PREDICTING NORTH CAROLINA THIRD GRADE END-OF GRADE TEST OF READING COMPREHENSION SCORES FROM FIRST, SECOND, AND THIRD GRADE VARIABLES

Sean B. Knuth

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Education (School Psychology).

Chapel Hill
2011

Approved by:
Steve E. Knotek, Ph.D.
Stephen R. Hooper, Ph.D.
William B. Ware, Ph.D.
Rune J. Simeonsson, Ph.D.
Barbara H. Wasik, Ph.D.
ABSTRACT

Sean B. Knuth: Predicting North Carolina Third Grade End-of-Grade Test of Reading Comprehension Scores From First, Second, and Third Grade Variables

(Under the direction of Dr. Steve Knotek and Dr. Stephen R. Hooper)

A significant body of research exists on the development of early literacy skills and their relationship to the development of literacy as a whole. Phonological awareness, orthographic processing, rapid automatized naming, phonological memory, and receptive vocabulary have all been shown to be predictive of early reading outcome measures. What is unknown, however, is whether or not these variables are predictive of the North Carolina third grade End of Grade Test of Reading Comprehension, a major outcome measure used in North Carolina to determine student, teacher, school, and district academic performance.

This study addressed the following research questions: 1) Do the variables identified through a review of the literature as being predictors of reading achievement in first, second, and third grade contribute to scores on the reading EOG? 2) Does growth on measures of these variables, identified through a review of the literature as being predictors of reading achievement in first, second, and third grade contribute to scores on the reading EOG?

Linear regressions were conducted on data consisting of 111 children. After controlling for age, IQ, and inherent characteristics of the data set, five predictor variables in grades one, two, and three were found to be significantly predictive of the outcome measure. A significant portion of variance was accounted for by receptive vocabulary at time point one; phonological awareness, RAN and receptive vocabulary at time point two; and
orthographic processing and RAN at time point three. Further examination indicated
phonological awareness and RAN were most predictive and time point two and orthographic
processing was most predictive at time point three. Phonological memory was never
significantly predictive at any given time point but contributed the most to outcome measure
prediction at time point two.

Findings suggest scores on the reading EOG are predicted by a child’s development
on skills key to the development of early literacy. An individual’s literacy skills can be used
to estimate later performance on this high stakes test of reading ability. The results of this
study suggest students can be screened for potential EOG failure and interventions can be
implemented to remediate key skills. This study also suggests a model for the evaluation of
other high stakes outcome measures.
ACKNOWLEDGEMENT

There are many people without whom this work would not be accomplished. I would like to thank in particular Dr. Stephen Hooper for his guidance, mentorship, and provision of motivation. I would also like to thank Kathleen Anderson, Donna Yerby, and the rest of the Writing Skills Development Project team for their tireless efforts throughout the length of the project that developed the database upon which this study is based. Additionally, I would like to thank the faculty at the University of North Carolina at Chapel Hill’s School Psychology program, especially Dr. Steve Knotek, for the mentorship, encouragement, and guidance. Beyond the scope of such direct support, however, lays an underlying network of encouragement, love, and motivation without which such an accomplishment would be difficult to obtain and for me much less meaningful: Steve, Sue, and Kevin Knuth; Roberta and Frank Kozeliski; and my many friends both professional and personal.

I’ve no doubt without the love, support, and encouragement of my wife, Dr. Corinne Zeller-Knuth, I would not have completed this project (nor many others). She deserves her own paragraph, and she always will.
# TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................ ix
ABBREVIATIONS .................................................................................................... x

CHAPTER

I. INTRODUCTION ............................................................................................... 1

  Overview ............................................................................................................. 1
  Current Study ..................................................................................................... 4
  Statement of Purpose .......................................................................................... 4
  Questions and Hypotheses .................................................................................. 5

II. LITERATURE REVIEW ...................................................................................... 9

  Introduction ......................................................................................................... 9
  Development of Reading .................................................................................... 10
  Reading Outcome Measures .............................................................................. 15

    End of grade tests ............................................................................................ 17
    Characteristics of the North Carolina EOG test of reading comprehension .......... 18

  Early Predictors of Reading Outcome .............................................................. 25

    Predicting reading success prior to first grade ............................................ 26
  Specific Predictors of Reading Skills ................................................................. 28

    Familial risk of dyslexia ............................................................................... 28
    Concepts of print ............................................................................................. 29
Phonological awareness .................................................................30
Orthographic processing ..............................................................33
Rapid automatized naming ............................................................34
Phonological memory .................................................................36
Receptive vocabulary .................................................................37

Summary ......................................................................................39

III. METHODS ...........................................................................41
Participants ..................................................................................41
Criterion Measure .........................................................................42
Predictor Variables .........................................................................44
Phonological awareness measures ............................................44
Orthographic processing measures .........................................46
Rapid automatized naming measures ........................................47
Phonological memory measures ...............................................48
Receptive vocabulary measures ...............................................49

Procedures ..................................................................................51
Data Analysis ................................................................................52
Preliminary data analysis .............................................................52

Research Question, Hypothesis, and Data Analysis ................56
Question one ................................................................................56
Question two ................................................................................57

IV. RESULTS ...........................................................................60
Preliminary Analysis ......................................................................60
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>North Carolina EOG Achievement Levels</td>
</tr>
<tr>
<td>1.2</td>
<td>Grade 3 Reading EOG Results by Level by Year</td>
</tr>
<tr>
<td>1.3</td>
<td>Grade 3 Reading EOG Concept Specifications</td>
</tr>
<tr>
<td>1.4</td>
<td>North Carolina Standard Course of Study, Language Arts, Grade 3</td>
</tr>
<tr>
<td>3.1</td>
<td>Demographic Information</td>
</tr>
<tr>
<td>3.2</td>
<td>Table of Reliability and Validity of Measures</td>
</tr>
<tr>
<td>4.1</td>
<td>Table of Predictor Variable Composition</td>
</tr>
<tr>
<td>4.2</td>
<td>Table of AR vs. TD Variable as Related to Age and IQ</td>
</tr>
<tr>
<td>4.3</td>
<td>Table of Means and Standard Deviations</td>
</tr>
<tr>
<td>4.4</td>
<td>Table ofRegression Data, Analyses for Question One</td>
</tr>
<tr>
<td>4.5</td>
<td>Table of Regression Data, Analyses for Question Two</td>
</tr>
<tr>
<td>5.1</td>
<td>North Carolina Third Grade EOG Test of Reading Comprehension Goals</td>
</tr>
<tr>
<td>5.2</td>
<td>Predictive Regression Equations with All Covariates</td>
</tr>
<tr>
<td>5.3</td>
<td>Predictive Regression Equations with Age as Covariate</td>
</tr>
<tr>
<td>5.4</td>
<td>Actual Achievement Levels vs. Predicted Achievement levels</td>
</tr>
</tbody>
</table>
**ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>At risk</td>
</tr>
<tr>
<td>CELF-4</td>
<td>Clinical Evaluation of Language Fundamentals, 4th edition</td>
</tr>
<tr>
<td>CFA</td>
<td>Confirmatory Factor Analysis</td>
</tr>
<tr>
<td>CPT</td>
<td>Continuous Performance Test</td>
</tr>
<tr>
<td>CREVT-2</td>
<td>Comprehensive Receptive and Expressive Vocabulary Test – Second Edition</td>
</tr>
<tr>
<td>CTOPP</td>
<td>Comprehensive Test of Phonological Processing</td>
</tr>
<tr>
<td>EOC</td>
<td>End of Course</td>
</tr>
<tr>
<td>EOG</td>
<td>End of Grade</td>
</tr>
<tr>
<td>EVT-2</td>
<td>Expressive Vocabulary Test, 2nd edition</td>
</tr>
<tr>
<td>FPG</td>
<td>Franklin Porter Graham Child Development Institute</td>
</tr>
<tr>
<td>IQ</td>
<td>Intelligence Quotient</td>
</tr>
<tr>
<td>NCDPI</td>
<td>North Carolina Department of Public Instruction</td>
</tr>
<tr>
<td>NCSCS</td>
<td>North Carolina Standard Course of Study</td>
</tr>
<tr>
<td>NELP</td>
<td>National Early Literacy Panel</td>
</tr>
<tr>
<td>NRP</td>
<td>National Reading Panel</td>
</tr>
<tr>
<td>OP</td>
<td>Orthographic Processing</td>
</tr>
<tr>
<td>PA</td>
<td>Phonological Awareness</td>
</tr>
<tr>
<td>PM</td>
<td>Phonological Memory</td>
</tr>
<tr>
<td>PAL-RW</td>
<td>Process Assessment of the Learner: Test Battery for Reading and Writing</td>
</tr>
<tr>
<td>PAL II</td>
<td>Process Assessment of the Learner: Test Battery for Reading and Writing - Second Edition</td>
</tr>
<tr>
<td>RAN</td>
<td>Rapid Automatized Naming</td>
</tr>
<tr>
<td>RV</td>
<td>Receptive Vocabulary</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>TD</td>
<td>Typically developing</td>
</tr>
<tr>
<td>TOWRE</td>
<td>Test of Word Reading Efficiency</td>
</tr>
<tr>
<td>VIF</td>
<td>Variance Inflation Factor</td>
</tr>
<tr>
<td>WIAT-II</td>
<td>Wechsler Individual Achievement Test – Second Edition</td>
</tr>
<tr>
<td>WISC-IV</td>
<td>Wechsler Intelligence Scale for Children - Fourth Edition</td>
</tr>
<tr>
<td>WISC-IV-PI</td>
<td>Wechsler Intelligence Scale for Children - Fourth Edition Process Integrated</td>
</tr>
<tr>
<td>WRAT-3</td>
<td>Wide Range Achievement Test, 3rd Edition</td>
</tr>
<tr>
<td>WRMT-R</td>
<td>Woodcock Reading Mastery Tests, Revised</td>
</tr>
<tr>
<td>WJ-III-Cog</td>
<td>Woodcock Johnson III Tests of Cognitive Abilities</td>
</tr>
<tr>
<td>WRAML-2</td>
<td>Wide Range Assessment of Memory and Learning – Second Edition</td>
</tr>
<tr>
<td>WASI</td>
<td>Wechsler Abbreviated Scale of Intelligence</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Overview

