings in context, main idea, summarizing, inferencing and author's purpose.

Today, the *Reading Framework for the 2013 National Assessment of Educational Progress* provides explicit requirements for states to use as guidelines in constructing reading

tests (National Assessment Governing Board, 2013). This document recommends the inclusion of specific “cognitive targets” which measure comprehension of both informational and literary text. The specific item types recommended by the NAEP reflect the unique reading processes students are expected to exhibit at various points in their development (in grades 4, 8 and 12). The three main domains include locate/recall, integrate/interpret and critique/evaluate items. These domains require students to demonstrate reading skills such as inferencing, main idea, author’s purpose, sequencing of events, and summarizing.

Many teachers take exception with such assessments, insisting that summarizing and identifying author’s purpose are strategies as opposed to skills. According to detractors, the strategies themselves should never be the ultimate goal of reading. Rather, they should simply facilitate a better understanding of text. Offering clarity on the issue, May (2011) notes that reading skills are employed automatically without conscious effort while strategies require deliberate, intentional effort to solve a problem or answer a question. This point is reiterated by Cain *et al.*, (2004) who describe the use of such strategies as a form of metacognition called comprehension monitoring.

The terms “strategies” and “skills” are often used synonymously amongst reading teachers. According to Afflerbach *et al.*, (2008), the continued proliferation of reading strategies as assessed items is the result of the increased reliance policymakers have on quantitative testing data. Beginning in the early 1990s, strategies such as main idea and author’s purpose regained renewed popularity as a result of their inclusion in basal readers. A decade later, such strategies were ubiquitous on standardized reading assessments following the authorization of No Child Left Behind.

In evaluating the most commonly assessed comprehension strategies, Pilonieta (2010)

compared the frequency of specific question types found in the elementary basal readers from five major publishers. The three most common categories included prediction, inferencing and summarizing. In a study of 20 elementary classrooms, similar results were reported by Ness (2011). Pilonieta (2010) further noted that textbooks were selected from California, Texas and Florida as 30% of our nation's basal reader purchases are from those states alone. This is important as many testing vendors rely on items found in reading textbooks when constructing comprehension items for standardized reading assessments. According to Cain *et al.*, (2004), making inferences is generally synonymous with drawing conclusions. In describing the importance of comprehension strategies, Ness (2011) states that questions which require students to make predictions, infer and summarize reflect important metacognitive processes which also enhance vocabulary and decoding skills.

The third grade Virginia SOL test of reading comprehension. In the state of Virginia, students are administered the third grade Standards of Learning (SOL) reading assessment. This test requires students to demonstrate proficiency in identifying the main idea of text, determining author's purpose and using context clues to determine the meaning of unfamiliar vocabulary. In their analysis of language arts curricula in all fifty states, Carmichael, Martino, Porter-Magee, and Wilson (2010) report that the reading standards in Virginia compare favorably with standards adopted by other states. Receiving a score of six out of seven in the area of content and rigor, the state is commended as the "standards for early reading are strong, addressing phonemic awareness, phonics, fluency, and comprehension" (p. 325). Assessing reading with increased rigor, the cover page of the 2010 grade 3 reading test blueprint states that the new standards "will be effective with the administration of the 2012-2013 English Standards of Learning (SOL) tests" (Virginia Department of Education, 2011).

The testing blueprint provides general information pertaining to the assessed skills and the number of items to be included for each reporting category. The third grade reading SOL test results are reported with a total score and separate sub-score for each reporting category. There are three reporting categories on the assessment. They include word analysis strategies and word reference materials, demonstrate comprehension of fictional texts and demonstrate comprehension of nonfiction texts. The multiple-choice assessment includes 40 items. These include 17 questions assessing ability to comprehend fiction, 16 items for nonfiction, and seven word analysis items.

Each of the three reporting categories incorporates several standards listed in an alphanumeric format. Although teachers are required to cover all of the standards, there are numerous items which are excluded from the assessment because the testing format is not conducive to evaluating mastery of those skills. Objectives not assessed include: making connections between previous experiences and reading selections and setting a purpose for reading. These two objectives reflect the Department of Education's attempt to integrate constructivist philosophy into the curriculum, although they are not easily measured in a multiple-choice assessment.

Each year, the Virginia Department of Education releases many versions of the same third grade reading test. However, the exact number of tests is unknown, and it may vary from one year to the next. In addition, each version is likely to contain a different sampling of testing items from the various reporting categories (Virginia Department of Education, 2011).

Reading selections for the third grade SOL resemble passages commonly found in age appropriate basal readers, trade books and magazines for children. Relevant standards that assess student ability to comprehend fiction include SOL 3.5 c-j. These include making

predictions, determining author's purpose, drawing conclusions, and identifying the main idea. Similarly, SOL 3.6 c-i test the ability to comprehend nonfiction. Although some of the standards are similar to those assessed while reading fiction, others include summarizing and contrasting biographies and autobiographies. Examples of excluded standards include SOL 3.5 b (making connections between previous experiences and reading selections) and SOL 3.5 m and 3.6 l (reading with fluency and accuracy). Word analysis standards include 3.4 b (using knowledge of roots, affixes, synonyms, and antonyms) and 3.4 d (using context to clarify meaning of unfamiliar words). Although other standards may occasionally be tested, these are the most commonly assessed (Virginia Department of Education, 2010; Virginia Department of Education, 2011).

Summary

The increased accountability of federal legislation governing education has tremendous implications for third grade reading. Today, raising achievement standards in reading has become a national priority. For third grade teachers, data-driven decision making is absolutely critical as there is often a lack of beneficial information to aid in the prediction of student outcome on standardized assessments of comprehension. The transition to third grade often coincides with an abrupt developmental shift. This can be difficult for students to negotiate as they move from decoding and fluency based instruction to reading for meaning.

As third grade teachers sift through relevant testing information at the beginning of the school year, they must use that data in the most strategic way possible. In utilizing assessments with a proven research-based foundation, reading teachers will be able to make sound judgments regarding each student's individual progress. This study incorporates instruments which evaluate student growth on critical subcomponents of comprehension identified through a survey

of the professional research literature. With reliable testing data, teachers can diagnose reading difficulty long before end-of-course assessments are administered. As the format of the SOL reading test is similar to other state instruments used throughout the country, the results of this study are expected to have wide generalizability beyond the Commonwealth of Virginia.

This research investigation incorporates three commonly used assessments as predictor variables of reading comprehension. More specifically, the second grade end-of-year PALS assessment, the STAR Reading Test and the QRI-5 each serve as independent variables. In addition, the third grade Virginia SOL reading test was administered as the criterion, or dependent variable.

While each of these assessments has been the subject of other investigations, this researcher has been unable to identify another study that has compared the strength of all three assessments as outcome predictors on the same test. In addition to comparing the strength of each independent variable, this study further examines the combined predictive strength of all three tests when used together. Given their prevalent use and widespread acceptance, this study addresses a significant gap in the professional literature.

This investigation also evaluates the extent to which each predictor is related to the other independent variables. Therefore, the results of this project demonstrate the degree to which the individual predictors measure the same reading constructs. In measuring the redundancy of predictor variables, this study provides insight concerning the relationship between vocabulary and other prerequisite skills of comprehension. Given that vocabulary has been implicated as an integral component of the Simple View of Reading (Gough & Tunmer, 1986) and the Lexical Quality Hypothesis (Perfetti & Hart, 2002), this investigation extends the work of Protopapas *et al.*, (2013) in blending the common elements of both models.

In 1986, Gough and Tunmer stated that their intent was to present their “case more clearly, in the hope that its truth or falsity might be decisively settled by future research” (p. 6). While their original model posited that decoding and language comprehension were of equal importance, later variants of the model established that the role of decoding weakens as students develop more sophisticated reading skills (Gough *et al.*, 1996; Hoover & Gough, 1990). In measuring the differential impact of decoding and vocabulary on comprehension at the end of third grade, this study also expands upon the research findings reported by Tilstra *et al.*, (2009) and Verhoeven & Van Leeuwe (2008).

By the time students reach the fourth grade, one of three students is already reading below grade level. A significant portion of those students remain unidentified because of their strong decoding skills (Katzir *et al.*, 2006). Sadly, the majority of our nation’s classroom teachers lack the appropriate training to provide adequate remedial assistance (Bailey & Drummond, 2006; Joshi *et al.*, 2009b). Therefore, it is imperative that reading teachers acquire a more thorough understanding of the emergence of comprehension, and of the assessments that are used to gauge developmental skills in reading. Even more disturbing, there is reason to believe that college faculty who help prepare educators for the classroom often have only a rudimentary understanding of early linguistic concepts, and many textbook publishers neglect the research implications of influential studies such as the National Reading Panel (Joshi *et al.*, 2009a; National Reading Panel, 2000). Therefore, the outcome of this investigation is expected to help bridge the divide between theory and classroom application.

This chapter has addressed the development of comprehension and prerequisite skills within the context of known theoretical models. In documenting the way in which commonly used assessments evaluate the predictive constituent skills of comprehension, this study is

expected to be of interest to a wide audience of instructional leaders. It is anticipated that the results of this investigation will help elementary reading teachers incorporate data-driven decision making as they plan for instruction and remediation commensurate with student needs. Finally, as a result of this study, school and division administration can better plan for effective professional development and the efficient allocation of fiscal resources. Chapter three provides a detailed presentation of the methodology of the current study.

CHAPTER THREE: METHODOLOGY

Third grade marks a transitional time for elementary school students. As stipulated by the federal No Child Left Behind legislation of 2001, students are assessed in reading each year in grades three through eight (Ravitch, 2009). Standardized reading assessments rely heavily on testing items designed to measure student comprehension. According to the Simple View of Reading put forth by Gough and Tunmer (1986), reading can be represented by the formula $R = D \times C$. Their theoretical construct stipulates that reading comprehension (R) requires the reader to both decode words in print (D) and recognize it as a meaningful word in their spoken language (linguistic comprehension, or C). A similar theoretical construct was put forth by Perfetti and Hart (2002) whose Lexical Quality Hypothesis holds that comprehension is contingent upon the quality of orthographic, phonological and semantic word representations. Each of these models continues to have tremendous appeal today.

This research investigation incorporates a sample of $n = 84$ third grade students from a semi-rural elementary school in Virginia. The design makes use of three diagnostic reading assessments as predictor variables. Each instrument assesses subcomponents of comprehension as outlined in the Simple View of Reading and the Lexical Quality Hypothesis. Using a non-experimental design, Pearson correlation coefficients are generated as a means of evaluating the magnitude of the relationship between each predictor and student scores on the third grade SOL reading test. Likewise, multiple regression analysis is utilized in an effort to investigate the relationship between a linear combination of all predictor variables and the criterion. Moreover, a stepwise multiple regression procedure helps to further refine the optimal regression equation between the predictors and the outcome variable. Finally, the unique contribution of each predictor is evaluated as a means of determining its predictive strength.

Design

The present study utilizes a correlational design. This investigation was conducted in an effort to examine the relationship between three predictor variables and an outcome measure (the criterion). Further analysis was also performed to reveal the significance of the model in predicting achievement scores on the third grade SOL test. According to Gall, Gall and Borg (2010), one purpose of non-experimental research designs “is to search for variables, measured at one point in time, that predict a criterion variable measured at a subsequent point in time” (p. 331). Thus, Pearson correlations were generated to determine the strength of the relationships between variables, and multiple regression analysis determined the predictive significance of the model when the variables are used simultaneously. The first predictor variable is the Phonological Awareness Literacy Screening. Otherwise known as PALS, this assessment is administered at the end of second grade. The other predictor variables include scores obtained on the Qualitative Reading Inventory - 5 (QRI-5) and the STAR Reading Test. Finally, the dependent variable is the score achieved on the Virginia Standards of Learning assessment of reading comprehension.

A multiple regression design is useful in predicting assessment scores in instances where there is more than one predictor variable. When there are multiple variables which are likely to be correlated with the outcome measure, this design is helpful if there is a need to “tease out the separate roles of the predictors” (Howell, 2011, p. 261). As it applies to this study, multiple regression analysis is used to evaluate the magnitude of the relationship between each of the predictor variables and student outcome on the third grade Standards of Learning reading assessment. Moreover, multiple regression analysis was used to determine if the linear combination of predictor variables is significantly related to achievement scores on the SOL test.

Questions and Hypotheses

The present study was designed to evaluate the following research questions and corresponding null hypotheses:

Research Question 1: What is the relationship between diagnostic reading measures obtained at the beginning of the year with achievement scores on the third grade Standards of Learning assessment in reading?

Null Hypothesis 1: There will be no significant correlation between scores on the Phonological Awareness Literacy Screening and scores on the Virginia Standards of Learning reading assessment.

Null Hypothesis 2: There will be no significant correlation between scores on the STAR Reading Test and scores on the Virginia Standards of Learning reading assessment.

Null Hypothesis 3: There will be no significant correlation between scores on the Qualitative Reading Inventory and scores on the Virginia Standards of Learning reading assessment.

Research Question 2: What is the relationship between a linear combination of three predictor variables (the joint model comprised of PALS, STAR and QRI) and achievement scores on the third grade Standards of Learning assessment in reading?

Null Hypothesis 4: There is no significant relationship between a linear combination of the three predictor variables and achievement scores on the third grade Virginia SOL test in reading.

Null Hypothesis 5: There will be no difference in the amount of SOL reading variance captured by the PALS, STAR and QRI reading measures.

Null Hypothesis 6: No linear combination of variables predicts achievement scores on the

third grade Virginia SOL reading test better than the linear combination of all three predictor variables utilized simultaneously.

Participants

The sample of students was recruited from a large elementary school in a semi-rural school division in Virginia. Because the combination of assessments serving as predictor variables was unique to this school, nonprobability convenience sampling was utilized. Although many other schools use one or more of these tests, no other identified school used the same battery of assessments in a uniform way with all third grade students.

The school attendance area is the largest in the division with a student enrollment of approximately 605. This study sample included all students who transitioned from second grade to third grade in the fall of 2012 ($n = 84$). Table 3 includes demographic information for the sample of students.

Table 3
Demographic Information

	Male	Female	
Ethnicity	African American	1	0
	Asian American	0	1
	Caucasian	30	28
	Hispanic	9	13
	Native American	2	0
English Language Learners	7	13	
Special Education Students	1	0	

The school serves a diverse student constituency, and this is somewhat atypical of the other schools in the division. The majority of the families in the attendance area have moved to

the region seeking employment in higher education, healthcare, pharmaceuticals and agriculture. While the school serves students from several affluent subdivisions, it is also a Title 1 school, and 48.5% of students are economically disadvantaged, receiving free or reduced lunch. Approximately 23% of the students in this school are English Language Learners. While most ELL students come from homes where Spanish is the native language, there are also many students whose first language is Russian. The heterogeneous nature of this student sample is expected to contribute to the generalizability of the research findings to the target population of third grade students throughout the country. Table 3 provides concise demographic information for the study sample.

The researcher presently serves as the assistant principal in this school. As this research investigation does not incorporate an experimental design with a treatment and control group, the subjects who comprise the sample were not made aware of the present study. Rather, the design incorporates archival testing data that was used in the course of everyday instructional practice. This approach was utilized as there was concern that subject awareness could adversely affect the outcome of the study resulting in a Hawthorne Effect (Gall *et al.*, 2010).

According to the recommendations of Tabachnick and Fidell (2007), the sample of $N = 84$ is sufficient in generalizing results to the larger population. The formula in determining required sample size for multiple regression is $N \geq 50 + 8m$. In this formula, m represents the number of predictor variables. Therefore, the required sample size for a study that includes three predictors is $N \geq 50 + 24 = 74$. Gall *et al.*, (2010) employ a more liberal standard, recommending inclusion of no fewer than 15 participants for each independent variable. Accordingly, this requires “a sample of at least 45 individuals for a multiple regression analysis involving three predictor variables” (p. 361).

Setting

The study took place in a mid-sized semi-rural school division in Virginia. There are 15 elementary schools, and 25 total schools in the division serving nearly 12,000 students. This investigation took place in the division's largest elementary school which opened in August of 2008. Although the school has approximately 605 students in grades PK - 5, the building was constructed with a capacity of 800. There are 105 employees in the school with 31 grade level teachers, one principal and one assistant principal. Roughly 38% of the licensed staff members hold a master's degree. Of the teachers, 65% have been teaching for fewer than 15 years while 35% having more than 15 years of experience.

A significant portion of the students at this school are English Language Learners. In addition, the school is also home to a regional hearing impaired program and a regional program for students with emotional disabilities. This school also has a program for early childhood special education (ages two through five).

Instrumentation

The present study incorporated reading assessment data obtained between May of 2012 and May of 2013. In all, three predictor variables were collected between the end of second grade and the beginning of third grade. The PALS assessment was administered in May of 2012, at the culmination of the second grade year. Likewise, QRI-5 and STAR Reading Tests were given in August and September of 2012. Finally, students took the Virginia reading SOL test (the criterion) in May of 2013.

Phonological Awareness Literacy Screening. Second grade teachers administered the PALS reading assessment in May of 2012. Student scores were compared against PALS data that had been collected during the first grading period. In the fall of 2012, third grade homeroom

teachers collected second grade end-of-year PALS data as they began to place students into homogeneous reading groups based on their individual needs. The PALS test has been used in the present study to assess student decoding skills, one of the subcomponents of reading comprehension common to both the Simple View of Reading put forth by Gough and Tunmer (1986) and the Lexical Quality Hypothesis of Perfetti and Hart (2002).

The PALS assessment is a criterion referenced test that is administered to students one-on-one by the second grade homeroom teacher. Test administration generally takes between 23-43 minutes per student (National Center on Response to Intervention, 2013). The test is given in a discrete location, free from distractions. Meanwhile, substitute teachers were hired to continue instruction with the remainder of the class. During test administration, students demonstrate mastery of concepts in several domains including orthographic knowledge, oral reading in context, alphabets and phonemic awareness. Each domain is subdivided into untimed performance-based tasks. When the test is completed, benchmark scores are calculated in order to determine a minimal competency level in each domain, while a total PALS score is calculated from the sum of each of the four domains.

In reporting reliability and validity data, it was noted that a broad sample of $n = 6,392$ students in grades one through three were included in the pilot and field testing process conducted in 2004. While pilot testing the assessment, an effort was made to incorporate a student sample which would closely approximate the demographic information reported for the state in terms of ethnicity, socioeconomic background, and gender.

As the PALS test is administered and scored by the classroom teacher, a fair amount of judgment is involved in the scoring process. This is further complicated by the fact that there are numerous subtasks, and the administration of the test can be time consuming. Therefore, inter-

rater reliability is a concern. While measures of inter-rater reliability are included for several of the individual subtasks, such data is not provided for the total PALS score. In addition, it should be noted that reliability data was collected and reported using different samples at different times.

In reporting on the inter-rater reliability of the various PALS subtasks, data from 1997-2002 was collected and evaluated using a Pearson correlation. With the exception of oral reading in context, all coefficients levels were both high and significant, ranging from .94 to .99 ($p < .01$). Inter-rater reliability coefficients for the second grade oral reading subtask was reported as .85 in 2002 ($n = 38$) with a higher score of .96 reported in 2000 ($n = 50$). These values were also determined to be significant at the $p < .01$ level. Therefore, it is believed that the test can be administered and scored accurately and reliably by two different testing proctors (Invernizzi *et al.*, 2013). It is also noteworthy that with a subsample of $n = 204$ students, estimates of test-retest reliability were also found to be both significant and high.

Considerable data was also reported for measuring the construct validity of the PALS test. This reflects the extent to which the assessment actually measures the underlying construct or theoretical model. In this case, the PALS test was designed to assess the intersection of speech sounds and knowledge of print. In order to evaluate construct validity, a principal components analysis and discriminant analysis was conducted on the theoretical model. According to Invernizzi *et al.*, (2013), PALS evaluates “the pronunciation of letter sounds, the ability to match letters and letter patterns to speech segments to produce a spelling, and the ability to recognize words” (p. 37). Taken together, there is considerable evidence suggesting that the PALS test assesses a single construct.

Measures of predictive validity were also included for the PALS assessment. Correlation

data was collected establishing a relationship between beginning of year PALS scores and end-of-year reading scores on the Virginia Standards of Learning test (SOL) in 2001. It was reported that high summed PALS scores from the fall were associated with strong SOL scores in the spring. A correlation of .60 ($p < .001$) was obtained with a sample of $n = 277$ students.

STAR. The STAR Reading Test was administered during the first week of school in an effort to establish baseline reading data for third grade students. Teachers took homeroom students to the computer lab, and all students took the test simultaneously. The ten minute assessment provided teachers with grade equivalencies, student percentile ranks and ZPD scores. This information was used to help place students into reading groups while providing a range of suggested literature that would appropriately challenge each student's continued growth in reading. In providing reliability and validity data for their product, Renaissance Learning (2011) notes that the STAR Reading Test has received the highest rating possible from the National Center on Response to Intervention, a federally funded organization which promotes data-driven instructional decision making (National Center on Response to Intervention, 2013). With an emphasis on word knowledge and vocabulary, the STAR test has been utilized as a measure of linguistic comprehension, an important subcomponent of reading comprehension as described by Gough and Tunmer (1986). The STAR test has also been included due to its ability to assess word meanings and semantic word knowledge, a critical factor in the Lexical Quality Hypothesis by Perfetti and Hart (2002)

In establishing test-retest reliability for STAR in 1999, more than 30,000 students were administered the assessment during the norming process (Renaissance Learning, 2011). This included students from nearly 300 schools in 47 different states. The assessment was renormed in 2008, including nearly 70,000 students in grades 1-12 from 48 states and the District of

Columbia. The most recently released test-retest reliability data for third grade students included a sample of $n = 223$, yielding a correlation coefficient of .87, indicating a very strong relationship. Likewise, a sample of $n = 476$ third grade students revealed an alternate forms reliability coefficient of .86 (Renaissance Learning, 2007).

In order for the STAR Reading Test to be a valid instrument, it must correlate well with other commonly used assessments of reading achievement. To establish concurrent validity, participating schools submitted STAR scores together with corresponding data from other assessments that are commonly administered throughout the country. The data was collected and disaggregated, reporting individual correlation coefficients with a wide array of various instruments. The available data varied from one grade level to another, with results published for all grades 1-12. Third grade data incorporated more than 200 students who had taken the Iowa Test of Basic Skills with correlation coefficients ranging from .70 to .82. Likewise, 117 students took the STAR and the TerraNova reading assessment yielding a correlation of .78. For 67 students who took the STAR Reading Test and the Stanford Achievement Test, a correlation of .79 was obtained. Other smaller samples were also included for third grade students, and all resulted in statistically significant relationships, thereby ensuring concurrent validity. Considerable data regarding the test's reliability and validity has also been reported by the National Center on Response to Intervention (2013).

QRI. Within the first month of school, the Qualitative Reading Inventory – 5 was administered to each student in the third grade. Homeroom teachers administered the test one-on-one for the purpose of obtaining independent reading levels for each student. Afterwards, the data was used for the purpose of placing students into homogenous reading groups. Each test took approximately 45 minutes to administer. Leslie and Caldwell (2011) provide detailed data

pertaining to inter-rater reliability and construct validity. The QRI-5 has been included in this study due to its ability to assess critical subcomponents of reading comprehension as outlined in the Simple View of Reading (Gough & Tunmer, 1986) and the Lexical Quality Hypothesis (Perfetti & Hart, 2002).

Owing to the fact that scoring a QRI involves an element of judgment on the part of the examiner, inter-rater reliability is critical. Leslie and Caldwell (2011) report that reliability data was obtained from a sample of $n = 3$ specialists who evaluated 49 oral reading passages. In analyzing inter-rater reliability data for the QRI-5, Cronbach's alpha levels were reported for total miscues (.99), meaning-change miscues (.99), explicit comprehension (.98) and implicit comprehension (.98).

Measures of construct validity were also reported for the QRI-5. This reflects the extent to which test scores on the QRI actually measure the construct being evaluated. In this case, scores pertaining to word recognition and comprehension were evaluated based on the conceptual framework of Gough and Tunmer (1986). Interrelationships among word identification, reading accuracy, semantic accuracy and reading rate were calculated. Regardless of grade level, correlation values ranged from $r = .34$ to $r = .59$, and all were found to be both positive and significant at the $p < .001$ level. Likewise, a statistically significant correlation was also calculated for prior knowledge and comprehension, with $r = .39$ at $p < .02$. In summary, Leslie and Caldwell (2011) note that "we have evidence that the QRI-5 measures at least two constructs that have been posited to be central to the reading process – word recognition and comprehension" (p. 487).

Virginia Standards of Learning (SOL). This instrument is the dependent variable of the present investigation. The third grade reading Standards of Learning Test is administered to

students in May of each school year. Similar tests are also administered to students in fourth and fifth grade in keeping with the No Child Left Behind legislation of 2001. According to the third grade testing blueprint, the recently adopted 2010 standards for reading were assessed for the first time in May of 2013 (Virginia Department of Education, 2011). This document states that the third grade curriculum is designed to evaluate the student's ability to employ reading comprehension strategies in reading fiction as well as nonfiction. In addition, students are also tested on word analysis strategies. The third grade SOL test in reading is a criterion-referenced test which includes forty multiple-choice items. Of those, 33 questions pertain to comprehension. As a measure of comprehension, the SOL test has been included in the present study as the outcome variable representing R (for reading) as outlined in the theoretical construct posited by Gough and Tunmer (1986).