A great deal of time and energy has been devoted to the study of reading development. Between 1966 and 2000 over 100,000 research studies have been published on the topic of reading (National Reading Panel, 2000), and a cursory search of the literature has uncovered at least an additional 52,000 since that time. This research has revealed many aspects of reading and learning to read. Specifically, different abilities contribute to the development of reading at different times and difficulties with one or more of these abilities can make learning to read difficult. The interaction between reading, its constituent abilities, and a child’s inherent cognitive and neuropsychological profile is complex. Fortunately, the development of reading follows predictable steps and several researchers have proposed models that provide for clearer conceptualization of the constituent parts.

The first such model was a Simple View of Reading, proposed by Gough and Tunmer (1986). This model is comprised of two factors: decoding, defined as the act of identifying written words as being parts of speech; and comprehension, defined as making sense of written words. Scarborough (2003) took the Simple View of Reading further and proposed a similar two-factor model with each model comprised of several skills, or “strands.” One of Scarborough’s two factors, Language Comprehension, is comprised of five strands: background knowledge, vocabulary, language structures, verbal reasoning, and literacy knowledge. The other factor, Word Recognition, is comprised of three strands: phonological
awareness, decoding, and sight recognition. Each of Scarborough’s factors, comprised of several strands, combines to produce skilled reading. In their metareview of all reading related research studies conducted between 1966 and 2000, the National Reading Panel (2000) identified three major areas identified as significantly contributing to the development of reading skills: alphabetics, including phonemic awareness and phonics; fluency; and comprehension, which includes vocabulary.

The skills described by these strands all contribute to the development of literacy and research has shown the development of literacy follows predictable steps. Before sense can be made of a sentence, a reader must first know how to read its words well enough to “free up cognitive resources for the comprehension process,” (Scarborough, 2003, p. 98). Reading a word involves not just having sufficient vocabulary to know its meaning but understanding the fact that letters used in creating a word represent sounds. These skills build on each other. Knowing the extent to which a child does or does not possess basic reading skills allows us to make fairly accurate predictions of the extent to which he will possess related skills at a future time. For example, in the absence of intervention, 56% of a child’s reading accuracy in third and fourth grade can be accounted for by grade two reading accuracy (Storch & Whitehurst, 2002). Current reading skills are predictive of future reading skills. Knowing this allows us to measure current skill levels and determine if intervention is needed.

A key part of being able to determine efficacy of reading interventions, and to explore reading skills in general, is through the use of reading outcome measures. Some of these, like teacher ratings of child proficiency, are informal measures often implemented to provide simple expert-opinion judgment of a child’s reading ability and progress. There are, however, significant drawbacks to using teacher judgment as a source of child evaluations, as teacher
expectations can effect achievement both positively and negatively (Hinnant, O’Brien, & Ghazarian, 2009). Though likely beneficial to students for whom expectation is high, this phenomenon can be problematic when students are viewed as being less likely to experience success.

More formalized measures, such as standardized psychoeducational measures of academic achievement, allow for more accurate comparisons between students and between time points. In order to measure student academic achievement in its schools North Carolina makes use of End-of-Grade (EOG) and End-of-Course (EOC) reading comprehension tests. EOG and EOC measures of reading comprehension are administered at the end of grades three through eight and again in high school in the form of the North Carolina High School Comprehensive Reading Test, and the English I End-of-Course Reading Test. These tests were developed by the North Carolina Testing Program (NCTP) and are used to assist with data-based decision making across many aspects of the educational system.

According to the North Carolina Department of Public Instruction (Office of Curriculum and School Reform Services, 2004), the NCTP utilizes results of EOG and EOC tests:

(1) to assure that all high school graduates possess the … skills and knowledge thought necessary to function as a member of society; (2) to provide a means of identifying strengths and weaknesses in the education process; and (3) to establish additional means for making the education system accountable to the public for results. (p. 10)

Assuming EOGs and EOCs are psychometrically sound measures of reading, and given the fact that early reading progress is predictable, consistent, and measurable, a student’s performance on the third grade EOG Test of Reading Comprehension should be estimatable by established predictors of reading ability. Determining the extent to which scores on the North Carolina third grade EOG Test of Reading Comprehension are
predictable through the assessment of skills deemed to be important to the development of early literacy will allow for the identification of students at-risk for reading failure. These students can then be provided with additional and more intensive interventions in order to improve their reading skills.

**Current Study**

The purpose of this study was to examine the relative contributions of educational and neuropsychological predictors measured in the first, second, and third grade to scores on the North Carolina third grade End-of-Grade (EOG) Test of Reading Comprehension. Data was collected from a sample of students enrolled in a single school district in North Carolina. A battery of standardized cognitive and achievement tests was administered during their first, second, and third grade year, and scores on the third grade EOG Test of Reading Comprehension were collected from school records. The data was analyzed to determine the relative contribution these selected variables make to EOG reading scores.

**Statement of Purpose**

This study examined the predictive value of selected early educational and neuropsychological measures on later reading achievement. By identifying variables that allow for accurate and quick prediction of later difficulties in reading educators and psychologists will be better able to identify individual students who would benefit from additional educational resources. Additionally, the predictive contribution of measures of early literacy skills to scores on the North Carolina third grade EOG Test of Reading Comprehension was investigated.

**Questions and Hypotheses**
*Question one.* Do the variables identified through a review of the literature as being predictors of reading achievement in first, second, and third grade contribute to scores on the 3rd grade North Carolina EOG Test of Reading Comprehension?

- *Hypothesis 1-1:* The phonological awareness variable, as measured in first grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.
- *Hypothesis 1-2:* The phonological awareness variable, as measured in second grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.
- *Hypothesis 1-3:* The phonological awareness variable, as measured in third grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.
- *Hypothesis 1-4:* The orthographic processing variable, as measured in first grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.
- *Hypothesis 1-5:* The orthographic processing variable, as measured in second grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.
- *Hypothesis 1-6:* The orthographic processing variable, as measured in third grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.
• **Hypothesis 1-7:** The rapid automatic naming variable, as measured in first grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

• **Hypothesis 1-8:** The rapid automatic naming variable, as measured in second grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

• **Hypothesis 1-9:** The rapid automatic naming variable, as measured in third grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

• **Hypothesis 1-10:** The phonological memory variable, as measured in first grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

• **Hypothesis 1-11:** The phonological memory variable, as measured in second grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

• **Hypothesis 1-12:** The phonological memory variable, as measured in third grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

• **Hypothesis 1-13:** The receptive vocabulary variable, as measured in first grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.
• **Hypothesis 1-14:** The receptive vocabulary variable, as measured in second grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

• **Hypothesis 1-15:** The receptive vocabulary variable, as measured in third grade, will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

*Question two.* Does the growth on measures of these variables, identified through a review of the literature as being predictors of reading achievement in first, second, and third grade contribute to scores on the 3rd grade North Carolina EOG Test of Reading Comprehension?

• **Hypothesis 2-1:** Change over time in phonological awareness will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

• **Hypothesis 2-2:** Change over time in orthographic processing ability will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

• **Hypothesis 2-3:** Change over time in rapid automatic naming ability will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

• **Hypothesis 2-4:** Change over time in phonological memory ability will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

• **Hypothesis 2-5:** Change over time in receptive vocabulary ability will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.
CHAPTER II
LITERATURE REVIEW

Introduction

The state of North Carolina makes use of standardized instruments to assess the reading ability of its students. Though not the only instrument employed in North Carolina to measure reading outcomes, the North Carolina End of Grade (EOG) Tests of Reading Comprehension are arguably the most significant as the results of these assessments are used to assist in making many decisions pertinent to students, teachers, and budgets. Administered to students at various points during their educational career, these instruments are used to determine if children are progressing in their skill development at an acceptable pace, if teachers are providing adequate instruction, and if schools and districts are performing as expected and desired (Office of Curriculum and School Reform Services, 2004).

These tests were developed to measure expected competencies set forth in the North Carolina Standard Course of Study (North Carolina Department of Public Instruction [NCDPI], 2004). Underlying these competencies are abilities believed to be key to the acquisition of literacy. As there is evidence said abilities have been demonstrated to be predictive of the development of reading skills, and since the third grade EOG Test of Reading Comprehension purports to measure reading skills, measures of these abilities should be predictive of third grade EOG Test of Reading Comprehension scores. This dissertation will help guide our understanding of the extent to which third grade EOG Test of Reading Comprehension scores are predictable given our current understanding of the
development of literacy. The following literature review examines our understanding of the
development of early literacy, and describes the different types of instruments used to assess
reading skills, including the North Carolina third grade EOG Test of Reading
Comprehension. Finally, several key early predictors of reading will be discussed.