Test items are developed by content experts who are employees of Educational Testing Service, or ETS. After field testing, those items are evaluated by members of content review committees, comprised of educators from throughout the state of Virginia. Employees of the Virginia Department of Education also scrutinize multiple-choice test items before they are included in testing. Finally, ETS works in conjunction with Pearson Psychometrics to provide the Virginia Department of Education with approved testing questions (Virginia Department of Education, 2010).

The SOL test is administered in a computerized testing format. The test is untimed, and students may have as much time as they need in one testing session. A perfect score on the third grade reading SOL test is a 600. A score of 500 or more is considered "pass/advanced." In order to achieve a passing, or "proficient," score, the student must earn a 400 or higher. These results are obtained by converting raw scores to scaled scores.

Reliability and Validity data were found in the most recent Standards of Learning Technical Report (Virginia Department of Education, 2010). As the third grade reading SOL test is a multiple-choice assessment, the VDOE is mainly concerned with the internal consistency reliability. In establishing internal consistency, a Cronbach's alpha coefficient was utilized in order to determine the degree that scores remain consistent when administered more than once. With a sample of $N = 38,824$ online testing students, a very desirable alpha level of $\alpha = .85$ was achieved.

As a measure of student achievement, the Virginia Department of Education includes data regarding content and construct validity for the third grade reading SOL. The test's content validity rests on its ability to assess the appropriate material from the prescribed curriculum. With this test, content validity was ensured by having educators from Virginia, VDOE employees, and specialists from ETS and Pearson negotiate item inclusion based on the SOL testing blueprint, the SOL curriculum framework as well as the actual Standards of Learning. The test's construct validity ensures that the third grade reading SOL is consistent with recognized theory in reading. In this case, student scores on the reading test were correlated with the Stanford 9 and the Literacy Passport Test, two widely used assessments which measure reading comprehension. The relationship with both tests was found to be statistically significant with correlation coefficients of $r = .76$ and $.78$ respectively.

Procedures

The present study incorporates testing data collected from third grade students in a public school in Virginia, thus necessitating approval from the university's Institutional Review Board. In addition, the study also required the consent of the division Assistant Superintendent as well as the Director of Student Assessment. Once permission was granted by all required parties, the

data set was collected.

At the beginning of the 2012-2013 school year, third grade students were assigned to one of four homeroom teachers in a heterogeneous format by school administrators. Prior to the first day of school, third grade teachers collected second grade end-of-year PALS data from student cumulative records. During the first week, each third grade teacher coordinated a time to take their class to the computer lab for the purpose of administering the STAR assessment to their homeroom students. Teachers collaborated with the school's computer lab assistant to ensure that the test was administered properly.

During the month of September, third grade teachers administered a QRI to each student in their homeroom. Initial training for the Qualitative Reading Inventory was conducted by the school's reading specialists to ensure uniformity and the accurate collection of reading data. Afterwards, school administrators instructed teachers to consolidate all reading assessment data into a spreadsheet that could be updated periodically. Once all information was available, the school's testing coordinator consolidated all assessment data into one spreadsheet.

Upon receiving approval to proceed with data collection, the researcher requested third grade assessment information from the school's testing coordinator. Second grade PALS data was collected from student cumulative files and incorporated into one master spreadsheet. Prior to receiving any information from the testing coordinator, all data identifiers were redacted so that it was impossible to identify individual pupils. With student anonymity preserved, access to data was password protected at all times. This study required no contact with individual students, and there was no treatment or intervention involved. In preparing the data for analysis, the Microsoft Excel spreadsheet was loaded into SPSS (Statistical Package for Social Sciences) version 19.0.

Data Analysis

In order to investigate the research questions of the current study, a multiple regression analysis was performed. The independent variables include the Phonological Awareness Literacy Screening, the STAR Reading Test, and the Qualitative Reading Inventory – 5. The dependent variable was the third grade Virginia Standards of Learning test of reading. Raw data was compiled by third grade teachers and collected in a Microsoft Excel spreadsheet. The data was then stored and shared amongst the teachers and the school’s testing coordinator on a web-based document storage site. Following IRB approval, the raw data was obtained from the testing coordinator and evaluated with the SPSS statistics software package.

According to Howell (2011), a multiple regression design is appropriate in evaluating the strength of the relationship between each predictor variable and the dependent variable. Likewise, this analysis is also useful in generating a regression equation, allowing the researcher to determine how a “linear combination of two, three, four, or more predictors will predict the criterion” (p. 269). Finally, a multiple regression design can evaluate the redundancy, or correlation between each of the independent variables. In cases where there is a concern for highly correlated predictors, a stepwise multiple regression analysis can be incorporated to further refine an optimal model, eliminating predictors that do not make a significant contribution to the regression equation (Gall *et al.*, 2010; Howell, 2011).

Pearson correlations were generated to determine the strength of the linear relationship between each of the three independent variables and the criterion, or dependent variable. Multiple regression analysis was conducted in an effort to evaluate the significance of the linear relationship between the model of predictors and the outcome variable, and the squared correlation coefficient (R^2) was reported as a measure of the variation in the criterion that is

attributable to the model of predictor variables taken together. Likewise, the multiple regression analysis was also useful in calculating the contribution of each predictor variable and its statistical significance (p). At the $p < .05$ level, each variable that was found to be significant was said to provide a significant contribution in predicting scores on the third grade Virginia Standards of Learning test in reading. Lastly, standardized regression coefficients and partial correlation values (β) were also calculated for each diagnostic instrument, providing an estimate of each independent variable's unique contribution in predicting the criterion. After conducting the initial multiple regression analysis, a stepwise multiple regression analysis was performed in an attempt to optimize the predictive quality of the regression equation.

Prior to conducting multiple regression analysis, assumptions of linearity, normality and homoscedasticity must be satisfied. In checking for the assumption of linearity, scatterplots of the raw data were generated to reveal the strength of the linear relationship between each of the independent variables and the dependent variable. A scatterplot of the standardized residuals with standardized predicted values was then generated to check for a homoscedastic distribution of the variance in error. A histogram of the residuals was constructed to check the assumption of normality, and collinearity statistics have also been included to check for variable redundancy. In ensuring that each predictor variable provides a unique contribution to the criterion, tolerance levels and variance inflation factors (or VIF) have been included for each predictor.

Summary

This research investigation incorporates a sample of students from a large elementary school in Virginia. This school was selected based on its uniform application of diagnostic reading instruments at the beginning of third grade. The data set was evaluated using Pearson correlations and multiple regression analysis, and the findings are presented in chapter four.

CHAPTER FOUR: FINDINGS

Chapter four presents the purpose of the present study together with a discussion of the research methodology and results of the statistical analysis. The research data is evaluated and results are discussed by research question and hypotheses. In this investigation, the dependent variable was student scores on the third grade Virginia Standards of Learning reading assessment. The study also utilized three diagnostic reading measures as predictor variables. Pearson correlations serve as an indicator of the strength of the relationship between each predictor and the outcome variable. Results were determined to be statistically significant at the $p < .01$ level. In addition, standard regression analysis was performed in an effort to determine how much of the outcome variance can be attributed to the model of the three predictors when used simultaneously. Likewise, the data was also evaluated to ascertain the unique contribution of each independent variable. Lastly, a stepwise multiple regression analysis was used to further refine the regression equation.

Purpose of the Study

The purpose of this investigation was to examine the strength of the relationships between three diagnostic reading measures and scores obtained on the third grade Virginia SOL reading test. Specifically, this study was designed to compare the practical utility of the Phonological Awareness Literacy Screening, the STAR Reading Test and the QRI-5. The study was also designed to evaluate the predictive strength of the model utilizing all three independent variables simultaneously.

Third grade coincides with a critical transition in reading as it is the first time federal mandates require states to assess reading comprehension. Therefore, the outcome of this study is expected to add to the body of research literature relevant to the effectiveness of diagnostic

reading assessments. As valuable instructional time and fiscal expense are consumed in assessing students, it is imperative that teachers and administrators make strategic use of screening instruments that are both efficient and effective.

Overview

The present study took place in a semi-rural public school division in Virginia. The school serves a diverse community, and the study sample incorporated secondary archival testing data for all students who took the third grade SOL reading test at the conclusion of the 2012-2013 school year. The data set included corresponding end-of-year second grade PALS scores from the previous school year together with STAR Reading and QRI data gathered at the beginning of third grade.

Data Analysis and Methodology

This research investigation utilized Pearson Product correlation coefficients to answer research question number one. This bivariate statistic is frequently used in educational studies due to its ability to generate continuous scores with minimal standard error (Gall, Gall & Borg, 2010). Because research question two attempts to address hypotheses involving a combination of predictor variables, a multiple regression analysis was performed. A preferred form of analysis for educational research, the main advantage of multiple regression is its versatility and the amount of useful data it produces (Gall *et al.*, 2010; Howell, 2011). Finally, a stepwise multiple regression was used with null hypothesis six. In combining stepup and stepdown multiple regression analysis simultaneously, a stepwise procedure is useful in generating an optimal regression equation (Howell, 2011).

Standard multiple regression analysis was performed utilizing the Statistical Package for Social Sciences (SPSS) version 19.0. Analysis was conducted between third grade SOL scores

(as the dependent measure) and three diagnostic reading instruments as independent variables (PALS, STAR and QRI). A preliminary examination of the data set revealed a small number of missing items. Of the 336 necessary data points, there were five missing values, or roughly 1.4% of the entire data set. Opting not to eliminate participants with incomplete data, SPSS imputed mean scores as estimates for missing data (Gall *et al.*, 2010).

Prior to conducting data analysis, assumptions were tested to check for violations of linearity, normality and residual homoscedasticity. Scatterplots were generated between each independent variable and the criterion (SOL test results). In each case, a line of best fit was superimposed on the distribution of data, indicating that the assumption of linearity was tenable. The following scatterplots reveal a linear relationship between each of the diagnostic measures and the third grade Standards of Learning test.

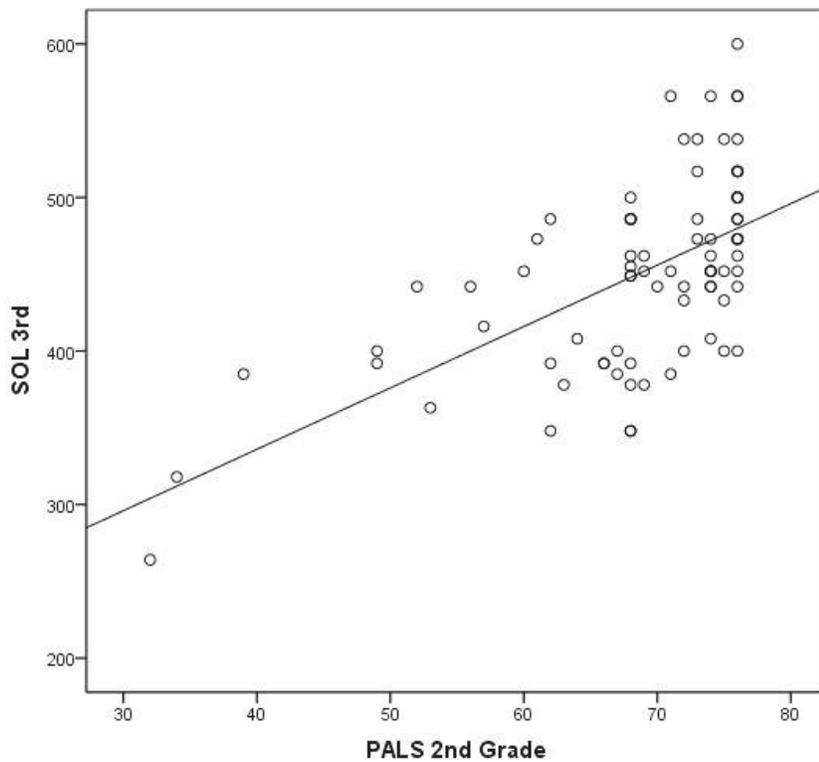


Figure 1. Scatterplot of the linear relationship between 2nd grade PALS and 3rd grade SOL.

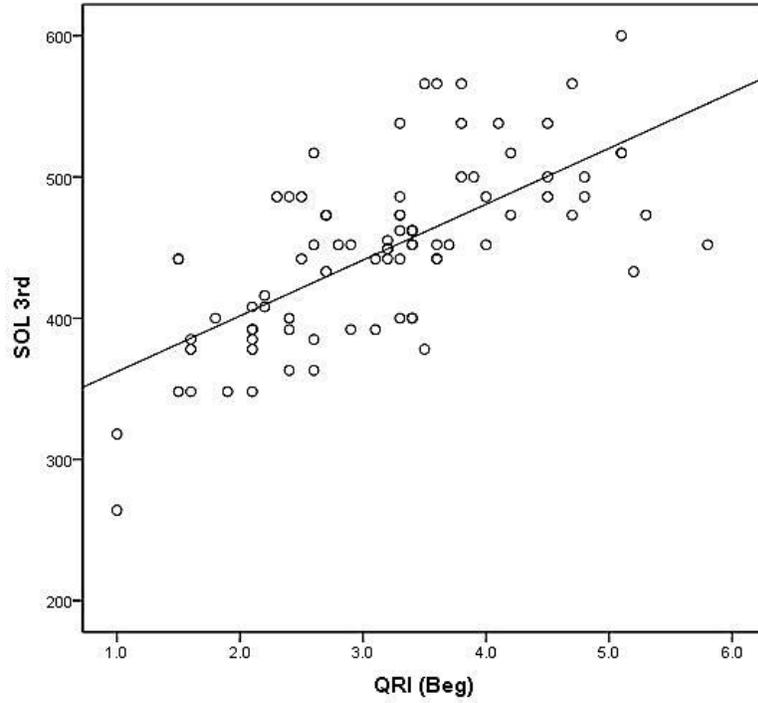


Figure 2. Scatterplot of the linear relationship between the QRI-5 and 3rd grade SOL.

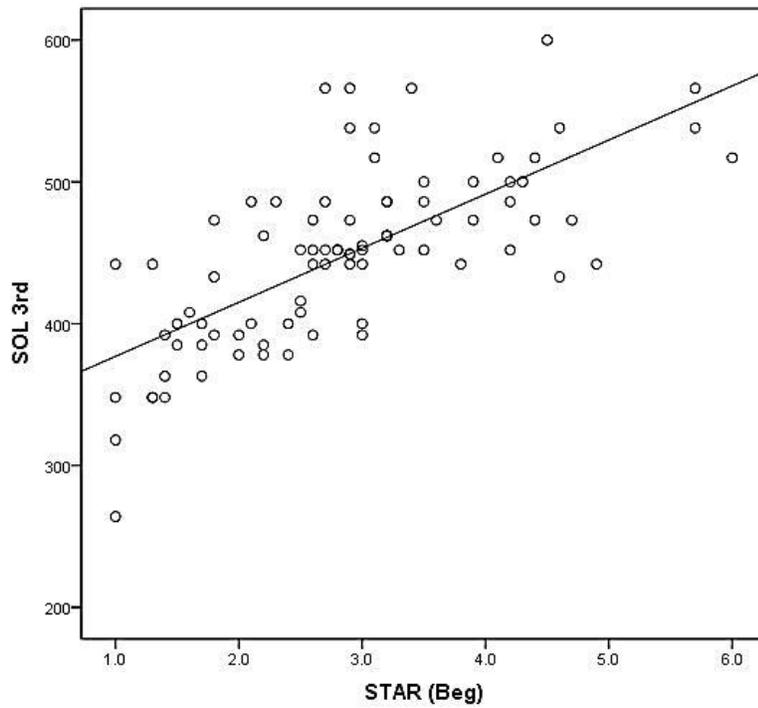


Figure 3. Scatterplot of the linear relationship between the STAR test and 3rd grade SOL.

In order to assess for the possibility of prediction errors, or residuals, a scatterplot of the residuals versus predicted values was constructed. The data was found to be evenly distributed, and the flat straight line in Figure 4 reveals that the variance in error is constant with the varying values found in the predicted variable. Therefore, the assumption of homoscedasticity has been satisfied.

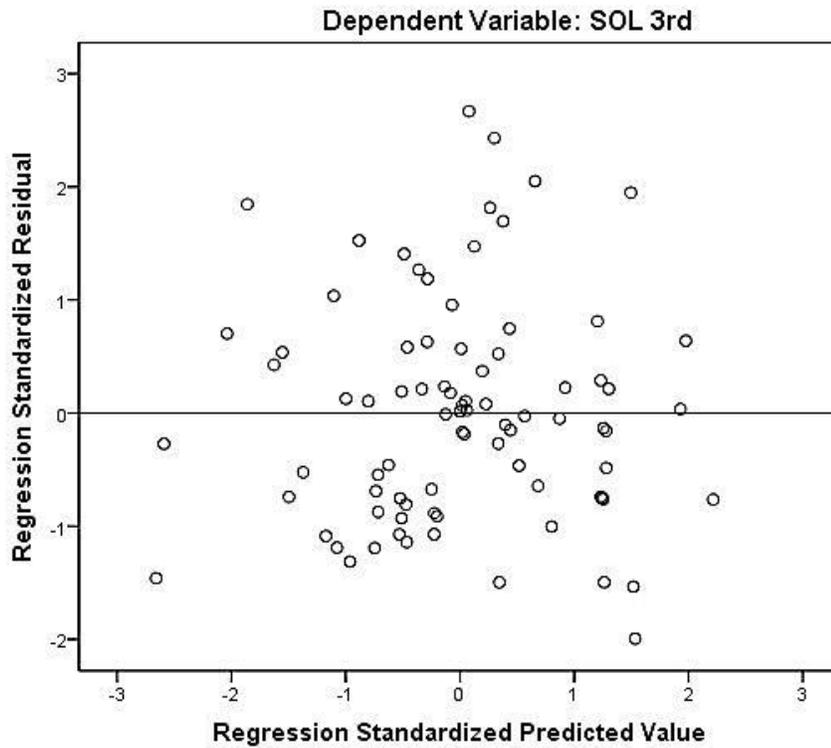


Figure 4. Scatterplot of residuals versus predicted values reveals homoscedastic data.

In an effort to check for the assumption of normality, a histogram of the residuals was generated. An examination of Figure 5 indicates that the data distribution is sufficiently normal. The figure was developed using the dependent variable, third grade SOL scores, together with the model of three independent variables. This histogram reflects the continuity of the collected data, resulting in the typical bell-shaped design characteristic of normally distributed data. Thus, the assumption of normality was also satisfied, and no violations were encountered.

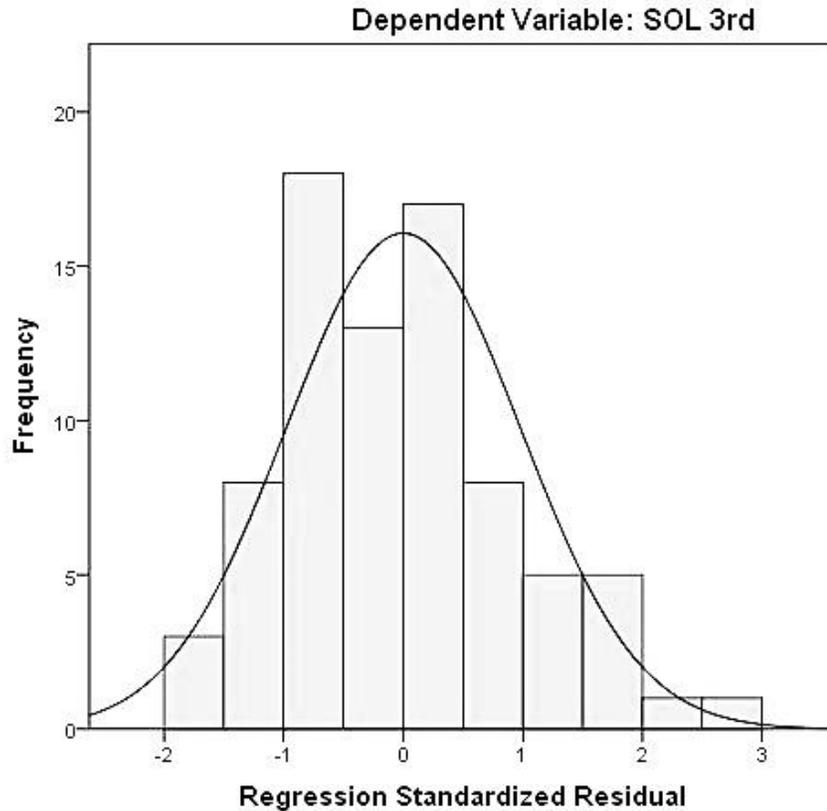


Figure 5. Normal probability plot of residuals.

Another critical step in performing multiple regression analysis is to check for the presence of multicollinearity amongst predictor variables. According to Gall *et al.*, (2010) this phenomenon occurs when there is considerable redundancy, or overlap, between the independent variables utilized in the multiple regression model. When predictor variables are highly correlated with one another, their contribution in predicting the criterion can be diminished. An intercorrelation matrix was constructed to determine the extent to which each of the diagnostic instruments was related to the others. Information found in Table 4 indicates that each of the predictor variables was positively correlated with the others. The strongest relationship was observed between the QRI-5 and the STAR Reading Test with a correlation of .853. However, none of the observed correlations exceeded the critical value of .90 established by Tabachnick

and Fidell (2013). When correlations surpass the .90 threshold, caution should be exercised in the inclusion of highly correlated predictors. In such instances, redundancy can be minimized and the regression equation can be improved through the removal of one or more of the independent variables (Tabachnick & Fidell, 2013).

Table 4
Intercorrelation Matrix

	SOL	STAR	QRI-5	PALS
SOL	1.0	.699	.676	.589
STAR Reading	.699	1.0	.853	.598
QRI-5	.676	.853	1.0	.635
PALS	.589	.598	.635	1.0

Although preliminary analysis revealed that there were no elevated concerns regarding the multicollinearity of variables, a strong relationship was observed between the QRI-5 and the STAR Reading Test. Therefore, further analysis of potential multicollinearity continued with an examination of the variance inflation factors (VIF) and tolerance values. Information found in Table 5 confirms that none of the observed tolerance levels were below the threshold of .10. Likewise, none of the VIF values were found to be greater than 10 (Tabachnick & Fidell, 2013).

Table 5
Collinearity Statistics for Predictor Variables

	VIF	Tolerance
PALS	1.708	.585
STAR Reading	3.752	.267
QRI-5	4.036	.248

Results

Pertinent descriptive statistics are found in Table 6. This includes minimum and maximum scores attained on each reading assessment. In addition, the mean score and standard deviation are also presented.

Table 6
Descriptive Statistics

	Min	Max	<i>M</i>	<i>SD</i>
SOL	264	600	448.2	62.140
PALS	32	76	68.57	9.145
STAR Reading	1.0	5.7	2.868	1.1397
QRI-5	1.0	5.8	3.177	1.0618

The present study was designed to evaluate two research questions and corresponding null hypotheses as follows:

Research Question 1: What is the relationship between diagnostic reading measures obtained at the beginning of the year with achievement scores on the third grade Standards of Learning assessment in reading?

Null Hypothesis 1: There will be no significant correlation between scores on the Phonological Awareness Literacy Screening and scores on the Virginia Standards of Learning reading assessment.

Null Hypothesis 2: There will be no significant correlation between scores on the STAR Reading Test and scores on the Virginia Standards of Learning reading assessment.

Null Hypothesis 3: There will be no significant correlation between scores on the Qualitative Reading Inventory and scores on the Virginia Standards of Learning reading

assessment.

Pearson Correlations were calculated as a means of evaluating the magnitude of the relationship between the third grade Standards of Learning assessment and each of the three diagnostic reading measures. Data included in Table 7 demonstrates that there is a significant positive correlation between each instrument and the SOL test. Of those, the STAR Reading Test was found to have the strongest relationship with student achievement scores on the SOL with a correlation of $r = .699$. A significant relationship was also found between the QRI-5 and the Standards of Learning test, yielding a correlation of $r = .676$. Although the effect size is not quite as strong, the PALS was also shown to have a significant relationship with outcome scores on the SOL test with a value of $r = .589$. All three tests were found to be significant at the $p < .01$ level (two-tailed test), thus minimizing the possibility of committing a Type 1 error. Having found a significant relationship between each predictor and SOL scores, each of the null hypotheses for research question one is rejected.

Table 7
Correlation Coefficients of Reading Measures with SOL

	<i>n</i>	<i>p</i>	<i>r</i>
STAR Reading	84	< 0.01*	.699
QRI-5	84	< 0.01*	.676
PALS	84	< 0.01*	.589

*significant

Research Question 2: What is the relationship between a linear combination of the three predictor variables (the joint model comprised of PALS, STAR and QRI) and achievement scores on the third grade Standards of Learning assessment in reading?