**Development of Reading**

Though written language is based on spoken language, learning to read is a much
more involved and difficult process than learning to understand spoken language (Lundberg,
2009). Learning to read is a dynamic process and the skills involved in reading are gained
gradually and over time. They are acquired in a predictable manner by children who have
normal or above-average language skills; who have had experiences in early childhood that
fostered motivation and provided exposure to literacy in use; who get information about the
nature of print through opportunities to learn letters and to recognize the internal structure of
spoken words, as well as explanations about the contrasting nature of spoken and written
language; and who attend schools that provide effective reading instruction and opportunities
to practice reading (Jansky & de Hirsch, 1972). The way these components interact to foster
literacy development is at the heart of reading development research.

Reading research theory has changed over the past 30 years, and the variables
investigated for their ability to predict reading ability have evolved as the theories have
changed (Gaffney & Anderson, 2000). For example, in the early days of reading research,
difficulty with reading was generally thought to be related to visual perception problems,
specifically as a result of perceived letter reversals (Catts & Hogan, 2003). Empirical
research into the acquisition of skills related to reading has not supported this theory,
however, as the predictive value of such problems has not been demonstrated (McCardle,
Scarborough, & Catts, 2001). In a study of visual-temporal processing, Olson and Datta (2002) did not find sufficient differences in such processing to account for much variance in reading ability between 356 reading-disabled and normal twins. In a longitudinal study of 540 children, Schatschneider et al. (2004) found perceptual skills as measured in kindergarten not to be good predictors of reading ability in first and second grade. Though the reason for the diminished assumption of the importance of perceptual measures is unclear (Schatschneider et al., 2004), there remain many more well investigated variables theorized to play important roles in the acquisition of reading skills.

There are many parts to reading, and skilled reading requires “that the processes involved in word recognition become so well practiced that they can proceed extremely quickly and almost effortlessly, freeing up the reader’s cognitive resources for comprehension processes,” (Scarborough, 2003, p. 98). Several researchers (e.g., Hirsch and Janskey, 1972; McCardle et al., 2001) report on other factors thought to contribute to the acquisition of reading skills, such as age, gender, IQ, neurological development, visual-motor integration, auditory processing, intersensory integration, emotional factors, and oral language development which includes such concepts as phonological awareness, vocabulary, writing, listening, and speaking.

All of these factors no doubt contribute, at least in part, to the development of reading skills. Gough and Tunmer (1986), however, proposed a Simple View of Reading, a way to conceptualize reading as being composed of two general factors: comprehension and decoding. Comprehension is the understanding of the meaning embedded in written language, and decoding is the ability to see written words and to recognize them as a part of speech. Most all variability among readers can be accounted for by measuring each of these
two broad skills (Torgesen, 2002), and each of these aspects are best conceptualized as developing along a continuum rather than in a binary, all-or-none fashion. (Whitehurst & Lonigan, 1998). The relative importance of each trait changes as the individual develops their reading skills; what is important to possess at one point in time becomes secondary to an additional skill as an individual’s reading skills develop (Compton, Fuchs, Fuchs, & Bryant, 2006).

To formally investigate and aggregate the research on reading development to-date, in 1997 the National Institute of Child Health and Human Development convened a National Reading Panel (NRP), “to assess the status of research-based knowledge” in the area of reading (National Reading Panel, 2000, p.1). Through their examination of all of the reading literature to date, they ascertained three general areas of reading skills development they determined to be key to the successful acquisition of literacy. These areas are alphabetics, which include such skills as phonemic awareness and letter-sound associations; fluency, which partially consists of orthography and sight-word recognition; and comprehension, which includes such skills as vocabulary and text comprehension. All of these general areas and their component skills were deemed by the NRP to be critically important to the development of literacy.

As a result of subsequent research, a robust model of reading development was proposed by Scarborough (2003). In his model, he incorporates the findings from the NRP and breaks down the Simple View of Reading’s two factors, describing each as being comprised of several strands. Within the comprehension factor are such strands as vocabulary, background knowledge, and verbal reasoning. Within the Word Recognition factor are such strands as phonological awareness, decoding, and sight recognition. While
each of these skills are crucial to the reading process in and of themselves, they all combine and influence each other to contribute to a child’s overall reading skills.

This integrationist view was largely supported by the findings of the National Early Literacy Panel (NELP). Convened in 2002 to conduct a meta-analysis of reading research pertaining to children from birth to age 5, the focus of the NELP was to establish, in a formal manner, the findings of decades of reading research pertaining to children younger than that examined by the NRP (National Early Literacy Panel, 2008). The findings from the NELP support early models in their identification of six primary variables identified as being correlated with later literacy development. These six skills are alphabetic knowledge, phonological awareness, rapid automatic naming of letters or digits, rapid automatic naming of colors, the ability to write one’s name, and phonological memory. In addition to these six variables, five additional variables were identified that contribute to early literacy development, but become of diminished importance when subsequent variables were accounted for. These five additional variables are concepts of print, print knowledge, reading readiness, oral language ability, and visual processing. These eleven variables were determined to be predictive of reading outcome measures in total. The latter five, however, were determined to be of greater predictive value when examining outcome measures in late kindergarten and early first grade, rather than further along in a child’s elementary school experience.

In all, reading research has identified several skills deemed important to the acquisition of literacy. These skills all contribute to a child’s ability to read, and difficulties in one or more of these skills can lead to reciprocal difficulties in others. For example, a reciprocal relationship has been established between attaining phonological awareness and
learning to decode print (Burt, 2006; Perfetti, Beck, Bell, & Hughes, 1987). Just as phonemic awareness contributes to an individual’s ability to learn to read, so too does an individual’s ability to read help develop his phonemic awareness. Another example of the reciprocal nature of reading skill acquisition is found in a study by Sprenger-Charolles, Siegel, Béchennec, & Serniclaes (2002), in which the development of phonological processing and orthographic processing was demonstrated to be reciprocally related, rather than to be “independent components of word recognition.” This reciprocal relationship between the development of skills important to the acquisition of literacy becomes even more noticeable when readers experience difficulties in more than one area. This has been referred to as a “double deficit” model. Children with difficulties in more than one area are at increased risk for additional reading difficulties (e.g., Catts & Hogan, 2003; McBride-Chang and Manis, 1996, McCardle et al, 2001; Wolf et al., 2000).

Fortunately, future difficulties in reading have been shown to be predictable. An individual’s reading skills in one year greatly determines reading skills in subsequent years. For example, Storch and Whitehurst (2002) examined oral language ability in 626 four year olds, and followed them through fifth grade. They found 90% of first grade oral language ability accounted for by preschool ability; 96% of grade one and two variance accounted for by kindergarten ability; and 88% of grade three and five variance accounted for by grade one and two ability. Additionally, a child’s reading accuracy in third and fourth grade was 56% accounted for by grade two reading accuracy. In another study, Boscardin, Muthen, Francis, and Baker (2008) administered measures of phonological awareness, word recognition, and rapid naming skills to 411 children four times a year for three years, starting in kindergarten. They found all three skills to be highly predictive of word reading, as measured in the second
grade, and that students who were identified as having reading difficulties in kindergarten demonstrated slower acquisition of word recognition skills throughout first and second grade.

Learning to read is a complicated but generally predictable process. It involves the acquisition of many skills that build upon each other, and each of these skills contributes differently at different times during a child’s development. If educators are to intervene effectively to help children whose reading skills are in danger of falling behind, it becomes important to be able to measure these skills, and to assess a child’s progress across all components important to the development of reading.

**Reading Outcome Measures**

There are two broad categories of instruments used to assess educational outcomes in general, and reading outcomes in particular. More formalized measures provide excellent reliability but are often criticized for attempting to measure cognition and learning decomposed into component parts and isolated, or decontextualized, from the situations in which the knowledge of interest is acquired and applied (Paris, Lawton, Turner, & Roth, 1991). These measures are also criticized for encouraging teachers to “teach to the test” (Salinger, 2001). Informal reading outcome measures such as portfolio assessments, teacher judgment, as well as performance measures such as curriculum based measurement address the issue of decontextualization by assessing students in situ and making judgments on skills based on real world performance. They do, however, suffer from a lack of reliability (Paris et al., 1991).

More formal outcome measures, such as the Iowa Test of Basic Skills (Hieronymous, Lindquist, & Hoover, 1980), or the Woodcock Johnson Test of Achievement, Third Edition (Woodcock, McGrew, & Mather, 2001), are constructed in a psychometrically sound
manner, in accordance with industry standards (e.g., *Standards for Educational and Psychological Testing*, AERA et al., 1999). These types of assessments have undergone rigorous norming procedures and have demonstrated reliability. They are, however, subject to some of the criticism mentioned above in that they decontextualize the assessed knowledge from the situation in which it is typically applied. An example of how such decontextualization can fail to accurately assess knowledge can be found in a study conducted by Saxe (1988) on twenty three 10-12 year old street vendors with no little or no formal education. Though these children had difficulties correctly answering traditional math problems, they were able to calculate relatively large numerical values in their head. Certainly they would have fared poorly on a traditional measure of mathematical achievement even though they were able to carry out analogous calculations within a framework that made sense to them. Their ability to solve mathematical problems within a familiar context belied their inability to solve decontextualized mathematical problems.