Null Hypothesis 4: There is no significant relationship between a linear combination of

the three predictor variables and achievement scores on the third grade Virginia SOL test in reading.

Student scores on the Standards of Learning test were regressed on the model of PALS, STAR and QRI used together. This analysis reveals that the model significantly predicts achievement on the third grade SOL reading test, $F(3,80) = 31.68$ with $p < .001$. The sample multiple correlation coefficient (R) was .737. This yields an R^2 value of .543, indicating that approximately 54% of the variance in third grade SOL scores is accounted for by the variability in the linear combination of the three diagnostic reading measures considered together. An adjusted $R^2 = .526$, indicates a slight loss in the predictive value of the model if applied to a different sample. As the linear combination of predictor variables was found to have a significant relationship with the criterion, Null Hypothesis 4 is rejected.

In conducting linear regression analysis of the combined model of PALS, STAR and QRI, a multiple regression equation was generated as an indicator of the model's strength in predicting outcome scores on the third grade reading SOL test. Having satisfied the assumptions of linearity, multicollinearity and homoscedasticity, the equation was deemed to be stable with strong predictive value. Unstandardized coefficients are generated for each independent variable, and a constant value is calculated in forming the multiple regression equation. The raw equation for predicting SOL scores utilizing the three diagnostic reading measures is:

$$\text{Predicted SOL score} = 22.017(\text{STAR}) + 10.888(\text{QRI}) + 1.556(\text{PALS}) + 243.796.$$

Null Hypothesis 5: There will be no difference in the amount of SOL reading variance captured by the PALS, STAR and QRI reading measures.

Standard multiple regression analysis was conducted in an effort to determine the predictive contribution of each independent variable to achievement scores on the SOL test. All

of the bivariate correlations between predictor variables and the criterion were found to be positive. Of the three indices, two were found to be statistically significant. Both the STAR Reading Test ($t = 2.758, p = .007$) and the second grade PALS assessment ($t = 2.317, p = .023$) were found to be significant predictors of SOL scores. In contrast, the QRI-5 ($t = 1.225, p = .224$) did not have a significant contribution in predicting outcome on the third grade SOL when used in conjunction with the STAR Reading test and PALS assessment simultaneously.

Table 8
Multiple Regression Coefficients and Significance Values

	B	β	Partial Correlation	t	p
PALS	1.556	.229	.251	2.317	.023*
STAR Reading	22.017	.404	.295	2.758	.007*
QRI-5	10.888	.186	.136	1.225	.224

*significant

Of the two significant predictors, analysis of the partial correlations in Table 8 reveals that the STAR Reading Test (.295) captures the greatest amount of unique variance in SOL scores while controlling for the impact of the other independent variables. Likewise, an analysis of standardized regression coefficients, or Beta weights, allows for a direct comparison of the relative contribution of each predictor variable. The corresponding Beta (or β) weights reveal that the STAR Reading Test is the strongest predictor in the multiple regression equation. The β values allow for a direct comparison of the variables as each is the result of standardized z -scores with a mean of zero and a standard deviation of one. In both cases, an analysis of partial correlations and Beta values indicate that the STAR Reading Test is the strongest predictor of outcome scores on the third grade SOL test. Therefore, Null Hypothesis 5 is rejected.

Null Hypothesis 6: No linear combination of variables predicts achievement scores on the third grade Virginia SOL reading test better than the linear combination of all three predictor variables utilized simultaneously.

In order to further refine the initial regression equation obtained with all three predictor variables, a stepwise regression analysis was performed to determine which linear combination of variables results in an optimal prediction of the achievement scores on the third grade SOL test. Although none of the bivariate correlations among predictor variables exceeded .90, the relationship between the STAR and QRI revealed a strong correlation of .853. In such instances when there is concern for redundancy among predictors, Tabachnick and Fidell (2013) recommend deleting one of the two highly correlated variables. According to Gall *et al.*, (2010), stepwise multiple regression incorporates both stepup and stepdown multiple regression analysis simultaneously in an effort to ascertain the best prediction of the criterion. Predictor variables are added to the multiple regression equation one at a time in order to detect the most significant increase in R . Likewise, predictor variables are deleted from the equation when they fail to contribute at a significant level.

Stepwise analysis revealed that the QRI-5 failed to make a significant contribution to the model when all three predictor variables were used together. In exploring the unique variance accounted for by each of the predictor variables in the previous hypothesis, the p value for the QRI-5 was only .224. As a result, the stepwise analysis removed the QRI from the multiple regression equation. The refined model eliminating the QRI-5 was found to significantly predict achievement scores on the third grade SOL reading test, $F(2,81) = 46.476$ with $p < .001$. The multiple correlation coefficient (R) for the new model was .731, resulting in a slight decrease from the .737 (R) reported for the model of PALS, STAR and QRI used together. The stepwise

regression further revealed an R^2 value of .534. Therefore, more than 53% of the variance in third grade SOL scores can be accounted for by the variability in the linear combination of STAR and PALS scores when used without the QRI. The adjusted $R^2 = .523$, demonstrating a very slight loss in the predictive value of the revised model when applied to a different sample. The raw equation for predicting SOL scores utilizing only PALS and STAR is as follows:

$$\text{Predicted SOL score} = 29.497(\text{STAR}) + 1.801 (\text{PALS}) + 240.149.$$

In comparing the results of the standard multiple regression analysis against the results of the stepwise analysis, a slight loss was observed in the predictive value of the model that only used two diagnostic reading measures (STAR and PALS) as predictors. Although the QRI-5 was not found to be a significant predictor, it did contribute a minute amount of unexplained variance in the criterion as part of the original model that utilized all three independent variables together. Therefore, as the original model captures the greatest amount of variance in outcome scores on the third grade reading SOL test, Null Hypothesis 6 cannot be rejected.

Summary

The purpose of the present study was to utilize common diagnostic reading assessments to predict outcome scores on a third grade test of reading comprehension. This study attempted to answer two research questions. While the first question was designed to evaluate the strength of the relationships between independent variables and outcome scores on the SOL test, the second research question attempted to determine the predictive value of the model as well as the predictive strength of the independent diagnostic reading measures.

An analysis of the data demonstrates that there is a significant relationship between each of the diagnostic instruments and achievement scores on the third grade SOL reading test. Likewise, a significant relationship was found between the linear combination of predictor

variables (the PALS, STAR and QRI assessments) and SOL scores. As a result, a regression equation was generated as a means of predicting achievement on the SOL. Of the three diagnostic reading measures utilized as predictor variables, the STAR Reading Test captured the greatest amount of unique variance in the criterion, and a stepwise regression analysis failed to find a model that was a stronger predictor than using all three diagnostic reading measures simultaneously. Chapter five includes a detailed presentation of the study conclusions together with instructional applications and implications for future research investigations.

CHAPTER FIVE: DISCUSSION

In the year 2001, President George W. Bush introduced landmark federal legislation aimed at closing the achievement gap between various demographic subgroups. Although the general public lauded the plan as a noble effort, very little would change in the decade that followed. While President Barack Obama pledged sweeping changes that would move away from measuring academic achievement through the use of multiple-choice standardized assessments, it is business as usual in public school classrooms throughout the United States. Rather than implementing an overhaul of the failed system, the U.S. Secretary of Education has implemented nothing more than “a few nips and tucks in the program” (Ravitch, 2009, p. 4). As a reauthorization of President Johnson’s Elementary and Secondary Education Act of 1965, No Child Left Behind has thus far failed in its civil rights mission as the great equalizer in American society. Today, the achievement gap persists in areas of reading and mathematics, and sixty years later, our country has yet to realize the promise of *Brown v. Board of Education* (Hewitt, 2011).

Third grade marks a transitional time period for elementary students and reading teachers alike. This developmental period is characterized by an instructional paradigm shift as children negotiate the transition from an emphasis on fluency to comprehension (Adlof *et al.*, 2006; Cartwright, 2006; Dooley, 2010). Moreover, federal mandates stipulate that states will evaluate reading achievement each year in grades three through eight (United States Department of Education, 2002). As teachers and administrators search for a magic formula to predict student success on these assessments for the first time, they are often forced to rely on subjective informal observations due to a lack of appropriate baseline data (Hinnant *et al.*, 2009; Keene, 2009). In this atmosphere of high stakes testing, however, teachers are expected to employ

research-based instruction and assessment (Invernizzi *et al.*, 2005; U.S. Department of Education, 2002). Therefore, the strategic use of diagnostic reading instruments which evaluate the known prerequisite skills of comprehension should enable teachers to effectively identify students who require remedial assistance prior to end-of-course test administration.

The present study is grounded within the conceptual framework of the Simple View of Reading (Gough & Tunmer, 1986) and the Lexical Quality Hypothesis (Perfetti & Hart, 2002). The Simple View of Reading holds that reading comprehension is the product of both decoding and linguistic comprehension. Similarly, the Lexical Quality Hypothesis emphasizes mental representations that result from interrelated orthographic, phonological and semantic word qualities. This investigation makes use of three diagnostic reading measures that are known to evaluate the sub-skills of comprehension identified in a thorough review of the research literature. These instruments serve as predictor variables of achievement scores on the third grade Standards of Learning reading test, the criterion. In the absence of useful baseline comprehension data, these instruments are expected to have considerable predictive utility.

Chapter five provides a summary of the present investigation and a discussion of the findings within the context of corresponding research questions. Following this overview of the study, a critical analysis of the research data will be utilized in addressing study limitations. Practical implications are also presented. Finally, this chapter will conclude with this researcher's recommendations for future research.

Summary of Findings

This research investigation incorporated a multiple regression design to evaluate the relationship between diagnostic reading measures and outcome scores on the third grade Virginia Standards of Learning assessment. Moving beyond simple relationships, this design was also

used to evaluate the utility of the collective model of variables in predicting achievement scores. Having determined the predictive value of the model, further analysis revealed the amount of unique variance in the dependent variable attributable to each of the predictor variables independently. Finally, a stepwise analysis was utilized in an effort to further refine the regression equation for practical applications.

This study utilized an archived data set from a heterogeneous sample of $n = 84$ third grade students at a large, semi-rural elementary school in Virginia. Three diagnostic reading measures were administered to each student between May and September of 2012. These include the second grade end-of-year PALS test, the STAR Reading Test and the QRI-5. Administered in May of 2013, the third grade reading SOL served as the criterion, or outcome measure. After all of the corresponding data had been collected for each student in one master spreadsheet, the school's testing coordinator stripped the data set of all identifying information. With all student identifiers redacted for the purpose of anonymity, the data set was supplied to the principal investigator, and statistical analysis commenced using SPSS version 19.0.

Multiple regression analysis revealed a significant relationship at the $p < .01$ level between each of the diagnostic reading measures and the third grade reading SOL test. With a correlation of $r = .699$, the STAR test exhibited the strongest relationship with the SOL. The QRI-5 and PALS assessments were also found to have a strong relationship with the SOL, yielding correlations of $r = .676$ and $r = .589$ respectively.

The linear combination of diagnostic instruments was also found to have a significant relationship with achievement scores on the SOL test. When utilized together, the model was found to significantly predict SOL outcome at $p < .01$. With a multiple correlation coefficient of $R = .737$, the collective model is said to account for more than 54% of the variance in third grade

SOL scores. As a predictor of SOL outcome, the following multiple regression equation was also generated: *Predicted SOL score* = 22.017(STAR) + 10.888(QRI) + 1.556 (PALS) + 243.796.

Of the three diagnostic reading measures used in this study, only the STAR and PALS tests were found to be significant SOL predictors when all three tests were used simultaneously. With a partial correlation of .295 and $\beta = .404$, the STAR test was significant at $p = .007$. In contrast, the PALS test was significant at $p = .023$, with a partial correlation of .251 and $\beta = .229$.

In an effort to further refine the initial multiple regression equation, a stepwise analysis was performed. As a means of eliminating variables that do not have a significant contribution to the initial equation, the QRI-5 was removed from the model of predictors. Although the new model of using only the STAR and PALS tests as predictors was found to be significant at $p < .01$, a slight loss was observed in the predictive value of the model with $R = .731$. Likewise, the refined model of predictors was found to account for more than 53% of the variance in SOL scores as opposed to the 54% accounted for by the original model incorporating the QRI.