There are pros and cons of formal and informal testing instruments. In its most basic sense a test administrator trades reliability for validity when he selects informal measures and trades validity for reliability when he administers formal measures. The importance of each of these traits, particularly with respect to the educational agency being assessed, should be evaluated before selecting methods of skills assessments

**End of grade tests.** The North Carolina End-of-Grade (EOG) and End-of-Course (EOC) tests are standardized measures used by the state of North Carolina to assess grade level performance across several different domains. The use of such measures are required by state law. The North Carolina Elementary and Secondary Reform Act (Elementary and Secondary School Reform Act of 1984, 1983) mandated the use of EOC tests, and its
reauthorization in 1989 introduced the requirement of EOG testing. The language of the legislation decreed EOG and EOC tests had two purposes: to measure individual skills and knowledge based on North Carolina Standard Course of Study (NCSCS), and to measure knowledge and skills attained by subgroups of the school population, the school system, and state accountability. These tests were developed by the NCDPI according to a process codified in 2003 by the State Board of Education (Office of Curriculum and School Reform Services, 2004, p. 14). This process was designed to ensure test instruments developed by the state met basic psychometric standards.

One such standard is the constitutional requirement established in Debra P. vs. Turlington (1984). In this case, it was established that achievement tests used for the purpose of determining grade promotion and diplomas need to be constructed in a psychometrically sound manner and requiring a student to pass a test for which he has not been prepared was deemed unconstitutional. In addition to the requirements set forth by this decision relevant associations have created lists of standards to which test construction should adhere in order to claim a test is psychometrically sound (AERA et al., 1999).

The NCDPI has published a technical report detailing the development of its language arts EOGs and EOCs, one of which is the third grade EOG Test of Reading Comprehension. The information in this document provides information useful for judging the psychometric validity of EOGs (Office of Curriculum and School Reform Services, 2004).

**Characteristics of the North Carolina EOG test of reading comprehension.** The third grade EOG Test of Reading Comprehension went through twenty discrete steps before being deemed as fully operational, a process that took between 44 and 49 months. Item construction was done through the generation of an item pool written by North Carolina
educators and selected based on knowledge of the current NCSCS. Additionally, some item
development was contracted to an external vendor who also was encouraged to employ North
Carolina educators to assist with item creation.

Each item was classified by item constructors into one of three categories of
difficulty. The three categories of item difficulty were easy, with a p-value of .7, defined as
questions approximately 70% of examinees could answer correctly; medium, with a p-value
of .5 to .6, defined as questions 50% to 60% of examinees could answer correctly; and hard,
with a p-value of .2 to .3, defined as questions 20% to 30% of examinees could correctly
answer. The percentage of each item difficulty type found in any given test was not made
available. The overall p-value for the third grade reading test item pool was .629. The
coefficient alpha for the third grade test was calculated to be .9245.

Items were examined for potential bias in two ways. The first was through educator
review. Each item was examined by selected educators for potential biases. Additionally,
items also were examined for differential item functioning using Mantel-Haenszel Log Odds
Ratio differential item functioning statistics. An item is said to exhibit differential item
functioning if it reliably favors individuals with traits other than those attempting to be
measured by the item itself (Penfield, 2001). An item that displays no differential functioning
has a log odds ratio of 1.0. If an item was found to have a log odds ratio greater than 1.5 or
less than .67 with respect to ethnicity or gender, or if an item was determined to be biased
through educator review, it was removed from the item pool.

Normative data for the third grade EOG Test of Reading Comprehension were
initially collected at the state level in 1993 and then again in 1998 and 1999. Normative data
for the third edition of End-of-Grade reading test were created in 2007 and the test was
implemented in 2008. Each student’s results are reported in four ways: as a scaled score, as a percentile, as either pass/fail, and at an achievement level. The scaled score is calculated by the DPI, and in the most recent year for which data were available, 2007, the mean scale score was 338.65 (SD = 12.57). From these scale scores, cut points are determined through a process described as “contrasting groups.” Ninety five percent of the students involved in the 2007 norming process were rated by their teachers as performing in class at one of four levels (Table 1.1) independent of their achieved score on the pilot test. The proportion of the sample rated as performing in each of the four categories determined the cut points for the third grade End-of-Grade Test of Reading Comprehension. An examination of third grade reading EOG test results from 1995 until 2003, the most recent year for which results were available, reveals gradually increasing scores (Table 1.2).

Table 1.1
North Carolina EOG Achievement Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
<th>Grade 3 Scale score range</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Students performing at this level do not have sufficient mastery of knowledge and skills in a particular subject area to be successful at the next grade level.</td>
<td>&lt;331</td>
</tr>
<tr>
<td>II</td>
<td>Students performing at this level demonstrate inconsistent mastery of knowledge and skills in the subject area and are minimally prepared to be successful at the next grade level.</td>
<td>331-337</td>
</tr>
<tr>
<td>III</td>
<td>Students performing at this level consistently demonstrate mastery of the grade level subject matter and skills and are well prepared for the next grade.</td>
<td>338-349</td>
</tr>
<tr>
<td>IV</td>
<td>Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient at grade level</td>
<td>&gt;349*</td>
</tr>
</tbody>
</table>

Note: the formula for transformation of scaled score to achievement levels was changed starting in 2007. These values represent the new ranges.
In order to establish that an instrument has been constructed in a psychometrically sound manner, evidence must show the instrument is both reliable and valid. Reliability refers to the consistency of test scores across measurements. Reliability for the third grade EOG Test of Reading Comprehension was established through the use of internal consistency measures; specifically, Cronbach’s coefficient alpha (Cronbach, 1951). The coefficient alpha for the test was determined to be .93, which is considered very good. Standard error of measurement for the test was calculated to range from two to five points, at the 95% confidence interval. Ranges at other confidence intervals were not reported.

A test is said to be valid if sufficient evidence is provided to demonstrate the test measures what it purports to measure. The establishment of validity is the single most important aspect of test construction (Angoff, 1988). In their report on the creation of the EOG, NCDPI offers several forms of validity for the EOG. By way of evidence for content validity, the Office of Curriculum and School Reform Services (2004) discusses the extent to which the third grade EOG Test of Reading Comprehension addresses four concepts each EOG is designed to measure: cognition, the most basic stage of reading comprehension; interpretation, the ability to more completely understand a text and expand on its ideas; critical stance, pertaining to a student’s ability to consider selections objectively; and connections, which involves relating selected readings to outside knowledge. Table 1.3

Table 1.2
*Grade Three Reading EOG Results by Level by Year (%)*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>12.9</td>
<td>11.3</td>
<td>11.0</td>
<td>8.6</td>
<td>6.9</td>
<td>6.2</td>
<td>5.7</td>
<td>4.2</td>
<td>3.9</td>
</tr>
<tr>
<td>II</td>
<td>23.7</td>
<td>23.9</td>
<td>23.2</td>
<td>19.8</td>
<td>19.5</td>
<td>19.4</td>
<td>17.9</td>
<td>16.0</td>
<td>13.5</td>
</tr>
<tr>
<td>III</td>
<td>37.2</td>
<td>37.9</td>
<td>37.6</td>
<td>36.3</td>
<td>36.7</td>
<td>38.0</td>
<td>38.4</td>
<td>38.8</td>
<td>37.1</td>
</tr>
<tr>
<td>IV</td>
<td>26.2</td>
<td>26.9</td>
<td>28.3</td>
<td>34.3</td>
<td>36.9</td>
<td>36.4</td>
<td>38.0</td>
<td>41.0</td>
<td>45.5</td>
</tr>
</tbody>
</table>

In order to establish that an instrument has been constructed in a psychometrically sound manner, evidence must show the instrument is both reliable and valid. Reliability refers to the consistency of test scores across measurements. Reliability for the third grade EOG Test of Reading Comprehension was established through the use of internal consistency measures; specifically, Cronbach’s coefficient alpha (Cronbach, 1951). The coefficient alpha for the test was determined to be .93, which is considered very good. Standard error of measurement for the test was calculated to range from two to five points, at the 95% confidence interval. Ranges at other confidence intervals were not reported.

A test is said to be valid if sufficient evidence is provided to demonstrate the test measures what it purports to measure. The establishment of validity is the single most important aspect of test construction (Angoff, 1988). In their report on the creation of the EOG, NCDPI offers several forms of validity for the EOG. By way of evidence for content validity, the Office of Curriculum and School Reform Services (2004) discusses the extent to which the third grade EOG Test of Reading Comprehension addresses four concepts each EOG is designed to measure: cognition, the most basic stage of reading comprehension; interpretation, the ability to more completely understand a text and expand on its ideas; critical stance, pertaining to a student’s ability to consider selections objectively; and connections, which involves relating selected readings to outside knowledge. Table 1.3
describes average proportions on each test form for the third grade EOG Test of Reading Comprehension devoted to each concept.

Table 1.3  
*Grade Three Reading EOG Concept Specifications*

<table>
<thead>
<tr>
<th>Grade 3</th>
<th>Average Number of Items Per Form</th>
<th>Average Percentage of Items Per Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognition</td>
<td>18.7</td>
<td>37.3%</td>
</tr>
<tr>
<td>Interpretation</td>
<td>18.3</td>
<td>36.7%</td>
</tr>
<tr>
<td>Critical Stance</td>
<td>9.7</td>
<td>19.3%</td>
</tr>
<tr>
<td>Connections</td>
<td>3.3</td>
<td>6.7%</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>100%</td>
</tr>
</tbody>
</table>

An instrument is said to have criterion validity if it is effective at predicting behavior similar to what the original instrument purports to measure. In the case of the third grade EOG Test of Reading Comprehension criterion validity was established through correlation between test results and other informal measures. Scale scores on the third grade EOG Test of Reading Comprehension correlated moderately with teacher judgment of achievement ($r = .63$) and expected course grades ($r = .67$). In addition, calculated achievement level, using the I, II, III, and IV rubric previously described in Table 1.1, correlated moderately with a student’s expected course grade ($r = .63$) and teacher judgment of achievement ($r = .63$). No evidence was provided for relationships between the North Carolina Reading End-of-Grade or End-of-Course tests and more traditional measures of reading achievement such as the Wechsler Individual Achievement Test-Second Edition (WIAT-II; Wechsler, 2002), or the Wide Range Achievement Test (WRAT-3; Wilkinson, 1993). In addition, a review of the literature revealed no studies investigating the relationship between the North Carolina reading EOGs and such measures.
The North Carolina Reading EOGs were constructed in a manner that provided some
evidence of criterion validity, a form of validity most formal achievement tests are criticized
as lacking (Paris, Lawton, Turner, & Roth, 1991). Efforts were made to ensure the test
content reflected what was being taught in the classroom. In order to accomplish this, North
Carolina teachers were recruited for the construction of items. After recruiting instructors to
create items for the reading EOGs, each writer was provided a three-day training session
which included materials pulled specifically from the state reading curriculum. These
materials included specific instructions that items be “based on the goals and objectives
outlined in the North Carolina Standard Course of Study (NCSCS) in Reading
Comprehension and written at the appropriate grade level” (Office of Curriculum and School
Reform Services, 2004, p. 76).

Criterion validity was also addressed through aligning the reading EOGs with the
language arts NCSCS. The language arts NCSCS consist of specific items of content
knowledge deemed important by the DPI and these items guide the creation and formation of
school courses. Table 1.4 lists Third Grade English Language Arts goals found in the
NCSCS, and the cognitive constructs to which they apply.
Table 1.4
North Carolina English Language Arts Standard Course of Study

<table>
<thead>
<tr>
<th>Competency Goal 1: The learner will apply enabling strategies and skills to read and write</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.01</td>
</tr>
<tr>
<td>1.02</td>
</tr>
<tr>
<td>1.03</td>
</tr>
<tr>
<td>1.04</td>
</tr>
<tr>
<td>1.05</td>
</tr>
<tr>
<td>1.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Competency Goal 2: The learner will apply strategies and skills to comprehend text that is read, heard, and viewed.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.01</td>
</tr>
<tr>
<td>2.02</td>
</tr>
<tr>
<td>2.03</td>
</tr>
<tr>
<td>2.04</td>
</tr>
<tr>
<td>2.05</td>
</tr>
<tr>
<td>2.06</td>
</tr>
<tr>
<td>2.07</td>
</tr>
<tr>
<td>2.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Competency Goal 3: The learner will make connections through the use of oral language, written language, and media and technology.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.01</td>
</tr>
<tr>
<td>3.02</td>
</tr>
<tr>
<td>3.03</td>
</tr>
<tr>
<td>3.04</td>
</tr>
<tr>
<td>3.05</td>
</tr>
<tr>
<td>3.06</td>
</tr>
</tbody>
</table>
Early Predictors of Reading Outcome

In order to efficiently provide children with services targeted at increasing their reading skills it is important to determine which children are most in need and in what skill area(s). According to La Paro and Pianta (2000) individual differences in a child’s academic, cognitive, social, and behavioral development can account for a small to moderate amount of variance in reading outcome measures. In their meta-analysis of 70 studies from 1985 to 1999 examining the extent to which academic, cognitive, social, and behavioral variables measured in preschool or kindergarten predicted academic outcomes measured in kindergarten, first, and second grade, correlations varied widely, from .08 to .72, with moderate effect sizes ranging from $r = .49$ to .51. What should be gleaned from this study is that the predictive contributions of variables depend on the age at which the variable is assessed and the outcome measure itself.

Some variables predict certain measures of reading better than others. For example, Catts et al. (1999) noted most studies supporting the predictive value of phonological awareness used word recognition as an outcome variable rather than reading comprehension. While using word recognition as an outcome measure makes sense when attempting to evaluate reading in younger children, it is not as useful an outcome measure when assessing more advanced reading skills. As the task of reading transitions from primarily single word recognition to sentence reading and on to passage comprehension, skills other than phonological awareness gain in importance. (Mason, 1992; Nation & Snowling, 1998; Whitehurst, 1997). There is a significant body of research demonstrating the predictive value of oral reading fluency for reading comprehension in young children as measured by state-mandated assessments (Hintze & Silberglitt, 2005; Silberglitt & Hintze, 2005; Stage &
Dacobsen, 2001), and in African American students (Hintze, Callahan III, Matthews, Williams, & Tobin, 2002; Kranzler, Miller, & Jordan, 1999). This makes sense as most instruments used to measure reading skills in young children assess an individual’s reading fluency, indirectly if not directly.

More importantly, however, is the ability to measure skills that allow for fluent oral reading itself and to determine the extent to which they allow prediction of later reading achievement. It has been suggested, and to an extent demonstrated (e.g., Speece, Roth, Cooper, & De La Paz, 1999, Whitehurst & Lonigan, 1998), that early prediction of later reading success depends on identifying which skills important to the development of reading contribute at different stages in literacy development. Catts et al. (2001) investigated the predictive value of measures of rapid naming, mother’s education, phonological awareness, sentence imitation, and letter identification as measured in 604 kindergarteners. They found all five variables to uniquely predict reading outcomes in second grade. These constructs, however, are not always found to be as predictive when measured earlier or later in a child’s life (Catts et al., 2001). Of particular interest to this study are variables determined to be predictive of later reading achievement when measured between first and third grade.

Predicting reading success prior to first grade. In a meta-analysis of 61 research studies, Scarborough (2003) examined the extent to which 21 different predictor variables, as measured in kindergarten and grouped into three general classifications, predicted later reading scores. Measures of nonverbal abilities, such as motor skills, visual memory, and visual discrimination, had the lowest overall correlations. These ranged from .16 for visual-motor integration to .31 for visual memory. Somewhat more predictive of future reading scores were measures of oral language proficiency such as phonological awareness, verbal
IQ, and receptive vocabulary. Correlations for these measures ranged from $r = .22$ for measures of speech perception to $r = .46$ for measures of phonological awareness. Most predictive were three variables which described an individual’s ability to process printed words. These three variables, letter-sound knowledge, letter identification, and print concepts, had the highest correlation to future reading scores, $r = .57$, .52, and .46, respectively.

Scarborough pointed out, however, that these correlations of variables as measured in kindergarten to later reading scores, the highest of which was $r = .57$, are relatively low when compared to the correlations between first and second grade variables to reading scores attained one, two, three, or four years later, at $r = .75$. It is only when individual variables are combined and multiple correlations with reading outcome measures are calculated that these variables become strong predictors, having an average $r$ of .75. Even if they do not strongly predict future reading when taken individually, when taken together their prediction of performance on future reading measures provides support for the theory that reading itself is comprised of many skills that interact and form strands that contribute to future literacy acquisition. This also suggests that first grade measures are more predictive of reading outcomes than later measures.

In a longitudinal study of 626 children followed from preschool through fourth grade Storch and Whitehurst (2002) demonstrated the predictability of first and second grade reading scores through measures of phonological awareness and print knowledge. In third and fourth grade, however, reading was better characterized as comprising two separate components similar to those identified in the Simple View of Reading: reading accuracy and comprehension, the latter best predicted by measures of oral proficiency rather than
phonological awareness (Schatschneider et al., 2004). In fact, several researchers have postulated emergent literacy skills contribute to reading outcome measures at different levels at different times (e.g., Speece, Roth, Cooper, & de la Paz, 1999; Whitehurst & Lonigan, 1998).

The question relevant to this study, then, is: what skills are most predictive in young children learning to read, and at what times? The literature suggests that the best time for predicting reading outcomes is during first grade, even if this is not the most ideal time for intervention. A review of the literature, too, informs the following discussion of reading skills identified as being the most predictive of reading outcomes during this time of a child’s development.