Discussion of Findings

The results of this correlational research investigation provide insight into the prediction of achievement scores on standardized assessments of reading comprehension. Although third grade teachers generally lack an accurate frame of reference, incorporating diagnostic measures that evaluate the prerequisite skills of comprehension makes perfect sense. The Simple View of Reading (Gough & Tunmer, 1986) and the Lexical Quality Hypothesis (Perfetti & Hart, 2002) each emphasize the importance of decoding and vocabulary. Therefore, diagnostic instruments assessing those skills would be expected to provide teachers with an accurate means of gauging success and failure on standardized reading tests such as the Virginia SOL. The insights

revealed in this section, therefore, demonstrate the instructional value of using beginning-of-year data to identify students in need of remedial assistance. The conclusions for this study are presented in relevant themes based on the corresponding research questions.

The relationship between diagnostic instruments and the third grade SOL. The data obtained in an analysis of research question one reveal strong relationships between each of the instruments and the SOL test. Although Renaissance Learning (2011) has demonstrated strong correlations between the STAR Reading Test and standardized reading assessments from 29 different states, Virginia was not among those that were listed. With a strong correlation of $r = .699$, the significant relationship observed in the present study supports the utility of the STAR test in predicting outcome scores on the third grade SOL. Serving as a measure of vocabulary, high scores on the STAR test generally correspond to high scores on the SOL. Although these findings are supported by the literature review (Adair, 2010; Churchwell, 2009), each study incorporated a significantly older sample of participants. Given its effectiveness and ease of implementation, the STAR Reading test would seem to be an ideal diagnostic instrument for teachers to administer at the beginning of the school year.

The strong relationship between the QRI-5 and the third grade reading SOL was also found to be significant. This is not surprising as test developers offer the QRI as a measure of word recognition (or linguistic comprehension), one of the central elements of the Simple View of Reading (Gough & Tunmer, 1986; Leslie & Caldwell, 2011). Although the test is administered one-on-one with the student, the observed correlation of $r = .676$, implicates the instrument as a satisfactory predictor of SOL achievement. Higher scores on the QRI-5 are typically associated with stronger scores on the SOL test. Although it has been argued that the inherent design of the QRI could yield an inflated measure of the student's ability to comprehend

(Epelbaum, 2007; Wolpert & Vacca-Rizopoulos, 2012), that is not supported in the present study. Rather, it seems as though the inclusion of narrative and expository passages are helpful in gauging student ability to comprehend both fiction and nonfictional excerpts found on the SOL test.

Of the three diagnostic instruments administered to students in the present study, only the PALS test was required. Administered at the conclusion of the second grade year, the PALS test is taken by nearly every second grade student in the state of Virginia (“PALS,” 2013). In assessing phonological awareness, the PALS test served as a measure of decoding in this investigation. Again, the correlation of $r = .589$ suggests that the second grade PALS test serves as a significant predictor of comprehension on the third grade SOL. According to Invernizzi *et al.*, (2013), PALS validity was evaluated in its ability to predict scores on the third grade reading SOL. With a sample of 277 third grade students in Virginia, a strong correlation of $r = .60$ was observed in 2001.

Although reported reliability and validity data was significant for all three diagnostic instruments, the STAR test was found to have the strongest relationship with third grade SOL scores. This is not surprising given the subjective testing protocols characteristic of the QRI and PALS tests. While the QRI and PALS appear to measure different underlying constructs, each is dependent upon time consuming assessment procedures between the testing proctor and the student. In contrast, the STAR Reading test is an efficient and object computer based instrument which alleviates concerns regarding the impact of observer effects such as bias, rating error, observer drift and reliability decay (Gall *et al.*, 2010). Therefore, concerns for test-retest reliability and inter-rater reliability are minimal. These findings indicate that while classroom teachers generally appreciate the value of face-to-face assessment procedures, computer

generated tests often yield more accurate results with incredible efficiency.

The relationship between the model of instruments and the third grade SOL. The findings of the present study reveal that a significant relationship exists between the linear combination of predictor variables (PALS, STAR and QRI) and student scores on the third grade reading SOL test. The multiple correlation coefficient of $R = .737$ yields an effect size of $R^2 = .543$. Therefore, more than 54% of the variability in third grade comprehension scores can be explained by variability in the model of predictors considered together.

These results suggest that a combination of diagnostic instruments assessing elements of the Simple View of Reading and the Lexical Quality Hypothesis are useful in predicting scores on comprehension assessments as early as third grade. Although this investigation utilized different instrumentation and a larger student sample, results are very similar to findings reported by Joshi and Aaron (2000) who noted that the Simple View of Reading accounts for roughly 48% of the variability in comprehension. Within the context of more recent literature on the topic, Georgiou *et al.*, (2009) found that the product of decoding and listening comprehension explains 45-47% of the variance in comprehension with third and fourth grade students. Likewise, Adlof *et al.*, (2006) consolidated findings from several investigations, reporting that the Simple View accounts for no less than 45% of comprehension variance. Nevertheless, the findings of this study fall far short of the results reported by Hoover and Gough (1990) who used linear regression analysis to generate a third grade estimate as high as 83%.

Given that these findings are in keeping with established empirical research, it would seem that the present combination of diagnostic instruments adequately predicts third grade comprehension scores. Therefore, with a significance level of $p < .001$, the raw equation for predicting SOL scores shows tremendous practical utility in identifying students who are at risk

of failure on the third grade SOL. In selecting candidates for remediation early in the school year, third grade teachers will be able to provide adequate assistance long before high stakes assessments are administered at the end of course.

SOL variance captured by the STAR Reading assessment. Although all three diagnostic instruments were found to have a significant relationship with the SOL test when used independently, regression analysis reveals that the STAR Reading test captured the greatest amount of variance in the criterion when the model is used to predict comprehension. This is not surprising as the STAR test also had the strongest correlation with the SOL with $r = .699$. Although the PALS test was also found to be a significant contributor at $p = .023$, the STAR assessment was significant at the more stringent level of $p = .007$. In contrast, when all three predictors were used together, the QRI was no longer found to have a significant contribution to outcome scores on the SOL. The inability of the QRI to capture a significant portion of the variance in the criterion is likely due to the overlap, or redundancy, between the STAR test and the QRI. Although the instruments seem to be incredibly different, it appears that the STAR test is able to effectively and efficiently capture all but a small portion of the same variance accounted for by the QRI-5.

In comparing the STAR Reading test and the PALS test, it is clear that each captures a fair amount of the unique variance in the outcome variable when utilized together. Nevertheless, as a measure of vocabulary, the STAR test clearly captures a more significant part of the variance in comprehension than PALS. Within the context of the Simple View of Reading, the measure of vocabulary generated a partial correlation of .295 while the measure of decoding (PALS) resulted in a partial correlation of .251.

The present investigation supports the work of Cartwright (2006) who observed that

students begin to “unglue” from print sometime between grade two and grade four. Although Gough and Tunmer (1986) offered decoding and linguistic comprehension as being roughly equal contributors in the Simple View of Reading, Hoover and Gough (1990) and Gough *et al.*, (1996) observed that student reliance on decoding begins a steady decline following the primary years in elementary school. In evaluating the fluctuating contribution of each subcomponent of the Simple View, Oullette and Beers (2010) found that decoding was the stronger predictor in grade one whereas vocabulary was the dominant skill for students in grade six. Similar results were found by Tilstra *et al.*, (2009) as well as Verhoeven and Van Leeuwe (2008). Their research findings demonstrate that as reliance on decoding begins to subside, vocabulary gradually becomes a stronger predictor of reading comprehension. More specifically, this transition is said to occur sometime between second grade and fourth grade.

The findings of this investigation also demonstrate that the STAR Reading test captures the greatest amount of variance in the third grade reading SOL. Although decoding and the underlying skill, phonological awareness, continue to account for unique variance in comprehension, vocabulary is the stronger measure by the time students enter the third grade. While teachers in Virginia are required to administer the PALS test at the conclusion of the second grade year, its utility is overshadowed by the STAR Reading test in its ability to predict third grade comprehension scores. However, as each assessment captures unique variance in outcome scores, it seems ideal to use both instruments together as a means of compensating for the lack of accurate baseline data for third grade teachers at the beginning of the school year.

The relationship between the refined model and the third grade SOL. A stepwise regression analysis was performed in an effort to identify a more efficient model for predicting third grade SOL scores. Not surprisingly, the QRI-5 was eliminated from the model as it failed

to account for a significant amount of the unique SOL variance in the presence of the other instruments. Curiously, the decrease in the power of the refined multiple regression equation is minimal. Whereas the model of three predictor variables was able to account for approximately 54% of SOL variance with an R value of .737, the refined model of using only the PALS and STAR captures more than 53% of the same variability with $R = .731$. A loss of less than 1%, therefore, seems entirely justified given the amount of lost instructional time consumed with the administration of the QRI-5. Although, the original model is slightly stronger, the amount of new information accounted for by the QRI is negligible. When practitioners have access to second grade PALS scores and STAR Reading data, there appears to be no justification for also administering the QRI. Therefore, the model of PALS and STAR accurately predicts comprehension scores as expected within the theoretical framework of the Simple View of Reading.

The relationship between the present study and previous research findings. The results of this study lend support to the Simple View of Reading as envisioned by Gough and Tunmer (1986). In utilizing a measure of phonological awareness to evaluate decoding ability and a vocabulary diagnostic measure in gauging linguistic comprehension, results were similar to research findings reported by Adlof *et al.*, (2006), Georgiou *et al.*, (2009) and Joshi and Aaron (2000) who found that the Simple View of Reading accounts for nearly 50% of the variance in comprehension.

In isolating the independent contributions of the individual predictor variables, this study also supports the findings of Hoover and Gough (1990) and Gough *et al.*, (1996) who proposed a revised model of the Simple View with a steadily declining contribution from decoding. Moreover, in providing evidence for the value of vocabulary measures in predicting

comprehension scores in third grade, this investigation supports the research findings of Cartwright (2006), Oullette and Beers (2010), Tilstra *et al.*, (2009) and Verhoeven and Van Leeuwe (2008).

Finally, in incorporating a vocabulary measure (STAR) in addition to the linguistic comprehension measure (QRI), this study supports the recent findings of Protopapas *et al.*, (2012) who described the Simple View of Reading as including both print-dependent and print-independent skills. As stepwise regression analysis revealed that the QRI failed to make a significant contribution in the presence of the vocabulary measure (STAR), it was removed from the model predicting comprehension. Clearly, the QRI was eliminated as a result of its correlation with the STAR assessment. Therefore, as a more robust predictor of print-independent reading skills, the STAR test serves as a more reliable measure in its ability to capture additional variance in reading comprehension.

Limitations of the Study

In evaluating the generalizability of the present study, several limitations must be addressed. To begin, the size and composition of the sample is a primary concern. Although the sample size satisfied the requirements outlined by Gall *et al.*, (2010) and Tabachnick and Fidell (2007), a larger sample would have allowed for greater statistical power in rejecting the null hypotheses. This investigation also made use of secondary archival data obtained through convenience sampling. While this eliminates potential concerns commonly associated with participant awareness and the impact of the Hawthorne Effect, it also makes the generalizability of the results suspect as the outcome could be due to preexisting differences between the sample and the target population. In this case, the sample of students included all third grade students in only one elementary school. This was due to the unique battery of diagnostic instruments that

were administered to the entire group. While other schools occasionally incorporated all three measures, it was done sporadically on an as needed basis. Thus the decision was made to limit the sample size to one school. Moreover, it is also thought that the atypical heterogeneous composition of this specific sample will help facilitate the generalizability of the results to the target population.

Although the present study incorporates three diagnostic instruments that are commonly used throughout the United States, the dependent variable is unique to the state of Virginia. While this represents a potential threat to the external validity of the present study, the results should be generalizable to the extent to which an assessment of similar composition is utilized as the outcome variable. To some degree, this concern is alleviated by policies established by the National Assessment Governing Board (2013), providing considerable continuity from one state to another in terms of the skills and strategies that are being evaluated on standardized reading tests throughout the country.

One threat to the internal validity of the present study is the potential impact of student effort. This extraneous variable is recognized as a potential confound. Clearly, students will not always exert an equal amount of effort on every assessment, and it is impossible to hold this variable constant. Although a significant relationship was found between the model of diagnostic instruments used as predictor variables and student scores on the SOL test, it is likely that an even more accurate regression equation could be generated if this extraneous variable could be controlled.

Implications

The statistical results of the present investigation lend support to the Simple View of Reading. Specifically, these findings expand upon the work of Gough *et al.*, (1996) who

observed that the impact of decoding begins a gradual decline as student reliance on linguistic comprehension strengthens. Within a developmental context, this study demonstrates that increased reliance on word knowledge and vocabulary manifests itself as early as the beginning of the third grade year. As this study compared the strength of linguistic comprehension (QRI-5) against a diagnostic measure of vocabulary (STAR) in predicting student comprehension, the performance of the latter measure supports the research of Protopapas *et al.*, (2012). The professional implications of this study, therefore, demonstrate the importance of cultivating a strong repertoire of vocabulary in the early elementary grades. Given the noteworthy struggles of English Language Learners and students from lower socioeconomic backgrounds, an instructional emphasis on word study and vocabulary development is likely to yield improved scores on tests of reading comprehension.