Specific Predictors of Reading Skills

**Familial risk of dyslexia.** Familial risk of dyslexia has been found to be predictive of future reading difficulties in young children. Puolakanaho et al. (2007) investigated this question in a study of 198 children measured at 3.5, 4.5, and 5.5 years of age, and again at the end of second grade. Measures of letter knowledge, phonological awareness, and rapid automatized naming were found to have a prediction probability above .80 when familial risk of dyslexia was also taken into account. Measures of expressive vocabulary and pseudoword repetition were not found to significantly contribute to predictive probability. This study provides evidence that prediction of future reading skills is possible before a child starts kindergarten. It also provides additional support for the well documented (Scarborough, 1990) phenomenon of familial risk factors for reading difficulties. Lyytinen and Lyytinen (2004) documented this fact as well in a longitudinal study of 200 Finnish children, 107 of whom were at familial risk for dyslexia. Though there were no group differences found on
measures of vocabulary and morphological skills at age two, subsequent measures at age 2.5, 3.5, and 5 indicated increased likelihood of reading difficulties for children both at familial risk and with low morphological skills. These results are applicable to reading studies in the United States as Finnish is even more reliant on phonetics than English (van der Hulst & van der Weijer, 1995). Both of these studies established this link in children who had yet to enter first grade, though not after. Though examination of this variable might be useful, it will not be included as part of the analysis as it is not part of the data set.

**Concepts of Print.** In an analysis of expert opinion and current research recommendations on methods to more rapidly increase reading skills in struggling readers, Reutzel and Smith (2004) stated that it is important to teach children concepts of print, also known as print conventions. A child who has a well developed concept of print is able to identify written words as being units of information, and is able to understand how to orient a book or page to correctly read it. These readers have an understanding that written English progresses from left to right and top to bottom, continuing on the next line or page after the end of the current line or page is reached. An understanding of concepts of print generally reflects familiarity with printed materials, though not necessarily an ability to read from them. Having a familiarity with concepts of print has been cited as important to learning to read as far back as the 1960s (Bond & Dykstra, 1967).

Evidence for the importance of concepts of print to learning to read has begun to emerge over the past 10 to 15 years. Scarborough (1998) found it to be an important kindergarten predictor of early elementary reading skills (median $r = .49$), second only to letter identification (median $r = .53$). Storch and Whitehurst (2002) found both print knowledge and phonemic awareness, as measured in pre-k and kindergarten, to be the best
predictors of grades one and two reading achievement. This did not hold up, however, for
reading achievement in grades three and four. In those grades the task of reading begins to be
less focused on decoding and more focused on comprehension. This evidence concurs with
the findings of the National Early Literacy Panel (2008), which stated that measures of such
skills as concepts of print and phonological awareness are less predictive of reading
outcomes as measured in later grades than in earlier grades. A hypothesized reason for this
decrease in predictive value is that children in third grade likely look very similar on these
measures. These skills are relatively easy skills to master, and by that point in their education
they have all developed the skill to a point of proficiency. That the population of third
graders is homogenous on these measures, almost by definition, makes such measures less
predictive. The concept of print variable will not be included as part of the analysis as it is
not part of the data set.

**Phonological processing.** Linguistic awareness, the ability to “discriminate units of
language (e.g., phonemes, words, propositions)” (Whitehurst & Lonigan, 1998), plays a large
part in the acquisition of reading skills. Whitehurst & Lonigan (1998) described the
hierarchical nature of linguistic awareness as well, describing syllabic sensitivity as an
important aspect of learning to discriminate these units of language (e.g., Liberman,
Shankweiler, Fischer, & Carter, 1974). This hierarchical unfolding also holds for a beginning
reader’s awareness of rhyme and intrasyllabic units (e.g., MacLean, Byrant, & Bradley,
1987). Despite this hierarchical nature, there is substantial evidence children learn to decode
words quicker if they are better at detecting parts of words, such as syllables, rhymes, and
phonemes, even after controlling for other contributing factors such as intelligence, receptive
vocabulary, memory abilities, and socioeconomic status. (e.g., MacLean, Byrant, & Bradley,
All of these skills are parts of the concept of phonological processing.

The smallest units of sound which make up spoken language are called phonemes (National Reading Panel, 2000). Stanovitch (1988) described phonological awareness as the explicit awareness of the sounds of speech within words, independent of their actual meaning. It is a unitary construct that varies on a continuum of complexity from preschool through early grade school (Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). It is the awareness that “words are composed of sequences of meaningless and somewhat distinct sounds” (Juel, 1988), and its acquisition is not a natural process (Lyon, 1998), but one that takes time and practice. Whitehurst and Lonigan (1998) and Sprenger-Charolles, Siegel, Béchennec, and Serniclaes (2002) cited phonological awareness as being important to the development of emergent literacy in elementary school children but of less importance prior to first grade. In fact, in some studies (e.g., Bryant et al., 1990) phonological awareness is the only variable investigated and was found to significantly predict reading acquisition. Phonological processing is often cited as being of crucial importance to the development of reading skills as a child enters elementary school.

There are many studies showing children with deficits in phonological processing in early elementary school have problems with reading (e.g., Fletcher et al., 1994; Fox & Routh, 1980; Olson, Wise, Conners, Rack, & Fulker, 1989, O’Connor, Bocian, Beebe-Frankenberger, & Linklater, 2008). In fact, tests of phonological processing often have been found to correlate more highly with early reading acquisition than more general tests of reading readiness or general intelligence (Stanovich, 1986), and a lack of phonological
processing and a subsequent inability to read or spell is a highly researched and replicated phenomenon (Lundberg, 2009).

In a 1988 study, Juel administered several measures designed to assess a child’s ability in several aspects theorized to be important to early reading: phonemic awareness, word decoding, word recognition, and listening comprehension. Children were given a test of reading comprehension at the beginning of first grade and at the beginning of fourth grade as outcome measures. Their findings were as follows: the probability a child would remain a poor reader in fourth grade, having been determined to be a poor reader in first grade, was .88. The probability a child would remain an average or better reader in fourth grade having been determined to be an average or better reader in first grade was .87. The largest and most significant factor separating average or better readers from poor readers was phonological awareness and word decoding skills. Better than average readers possessed better phonological awareness and word decoding skills by the end of second grade than poor readers at the end of fourth grade.

In their investigation into the predictive contribution of eight different variables to later reading achievement, Schatschneider, Fletcher, Francis, Carlson, and Foorman (2004) measured several constructs theorized to contribute to reading skills development in 540 kindergartners and examined the extent to which these constructs predicted reading outcomes in first and second grade. They found phonological awareness to be one of the biggest three predictors of reading outcomes. Knowledge of the importance of phonemic awareness also has contributed to improved outcomes in intervention. For example, in a study by Foorman, Fletcher, Francis, Schatschneider, and Mehta (1998), 285 children in first and second grade were provided with different reading instruction interventions. Children provided with direct
instruction in letter-sound correspondence improved in word reading skill at a significantly faster rate than those provided with indirect instruction consisting of exposure to literature. This variable will be included as part of the analysis.

**Orthographic processing.** According to McCardle, Scarborough, and Catts (2001), the first step of learning to read is to recognize that spoken words are themselves composed of discernable sounds. This is referred to as phonological awareness. The next step is to understand the fact that these sounds can be represented by the use of letters. This is the alphabetic principle, and the application of the alphabetic principal to written words is orthographic processing. It is only after these aspects of printed language are understood that individuals can apply their knowledge of phonemes to aid in the comprehension of text.

Most poor readers have significant problems learning to decode words (Gough & Tunmer, 1986), which is only partially attributed to problems with phonological processing (e.g., Wagner & Torgesen, 1987). The act of decoding words requires awareness of the alphabetic principle: the concept that letters correspond to certain sounds. These sounds make up phonemes, which are the basic building block of spoken words. Orthographic processing is the ability to derive the sounds of language (phonemes) from written symbols (graphemes, Burt, 2006), and is an important part of learning how to read. This skill has been demonstrated to contribute to word reading above and beyond the contribution provided by phonological awareness (Burt, 2006).

There also is evidence that children with higher levels of knowledge of phoneme-grapheme correspondence exhibit higher levels of reading achievement (Hoover & Gough, 1990; Jorm, Share, MacLean, & Matthews, 1984; Tunmer et al., 1988; Vellutino et al., 1996). In the same longitudinal study that illustrated the importance of kindergarten
phonemic awareness to first and second grade reading outcome measures, Schatschneider, Fletcher, Francis, Carlson, and Foorman (2004) found letter-sound knowledge in kindergarten to account for unique variance in the same outcome measures. In a longitudinal examination of 54 elementary school students, Juel (1988) found poor readers grew on measures of orthographic processing at a much slower rate than good readers. Students who were deemed to be poor readers in first grade scored, on average, one full standard deviation below their adequate-reader counterparts on fourth grade measures of orthographic processing. This variable will be included as part of the analysis as it is part of the data set.

**Rapid automatized naming.** In 1972, Jansky and de Hirsch found letter naming in kindergarten to be the best single predictor of second grade reading success. Whitehurst and Lonigan (1998) also reported knowledge of letters as being important in the acquisition of reading skills, as a beginning reader who is unable to name the letters of the alphabet cannot learn to which sounds letters or combinations of letters are associated (Bond & Dykstra, 1967; Mason, 1980). Stevenson and Newman (1986) identified knowledge of the alphabet at entry into school as one of the strongest predictors of literacy success.