The early identification of struggling readers is paramount as third grade teachers prepare students to take norm referenced standardized reading assessments for the first time. This study compares the utility of three diagnostic instruments in gauging student ability to read for meaning. According to Sloat *et al.*, (2007), students who struggle to comprehend in third grade frequently experience a host of social and academic challenges. For instance, students with inadequate reading abilities typically experience deficits in curricular understanding and emotional and behavioral difficulties which are frequently manifested as occupational and economic challenges in adulthood. Effective literacy screening can help to forestall these greater sociocultural issues.

Although the QRI-5 appears to be an adequate measure of comprehension, it no longer makes a significant contribution to the variance in comprehension when used in conjunction with the STAR Reading test. Given the ease of implementation and minimal loss of instructional

time, the STAR Reading test is an ideal diagnostic instrument. The efficiency of the STAR test helps to alleviate the need for time consuming diagnostic instruments that appear to be measuring the same underlying construct. The reduction in wasted time then provides valuable opportunities for remedial assistance

While many teachers prefer diagnostic instruments that are administered one-on-one, school leaders must consider their cost effectiveness. This is especially critical in a weak economy when there is often a lack of funding for public schools. In order to administer one-on-one reading assessments to a group of 20 students, classroom teachers would lose approximately 13 hours, or nearly 3 days, of instructional time. On a five teacher team, this is a total loss of approximately 15 instructional days. The net loss, therefore, is measured not only in the amount of time the classroom teacher is out of the classroom, but also in the financial expense of paying substitutes.

When utilized together, the PALS test and the STAR Reading Test can help third grade teachers compensate for the lack of available comprehension data for their students. Grounded in the Simple View of Reading and the Lexical Quality Hypothesis, this combination of diagnostic instruments incorporates a measure of print-dependent and print-independent skills. This has obvious practical utility. Clearly, students who struggle on diagnostic assessments measuring phonological awareness, such as PALS, require additional help with decoding and fluency. Failure to provide those students with appropriate remedial assistance would likely place those students at risk of failure on third grade comprehension assessments. As a means of identifying students most in need of intensive assistance, the multiple regression equation would be useful in predicting comprehension achievement scores months in advance. Nevertheless, in capturing a more significant amount of the unique variance in comprehension, special attention

should be given to the student's STAR score as an objective means of forecasting SOL achievement.

The results of this study have tremendous practical implications for school administrators. In an age of increased accountability, building principals are expected to be instructional leaders with knowledge of best practices. In this way, school leaders have the ability to bridge the divide between the world of academia and the practical setting of the elementary classroom. Likewise, it is incumbent upon school leaders to be good stewards of public funds in implementing research-based pedagogy. The results of this dissertation can help building principals achieve both goals simultaneously.

This investigation illustrates the real world applicability of the Simple View of Reading. In recommending the strategic use of the PALS assessment and the STAR Reading test as measures of decoding and vocabulary, school leaders promote practices which allow for the accurate identification of at-risk students who struggle to demonstrate adequate comprehension in reading. In providing for the training and implementation of explicit vocabulary instruction, school administrators will do much to reduce the achievement gap, all while providing more opportunities for authentic and holistic educational experiences for the children they serve.

Recommendations for Future Research

Whereas this investigation was able to address each of the research questions and corresponding hypotheses, many questions remain unanswered. A synthesis of these findings has thus generated the following recommendations for best practice in the instruction and assessment of reading comprehension. In order to maximize generalizability of the research findings, it is recommended that future efforts attempt to replicate this study with a sample of third grade students in other states. Although many standardized assessments throughout the

country are similar in composition, it would be useful to know if similar results would be found when using the same battery of diagnostic measures in predicting outcome scores on a different criterion.

This study expands upon the body of research supporting the Simple View of Reading as a useful construct in understanding comprehension. Although other studies have shown that reliance on decoding begins to subside as vocabulary becomes the stronger predictor, this study demonstrates that that shift occurs as early as the beginning of third grade. Future research should continue to explore this shift with younger students as they may benefit from a structured, comprehension-based reading instruction prior to third grade. This gradual transition would represent a contrast to the abrupt instructional shift that students typically experience as they move from second to third grade.

As this study implicates the significant role of vocabulary in reading comprehension, it is recommended that future studies investigate the effects of a structured and explicit form of vocabulary instruction. Specifically, it is recommended that the regression equation of PALS and STAR be used for the purpose of identifying students at risk. In providing an intervention designed to enrich oral and receptive vocabulary, future research will help determine if the early identification and remediation of at-risk students makes a significant difference in reading comprehension. As vocabulary deficits are implicated as an obstruction to reading for meaning, the net result of such intensive enrichment should be reflected in higher achievement scores on standardized reading assessments such as the Virginia Standards of Learning.

Finally, an emerging body of research has demonstrated the effectiveness of the Simple View of Reading when used with English Language Learners (Gottardo & Mueller, 2009; Leider *et al.*, 2013; Proctor *et al.*, 2005). However, it is recommended that future research compare how

the model accounts for comprehension in differing demographic subgroups. Although the sample of participants in the present study was comprised of a significant number of ELL students, the number was not sufficient in allowing for an adequate comparison study. As the number of English Language Learners continues to grow in the United States, it would be useful to know if the model predicts comprehension differently in ELL students and their peers who speak English as their first language.

References

- Adair, A. (2010). *A validation study of the grade six reading Texas assessment of knowledge and skills using the test of reading comprehension-fourth edition and the STAR reading test as criterion measures*. Available from ProQuest Dissertations and Theses database. (UMI No. 3438285).
- Adams, M. J., Foorman, B. R., Lundberg, I., & Beeler, T. (1998). *Phonemic awareness in young children: A classroom curriculum*. Baltimore, MD: Paul H Brookes Publishing Co.
- Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities, 43*(4), 332-345.
- Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing, 19*(9), 933-958.
- Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher, 61*(5), 364-373.
- Alvarez, L., & Corn, J. (2008). Exchanging assessment for accountability: The implications of high-stakes reading assessments for English learners. *Language Arts, 85*(5), 354-365.
- Andrews, S., & Bond, R. (2009). Lexical expertise and reading skill: Bottom-up and top-down processing of lexical ambiguity. *Reading and Writing, 22*(6), 687-711.
- Bailey, A. L., & Drummond, K. V. (2006). Who is at risk and why? Teachers' reasons for concern and their understanding and assessment of early literacy. *Educational Assessment, 11*(3), 149-178.
- Barth, A. E., Catts, H. W., & Anthony, J. L. (2009). The component skills underlying reading fluency in adolescent readers: A latent variable analysis. *Reading and Writing, 22*(5), 567-590.

- Bear, D. R., Invernizzi, M., Templeton, S., & Johnston, F. (2008). *Words their way: Word study for phonics, vocabulary, and spelling instruction*, 4th edition. Upper Saddle River, NJ: Prentice Hall.
- Blackwell-Bullock, R., Invernizzi, M., Drake, E. A., & Howell, J. L. (2008-2009). Concept of word in text: An integral literacy skill. *Journal of the Virginia State Reading Association*, 31, 30-39.
- Booker, K. C., Invernizzi, M. A., & McCormick, M. (2007). "Kiss your brain": A closer look at flourishing literacy gains in impoverished elementary schools. *Reading Research and Instruction*, 46(4), 315-340.
- Boucher, D. (2005). Can a ten-minute reading assignment deliver on its claims? *California English*, 10(4), 22-24.
- Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities*, 40(3), 226-243.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96(1), 31-41.
- Carmichael, S. B., Martino, G., Porter-Magee, K., & Wilson, W. S. (2010). *The state of state standards - and the common core - in 2010*. Washington, DC. Retrieved from Thomas B. Fordham Institute website:
http://www.edexcellence.net/sites/default/files/publication/pdfs/SOSSandCC2010_FullReportFINAL_6.pdf
- Cartwright, K. B. (2006). Fostering flexibility and comprehension in elementary students. *The Reading Teacher*, 59(7), 628-634.

- Casbergue, R. M. (2011). Assessment and instruction in early childhood education: Early literacy as a microcosm of shifting perspectives. *Journal of Education, 190*(1), 13-20.
- Churchwell, D. E. (2009). *The impact of reading achievement on overall academic achievement*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3389306).
- Clark, R., Morrison, T. G., & Wilcox, B. (2009). Readers' theater: A process of developing fourth-graders' reading fluency. *Reading Psychology, 30*(4), 359-385.
- DeVries, B. A. (2012). Vocabulary assessment as predictor of literacy skills. *New England Reading Association Journal, 47*(2), 4-9.
- Diehl, H. L., Armitage, C. J., Nettles, D. H., & Peterson, C. (2011). The three-phase reading comprehension intervention (3-RCE): A support for intermediate-grade word callers. *Reading Horizons, 51*(2), 149-170.
- Dooley, C. M. (2010). Young children's approaches to books: The emergence of comprehension. *The Reading Teacher, 64*(2), 120-130.
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*(2), 167-188.
- Epelbaum, D. (2007). Multiple intelligence assessments give insight into reading comprehension difficulties and potential: A case study. *The International Journal of Learning, 14*(5), 243-251.
- Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, V. J., Hamlett, C. L., & Lambert, L. (2012). First grade cognitive abilities as long-term predictors of reading comprehension and disability status. *Journal of Learning Disabilities, 45*(3), 217-231.
- Gaither, A. C. (2008). *A comparison of two early education literacy benchmark assessments*.

- Available from ProQuest Dissertations and Theses database. (UMI No. 3308690).
- Gall, M. D., Gall, J. P., & Borg, W. R. (2010). *Educational research: An introduction* (8th ed.). New York: Allyn & Bacon.
- Ganske, K. (2000). *Word journeys: Assessment-guided phonics, spelling, and vocabulary instruction*. New York, NY: Guilford Press.
- Gattis, M. N., Morrow-Howell, N., McCrary, S., Lee, M., Jonson-Reid, M., McCoy, H., Tamar, K., Molina, A., & Invernizzi, M. (2010). Examining the effects of New York experience corps program on young readers. *Literacy Research and Instruction, 49*, 299-314.
- Georgiou, G. K., Das, J. P., & Hayward, D. (2009). Revisiting the “simple view of reading” in a group of children with poor reading comprehension. *Journal of Learning Disabilities, 42*(1), 76-84.
- Gottardo, A., & Mueller, J. (2009). Are first- and second-language factors related in predicting second-language reading comprehension? A study of Spanish-speaking children acquiring English as a second language from first to second grade. *Journal of Educational Psychology, 101*(2), 330-344.
- Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In Cornoldi, C. & Oakhill, J. (Eds.), *Reading comprehension difficulties: Processes and interventions* (pp. 1-13). Hillsdale, NJ: Erlbaum.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6-10.
- Helman, L. A. (2005). Using literacy assessment results to improve teaching for English-language learners. *The Reading Teacher, 58*(7), 668-667.
- Hewitt, D. (2011). Reauthorize, revise and remember: Refocusing the No Child Left Behind Act to

- fulfill Brown's promise. *Yale Law & Policy Review*, 30(1), 169-194.
- Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, 101(3), 662-670.
- Hoiem-Tengesdal, I. (2010). Is the simple view of reading too simple? *Scandinavian Journal of Educational Research*, 54(5), 451-469.
- Hoover, W. A. & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127-160.
- Howell, D. C. (2011). *Fundamental statistics for the behavioral sciences* (7th ed.). Belmont, CA: Thomson/Wadsworth.
- Howell, J. L., Partridge, H.A., Landrum, T.J., & Invernizzi, M.A. (2004). A look back on seven years of PALS and Virginia's EIRI. *Journal of the Virginia State Reading Association*, 27, 74-77.
- Hulslander, J., Olson, R. K., Willcutt, E. G., & Wadsworth, S. J. (2010). Longitudinal stability of reading-related skills and their prediction of reading development. *Scientific Studies of Reading*, 14(2), 111-136.
- Invernizzi, M. A., Landrum, T. J., Howell, J. L., & Warley, H. P. (2005). Toward the peaceful coexistence of test developers, policymakers, and teachers in an era of accountability. *The Reading Teacher*, 58(7), 610-618.
- Invernizzi, M., Meier, J., & Juel, C. (2013). *PALS 1-3 technical reference form A*. Charlottesville, Virginia: The University of Virginia.
- Iwata, T., Kojiri, T., Yamada, T., & Watanabe, T. (2011). Recommendation for English multiple-choice cloze questions based on expected test scores. *International Journal of Knowledge-*