In a longitudinal study of 55 children, Scarborough (1998) found second grade measures of rapid automatic naming (RAN) of common objects to contribute substantially to prediction of grade two and grade eight reading outcomes of children determined to have reading disabilities at age eight. In a study of 255 children, results from letter naming measures administered to kindergartners were found to correlate more highly with fifth and tenth grade scores on reading achievement tests than any other measure, which included receptive vocabulary, auditory discrimination, spelling, auditory memory, and word recognition (Stevenson & Newman, 1986). Not only does letter knowledge influence the
acquisition of phonological sensitivity (e.g., Bowey, 1994; Johnston, Anderson, & Holligan, 1996; Stahl & Murray, 1994), but the rapidity of naming matters as well. Most studies examining tasks of speeded serial naming have provided evidence of reading deficits in children (Catts, Hogan & Fey, 2003; Wolf, 1984), and measures of RAN can explain variance in reading achievement above and beyond what can be explained by phonological awareness (Baidan, 1994; Catts; 1993).

In their study examining predictive contributions of various skills important to the acquisition of reading to later measures of reading achievement, Schatschneider et al. (2004) found speeded naming tasks to account for unique variance across reading outcomes in both grades one and two, though not in subsequent grades. In a longitudinal study of 226 Dutch children, Verhagen, Aarnoutse, and Van Leeuwe (2010), naming speed of digits, colors, and pictures, along with measures of phonological awareness, were significant predictors of word spelling accuracy and word recognition at the end of grades one and two. The results of this study are relevant to studies conducted with American children as Dutch is consistent with English with regards to its reliance on phonemes (Gussenhoven, 1992). In a study by McBride-Chang & Manis (1996), 125 readers in third and fourth grade were administered assessments of verbal intelligence, speeded naming of letters and numbers, and phonological awareness. For poor readers, phonological awareness and speeded naming both were significantly associated with an outcome measure of word reading, but not verbal intelligence. For good readers, only phonological awareness and verbal intelligence were associated with reading outcomes. This variable will be included as part of the analysis.

**Phonological memory.** An additional skill often cited as a predictor of later reading achievement is phonological memory, or an individual’s capacity to retain and repeat back
verbal information in memory. Children with reading disabilities generally perform poorly on these types of tasks (Mann, Liberman, & Shankweiler, 1980; Stone & Brady, 1995). In their review of emergent literacy research, Whitehurst and Lonigan (1998) highlighted the relationships established between phonological memory and vocabulary acquisition (Gathercole, Willis, Emslie, & Baddeley, 1992), phonological memory and general reading skills development (Gathercole, Willis, & Baddeley, 1991; Rohl & Pratt, 1995; Wagner et al., 1994), and difficulties with phonological memory and the presence of reading disabilities (Cohen & Netley, 1981).

In an examination of 477 children, Jorm, Share, Maclean, and Matthews (1984) assessed each child’s short term memory upon entry to kindergarten. The children’s reading ability and short term memory were then assessed after they finished first grade. Initial measures of short term memory were found to correlate significantly with both outcome measures. Short term memory also has been found to contribute to reading ability in older children as well. For example, in a longitudinal study of 102 children assessed at ages eight, nine, and eleven, Cain, Bryant and Oakhill (2004) obtained measures of reading ability, vocabulary and verbal skills, and performance on two auditory working memory assessments. At each time point the children were measured, auditory working memory was found to predict unique variance in reading comprehension, even after controlling for vocabulary and verbal ability. They concluded auditory working memory “should be regarded as one of several factors that can influence comprehension ability and comprehension development,” (p. 40) but that good working memory alone is not enough to allow for skilled reading. This variable will be included as part of the analysis.
Receptive vocabulary. Another factor commonly cited as being a contributor to the acquisition of reading skills is receptive vocabulary. Receptive vocabulary is the collection of words to which a child is able to assign meaning, and is distinct from expressive vocabulary, which is the collection of words a child is able to utilize when speaking or writing. Vocabulary is important in the process of literacy acquisition (Whitehurst & Lonigan, 1998). A small vocabulary limits the number of words available for immediate recognition, and increases processing load (Catts, Hogan & Fey, 2003). A larger vocabulary aids in recognition, and allows for more cognitive processing to be devoted to extracting meaning from recognized words. Several studies have found an association between reading disabilities in children and their ability on picture naming tasks (Catts, 1986; German, 1979; Wolf, 1984).

A child’s vocabulary becomes more important to his or her reading ability as they begin to engage in more complex forms of reading. Research by Mason (1992), Snow, Barnes, Chandler, Hemphill, & Goodman (1991), and Whitehurst (1996) showed that semantic (and syntactic) abilities become more important as a child’s reading development shifts away from sounding out words towards deriving meaning from passages.

Receptive vocabulary as a predictor of future reading ability begins to play a more important role as a child’s reading skills improve. Schatschneider, Fletcher, Francis, Carlson, and Foorman (2004) found oral language (of which receptive vocabulary is a part), as measured in kindergarten, to be poor predictors of reading outcomes measured in grades one and two. These measures, however, were more phonologically oriented and less related to extracted meaning. Storch and Whitehurst (2002) found measures of oral language only related to reading comprehension in grades three and four; reading activities of a more basic
nature, that are assessed prior to these grades, rely less heavily on oral language skills. They concluded, “Although oral language abilities do not appear to make a direct contribution to reading during grades one and two, a child’s skill with spoken language does play an essential, albeit an indirect role in reading achievement during the early stages of reading acquisition” (p. 943).

The importance of receptive vocabulary begins to be revealed when older children are the focus of investigations. Ricketts, Nation, and Bishop (2007) investigated the relationship between vocabulary and literacy in 81 children aged eight to ten years. They found measures of vocabulary to predict reading comprehension and exception word reading (exception words being words such as “break,” rather than “brake,” that do not follow typical phonological rules). In a similar study by Nation and Snowling (2004), vocabulary was found to positively contribute to reading ability in 72 children followed from ages 8.5 to 13. DeThorne, Petrill, Schatschneider, and Cutting (2010) found predictive evidence of formally measured vocabulary to early reading ability in a study of 380 twins across a three year period (mean age of participant 7.13 years, SD = .65). Of note, however, was mention in the study that additional measures of conversational language (systematic decoding of 15-minute conversations between researcher and child, accounting for total number of words, total number of different words, total number of conjunctions) contributed to predictions of reading ability above and beyond that predicted by measures of vocabulary. This variable will be included as part of the analysis.

The preceding seven skills have all been found to be effective predictors of reading ability. Risk of dyslexia and concepts of print are both important predictors early in a child’s development, but fail to hold up in importance as a child enters first grade. Phonological
awareness, orthographic processing, rapid automatized naming, phonological memory, and receptive vocabulary have been found to be consistent predictors of reading ability through early elementary school, and so will be investigated in this study.

**Summary**

A significant amount of research has been conducted on the development of reading. Models for literacy acquisition have become more complex and more complete in recent years. This has allowed for robust conceptualization of the relative contributions of different skills to the learning of reading at different levels of mastery. Understanding how reading as a skill develops provides educators with more targeted interventions and more accurate tools for measuring reading outcomes.

One such tool utilized in North Carolina is the third grade EOG Test of Reading Comprehension. Though this test was developed in accordance with guidelines set forth by the North Carolina Board of Education in an attempt to create a psychometrically sound instrument, there exists only a moderate amount of evidence for its appropriateness as such a tool. It is a reliable measure but it is not clear that it is a valid measure. Additional evidence in support of the third grade EOG Test of Reading Comprehension’s validity is needed. Such support for its validity could come from a study examining the extent to which performance on the third grade EOG Test of Reading Comprehension is predicted by skills identified by the literature as being important to the acquisition of reading. These skills have all shown to be reliable predictors of reading outcomes. While there are earlier skills shown to be predictive, those measured in first, second, and third grade are generally stronger predictors and as such are more relevant to assessing the validity of an outcome measure.
This study attempts to provide such evidence by examining the predictive contribution of measurements of five research-identified variables to scores obtained on the third grade EOG Test of Reading Comprehension. Testing results from this instrument are already being used to make decisions on student promotion, teacher efficacy, and district-wide resource allocation. If these variables do predict subsequent scores on the third grade EOG Test of Reading Comprehension, it will provide additional evidence in support of the use of this instrument for such decisions.
CHAPTER III

METHODS

Participants

The current study examined the relative contributions of educational and neuropsychological predictors to scores on the North Carolina third grade End-of-Grade (EOG) Test of Reading Comprehension. Each of the five variable’s predictive value was assessed at three different time points: first grade, second grade, and third grade. In addition, the predictive contribution of the change in each variable across time points was assessed. The sample used in this study was obtained from a larger sample recruited in accordance with the University of North Carolina Institutional Review Board procedures.

The sample was recruited from a single rural school district in North Carolina to participate in a study examining the effects of a writing intervention. Children were recruited in two cohorts and each cohort was followed for at least three years. Cohort one was established through screenings that took place in the Fall of 2006. Cohort two was established through screenings that took place in the Fall of 2007. These screenings consisted of administering the Wechsler Individual Achievement Test-II (WAIS-II; The Psychological Corporation, 2002) Written Expression subtest to every first grade student in the school district. The sample consisted of two groups: an at-risk (AR) group (n = 43) and a typically developing (TD) group (n = 68). Children were classified as either AR or TD based on their score on the Written Expression subtest of the WIAT-II Written Expression Subtest. See Table 3.1 for demographic information of the sample. All of the students included in the final
sample met criteria defined as necessary by the primary study: all children were primarily placed in a regular education setting, all children had completed kindergarten, and all children spoke English as a primary language. In total, data from 111 children were used for this analysis.