- based and Intelligent Engineering Systems, 15(1), 15-24.*
- Johnson, G. M. (2004). Constructivist remediation: Correction in context. *International Journal of Special Education, 19(1), 72-88.*
- Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21(2), 85-97.*
- Joshi, R. M., Binks, E., Graham, L., Ocker-Dean, E., Smith, D. L., & Boulware-Gooden, R. (2009a). Do textbooks used in university reading education courses conform to the instructional recommendations of the national reading panel? *Journal of Learning Disabilities, 42(5), 458-463.*
- Joshi, R. M., Binks, E., Hougen, M., Dahlgren, M. E., Ocker-Dean, E., & Smith, D. L. (2009b). Why elementary teachers might be inadequately prepared to teach reading. *Journal of Learning Disabilities, 42(5), 392-402.*
- Katzir, T., Kim, Y., Wolf, M., O'Brien, B., Kennedy, B., Lovett, M., & Morris, R. (2006). Reading fluency: The whole is more than the parts. *Annals of Dyslexia, 56(1), 51-82.*
- Keene, E. (2009). The small changes that make a big difference. In K. Ganske & D. Fisher (Eds.), *Comprehension across the curriculum: Perspectives and practices K-12* (pp. 7-22). New York: Guilford Press.
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12(3), 281-300.*
- Kesler, T. (2010). Shared reading to build vocabulary and comprehension. *The Reading Teacher, 64(4), 272-277.*
- Kim, Y., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading

- fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology*, 102(3), 652-667.
- Kim, Y. S., Wagner, R. K., & Foster, E. (2011). Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first-grade readers. *Scientific Studies of Reading*, 15(4), 338-362.
- Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology*, 95(3), 453-464.
- Kirby, J. R., & Savage, R. S. (2008). Can the simple view deal with the complexities of reading? *Literacy*, 42(3), 75-82.
- Knuth, S. B. (2011). *Predicting North Carolina third grade end-of-grade test of reading comprehension scores from first, second, and third grade variables* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3477474)
- Kucan, L. (2012). What is most important to know about vocabulary? *The Reading Teacher*, 65(6), 360-366.
- Leider, C. M., Proctor, C. P., Silverman, R. D., & Haring, J. R. (2013). Examining the role of vocabulary depth, cross-linguistic transfer, and types of reading measures on the reading comprehension of Latino bilinguals in elementary school. *Reading and Writing*, 26(9), 1459-1485.
- Lervag, A., & Aukrust, V. G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *The Journal of Child Psychology and Psychiatry*, 51(5), 612-620.
- Leslie, L., & Caldwell, J. S. (2011). *Qualitative reading inventory – 5*. Boston, MA: Pearson.
- Luttenegger, K. C. (2009). Research in progress: Formative assessment practices in reading

- instruction in pre-service teachers' elementary school classrooms. *Journal of Education for Teaching*, 35(3), 299-301.
- May, L. A. (2011). Situating strategies: An examination of comprehension strategy instruction in one upper elementary classroom oriented toward culturally relevant teaching. *Literacy Research and Instruction*, 50(1), 31-43.
- McKenna, M. C., & Walpole, S. (2008). *The literacy coaching challenge: Models and methods for grades K-8*. New York: The Guilford Press.
- Mokhtari, K., Hutchison, A. C., & Edwards, P. A. (2010). Organizing instruction for struggling readers in tutorial settings. *The Reading Teacher*, 64(4), 287-290.
- Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, 27(4), 342-356.
- National Assessment Governing Board. (2013). *Reading framework for the 2013 national assessment of educational progress*. Washington, DC: Author.
- National Center for Education Statistics (2011). *The Nation's Report Card: Reading*. Retrieved September 14, 2013 from <http://nationsreportcard.gov/>
- National Center on Response to Intervention (2013) *Screening Tools Chart*. Retrieved October 6, 2013 from <http://www.rti4success.org/screeningTools>
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Report of the subgroups*. Bethesda, MD. Author.
- Ness, M. (2011). Explicit reading comprehension instruction in elementary classrooms: Teacher use of reading comprehension strategies. *Journal of Research in Childhood Education*, 25(1), 98-117.

- Nilsson, N. L. (2008). A critical analysis of eight informal reading inventories. *The Reading Teacher, 61*(7), 526-536.
- Nithart, C., Demont, E., Metz-Lutz, M. N., Majerus, S., Poncelet, M., & Leybaert, J. (2011). Early contribution of phonological awareness and later influence of phonological memory throughout reading acquisition. *Journal of Research in Reading, 34*(3), 346-363.
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554-566.
- Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing, 23*(2), 189-208.
- PALS (2013). Retrieved June 2, 2013, from <http://pals.virginia.edu/>
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213), Amsterdam, Netherlands/Philadelphia, PA: John Benjamins.
- Pilonieta, P. (2010). Instruction of research-based comprehension strategies in basal reading programs. *Reading Psychology, 31*(2), 150-175.
- Proctor, C. P., Carlo, M., August, D., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology, 97*(2), 246-256.
- Protopapas, A., Mouzaki, A., Sideridis, G. D., Kotsolakou, A., & Simos, P. G. (2013). The role of vocabulary in the context of the simple view of reading. *Reading and Writing Quarterly, 29*(2), 168-202.

- Protopapas, A., Simos, P. G., Sideridis, G. D., & Mouzaki, A. (2012). The components of the simple view of reading: A confirmatory factor analysis. *Reading Psychology, 33*(3), 217-240.
- Pufpaff, L. A. (2009). A developmental continuum of phonological sensitivity skills. *Psychology in the Schools, 46*(7), 679-691.
- Rasinski, T., Homan, S., & Biggs, M. (2009). Teaching reading fluency to struggling readers: Methods, materials, and evidence. *Reading and Writing Quarterly, 25*(2), 192-204.
- Ravitch, D. (2009). Time to kill no child left behind. *Education Week, 75*(1), 4-6.
- Renaissance Learning (2007). *Understanding STAR Assessments*. Retrieved June 1, 2013 from <http://www.wpsok.org/BurchamWeb/Assessment%20Benchmarks%202011/Understanding%20Star%20Assessments.pdf>
- Renaissance Learning. (2011). *The Foundation of the STAR Assessments*. Retrieved June 1, 2013 from <http://doc.renlearn.com/KMNet/R003957507GG2170.pdf>
- Riccards, P. (2012). Read all about it! What states are doing to ensure kids are reading by third grade. Retrieved from: <http://files.eric.ed.gov/fulltext/ED540344.pdf>
- Ryder, J. F., Tunmer, W. E., & Greaney, K. T. (2007). Explicit instruction in phonemic awareness and phonemically based decoding skills as an intervention strategy for struggling readers in whole language classrooms. *Reading and Writing, 21*(4), 349-369.
- Savage, R., Carless, S., & Ferraro, V. (2007). Predicting curriculum and test performance at age 11 years from pupil background, baseline skills and phonological awareness at age 5 years. *Journal of Child Psychology and Psychiatry, 48*(7), 732-739.
- Savage, R., & Wolforth, J. (2007). An additive simple view of reading describes the performance of good and poor readers in higher education. *Exceptionality Education Canada, 17*(2),

243-268.

- Schwanenflugel, P. J., Meisinger, E. B., Wisenbaker, J. M., Kuhn, M. R., Strauss, G. P., & Morris, R. D. (2006). Becoming a fluent and automatic reader in the early elementary school years. *Reading Research Quarterly, 41*(4), 496-522.
- Shanahan, T. (2003). Research-based reading instruction: Myths about the national reading panel. *The Reading Teacher, 56*(7), 646-655.
- Shippen, M. E., Houchins, D. E., Crites, S. A., Derzis, N. C., & Patterson, D. (2010). An examination of the basic reading skills of incarcerated males. *Adult Learning, 21*(3), 4-12.
- Siegest, J., & Van Patten, J. (2007). No child left behind: Catalyst and controversy. *Journal of Philosophy and History of Education, 57*(1), 146-157.
- Sloat, E. A., Beswick, J. F., & Willms, J. D. (2007). Using early literacy monitoring to prevent reading failure. *Phi Delta Kappan, 88*(7), 523-529.
- Stapleton, P. D. (1999). Superintendent's memo no. 26. Retrieved from <http://www.doe.virginia.gov/testing/>
- Stout, R. (2009). Putting literacy centers to work: A novice teacher utilizes literacy centers to improve reading instruction. *Networks, 11*(1), 1-6.
- Stuart, M., Stainthorp, R., & Snowling, M. (2008). Literacy as a complex activity: Deconstructing the simple view of reading. *Literacy, 42*(2), 59-66.
- Stuebing, K. K., Barth, A. E., Cirino, P. T., Francis, D. J., & Fletcher, J. M. (2008). A response to recent reanalyses of the national reading panel report: Effects of systematic phonics instruction are practically significant. *Journal of Educational Psychology, 100*(1), 123-134.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA:

- Pearson.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.
- Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading*, 32(4), 383-401.
- Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, 45(5), 453-466.
- U.S. Department of Education. (2002). *Elementary and Secondary Education: Part A – Improving Basic Programs Operated by Local Educational Agencies*. Retrieved June 10, 2013, from <http://www2.ed.gov/policy/elsec/leg/esea02/pg2.html#sec1116>
- U.S. Department of Education. (2002, April). *Guidance for the Reading First program*. Retrieved June 10, 2013, from <http://www2.ed.gov/programs/readingfirst/guidance.pdf>
- Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, 22(3), 407-423.
- Virginia Department of Education. (2010). *English Standards of Learning for Virginia Public Schools: Grade Three*. Retrieved June 10, 2013, from http://www.doe.virginia.gov/testing/sol/standards_docs/english/2010/stds_english3.pdf
- Virginia Department of Education. (2010). *Virginia Standards of Learning Assessments: Technical Report 2010-2011 Administration Cycle*. Retrieved June 29, 2013, from http://www.doe.virginia.gov/testing/test_administration/technical_reports/sol_technical_report_2010-11_administration_cycle.pdf

- Virginia Department of Education. (2011). *Virginia Standards of Learning Assessments Test Blueprint: Grade 3 Reading*. Retrieved June 10, 2013, from http://www.doe.virginia.gov/testing/sol/blueprints/english_blueprints/2010/2010_blueprint_gr3_reading.pdf
- Wolpert, G., & Vacca-Rizopoulos, L. A. (2012). Assessing narrative and expository reading passages by text and online presentation. *American International Journal of Contemporary Research*, 2(4), 39-46.
- Wright, D., & Ehri, L. (2007). Beginners remember orthography when they learn to read words: The case of doubled letters. *Applied Psycholinguistics*, 28(1), 115-133.
- Young, L. E. (2013). Poor school performance linked to incarceration. Retrieved from: <http://bursapdf.info/PDF-poor-school-performance-linked-to-incarceration-five-part-series-1412>

APPENDIX A: PERMISSION LETTER FROM SCHOOL DIVISION

PUBLIC SCHOOLS

April 10, 2014

Mr. Todd Johnson
Elementary School
, VA

Dear Mr. Johnson:

You have permission to utilize second grade PALS scores together with third grade scores on the STAR Reading Test, the QRI and SOL reading tests. This testing data will be provided to you after it has been completely stripped of all identifying information pertaining to individual students. I am aware that this request is being made for the purpose of your dissertation, and the school name and county will not be revealed in your study.

Sincerely,

, Ed.D.
Assistant Superintendent
County Public Schools

APPENDIX B: INSTITUTIONAL REVIEW BOARD APPROVAL

LIBERTY UNIVERSITY
INSTITUTIONAL REVIEW BOARD

April 11, 2014

Todd Allen Johnson

IRB Application 1845: Using Beginning-of-Year Diagnostic Reading Measures to Predict Third-Grade Comprehension Scores in Virginia

Dear Todd,

The Liberty University Institutional Review Board has reviewed your application in accordance with the Office for Human Research Protections (OHRP) and Food and Drug Administration (FDA) regulations and finds your study does not classify as human subjects research. This means you may begin your research with the data safeguarding methods mentioned in your approved application.

Your study does not classify as human subjects research because your study involves deidentified archival data.

Please note that this decision only applies to your current research application, and that any changes to your protocol must be reported to the Liberty IRB for verification of continued non-human subjects research status. You may report these changes by submitting a new application to the IRB and referencing the above IRB Application number.

If you have any questions about this determination, or need assistance in identifying whether possible changes to your protocol would change your application's status, please email us at irb@liberty.edu.

Sincerely,

Fernando Garzon, Psy.D.
Professor, IRB Chair
Counseling

(434) 592-4054

LIBERTY
UNIVERSITY.