Table 3.1
Demographic Information

<table>
<thead>
<tr>
<th>Category</th>
<th>Total sample (n = 111)</th>
<th>Typically Developing (n = 68)</th>
<th>At Risk (n = 43)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age in months</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1*</td>
<td>82.30 (5.42)</td>
<td>81.28 (4.56)</td>
<td>83.91 (6.29)</td>
</tr>
<tr>
<td>Year 2*</td>
<td>94.40 (5.44)</td>
<td>93.34 (4.60)</td>
<td>96.06 (6.24)</td>
</tr>
<tr>
<td>Year 3*</td>
<td>106.79 (5.50)</td>
<td>105.40 (4.56)</td>
<td>107.98 (6.50)</td>
</tr>
<tr>
<td><strong>IQ</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1*</td>
<td>99.51 (13.66)</td>
<td>102.91 (12.77)</td>
<td>94.14 (13.42)</td>
</tr>
<tr>
<td>Year 2**</td>
<td>100.04 (15.16)</td>
<td>103.97 (14.18)</td>
<td>93.81 (14.72)</td>
</tr>
<tr>
<td>Year 3**</td>
<td>106.65 (15.00)</td>
<td>111.01 (11.17)</td>
<td>99.60 (17.65)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>61</td>
<td>35</td>
<td>26</td>
</tr>
<tr>
<td>Female</td>
<td>50</td>
<td>33</td>
<td>17</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian American</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>African American</td>
<td>20</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Native American</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Multi-racial</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Caucasian</td>
<td>85</td>
<td>54</td>
<td>31</td>
</tr>
<tr>
<td>Latino identified</td>
<td>12</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

*Note: Continuous variables presented as: Mean (Standard Deviation).
*denotes statistically significant differences between TD and AR groups at $p < .05$
**denotes statistically significant differences between TD and AR groups at $p < .001$

**Criterion Measure**

All students were administered the North Carolina third grade EOG Test of Reading Comprehension. This test is used by the state of North Carolina to “know the extent to which (students) have mastered expected knowledge and skills and how they compare to others.” (Office of Curriculum and School Reform Services, 2004, p. 10). Scores from the third grade EOG Test of Reading Comprehension are reported in four different ways: as a scaled score, as a percentile rank, as one of four achievement levels, and as pass/fail. The scale scores are
the result of a transformation of raw scores making use of a program developed at the L.L. Thurstone Psychometric Laboratory at the University of North Carolina at Chapel Hill (Office of Curriculum and School Reform Services, 2004). Scale scores for the third grade EOG Test of Reading Comprehension have a mean of 338.65 and a standard deviation of 12.57. Analyses in this study of the third grade EOG Test of Reading Comprehension was conducted using scaled scores.

When reporting the results of EOG tests, achievement levels are also used. They are calculated from scale scores, as shown in Table 1.1. Achievement levels range from one to four. Students scoring a one or a two are deemed to have failed the test and students scoring a three or a four are deemed to have passed the test. Instructional information can be found within these levels of achievement. Within the passing range, an achievement level of four indicates superior skills beyond what is expected, compared to the expected mastery denoted by an achievement level of three. Though both achievement levels one and two are considered failing, level two denotes minimal preparation for success in subsequent grades, whereas level one indicates insufficient preparation. Though scale scores were used as the outcome measure for this study, relating scaled scores back to achievement levels provided additional information with regards to the predictive value of the selected variables.

The third grade EOG Test of Reading Comprehension was chosen as the outcome measure for this study as it is mandated through North Carolina state policy as the outcome measure of interest for making policy decisions. The results of EOGs are used to make significant decisions about student learning, teacher effectiveness, district policy, and resource allocation. Irrespective of whether the instrument is the most accurate outcome
measure of reading achievement, it is arguably the outcome measure whose results have the widest ranging effects in the state of North Carolina.

The coefficient alpha for the third grade EOG Test of Reading Comprehension was reported to be .93. Two types of validity were described as being present in this instrument. Content validity was addressed through alignment with cognitive concepts set forth as important by the Office of Curriculum and School Reform Services (2004). Criterion validity was addressed through alignment with instructional practices as defined by teachers and through alignment with the NCSCS.

**Predictor Variables**

**Phonological awareness measures.** During cohort one’s first year of assessment, each student was administered subtests from the Comprehensive Test Of Phonological Processing (CTOPP, Wagner, Torgesen, & Rashotte, 1999) The CTOPP is a widely used test designed to assess various aspects of phonological awareness, phonological memory, and rapid naming in children ages five and older. The Elision subtest of the CTOPP was administered in year one of the study to assess phonological awareness. During this 5-minute task students are asked to repeat a word and then repeat the word with a part omitted. Raw scores are transformed into a scaled score using age-based normative tables. Coefficient alphas reported for the Elision subtest range from .92 at age 6 to .89 at age 9 (Wagner et al. 1999). Concurrent validity for the CTOPP Elision subtest includes a correlation of .73 and .74 with the Woodcock Reading Mastery Tests, Revised (WRMT-R) Word Identification subtest and Word Analysis subtest, respectively, and of .67 and .68 with the Test of Word Reading Efficiency (TOWRE) Sight Word Efficiency and Phonetic Decoding Efficiency, respectively (Wagner et al., 1999)
After its use with cohort one in year one use of the CTOPP was discontinued at the request of the school district in which the study was being conducted. Beginning with the first year assessment of cohort two the Process Assessment of the Learner (PAL, Berninger, 2001) Syllables and Phonemes subtests were used in place of the CTOPP Elision subtest to assess phonological processing. The PAL is a standardized measure designed to assess processes related to reading and writing in students in grades kindergarten to six. The instrument generates grade-based scale scores. During the Syllables and Phonemes subtests children are asked to demonstrate phonological awareness of syllables and phonemes by identifying syllables and sounds in words. Examinees are asked to repeat a word with a syllable or sound omitted or to identify the remaining phonemes when one phoneme has been omitted. Raw scores are transformed to decile scores using grade-based normative tables. Coefficient alphas were reported to range from .81 in grade one to .73 in grade three for the Syllables subtest, and from .84 in grade one to .81 in grade three for the Phonemes subtest. Though no concurrent validity statistics were reported, construct validity was addressed through intercorrelation between the two subtests, ranging from .69 in first grade to .56 in third grade, and through face and content validity through the use of empirical item analysis and expert judges (Berninger, 2007). For the third assessment of the second cohort the PAL was replaced with the Process Assessment of the Learner, Second Edition (PAL-II, Berninger, 2007). Though the administration of subtests was the same, scoring was changed from deciles to scale scores. For the normative groups used in both PAL instruments means and standard deviations are available.

**Orthographic processing measures.** Two measures were used to estimate a child’s orthographic processing ability: The Wechsler Individual Achievement Test-Second Edition
The WIAT-II is a standardized measure of academic and problem-solving skills that measures various aspects of reading, writing, and arithmetic. Normative data can be calculated based on age or grade; for this analysis all WIAT-II subtest scores were calculated based on age. This subtest included items designed to provide estimates of knowledge of single letter sounds, letter blends, and single words. Raw scores were transformed to scaled scores based on age-based normative tables. Reliability for the WIAT-II Spelling subtest, using the Spearman-Brown split-half formula, ranges from .93 at age 5, to .94 at age 9 (Wechsler, 2002). Concurrent validity for the WIAT-II Spelling subtest includes a correlation of .86 with the Wechsler Individual Achievement Test Basic Reading subtest, .79 with the Process Assessment of the Learner, Reading and Writing (PAL-RW) Word Choice subtest, and .78 and .76 with the Wide Range Achievement Test - 3rd edition Spelling subtest and Written Language composite, respectively. (Wechsler, 2002). Age-based scale scores were used for analysis.

The PAL Word Choice subtest was designed to represent a direct challenge to an individual’s orthographic processing skill as well. Scores are calculated based on grade and reported as standard scores. Reliability for the PAL Word Choice subtests is .83 in first grade, .88 in second grade, and .58 in third grade (Berninger, 2007). Concurrent validity for the Word Choice subtest includes a correlation of .34 with the Differential Abilities Scale, Second Edition Verbal Cluster, .35 with the NEPSY-II Phonological Processing subtest, .39 with the NEPSY-II Word List Interference subtest (Berninger, 2007), and .80 with the WIAT-II Pseudoword Decoding subtest.
**Rapid automatized naming measure.** The Process Assessment of the Learner (PAL) Rapid Automatized Naming (RAN) Letters subtest requires children to rapidly name visually presented letters and letter groups. Times for letter naming and letter group naming were added together, and decile scores were generated using grade-based normative tables. Coefficient alphas were reported to be .79 for RAN Letters for grades 1 through 3, and .85 for RAN Letter Groups for grades 1 through 3. Content validity for this measure was established through the use of empirical item analysis and expert judges. Test-retest correlations for RAN Letters, grades K through three, were reported to be .79 for letters and .85 for letter groups. Though no concurrent validity statistics were reported, face and content validity were addressed through the use of empirical item analysis and expert judges (Berninger, 2007).

For the third assessment of the second cohort, the PAL was replaced with the Process Assessment of the Learner, Second Edition (PAL-II, Berninger, 2007). Though the administration of subtests was the same, scoring was changed from deciles to scale scores. Additionally, rather than summing the total time for the letter and letter group portions of the subtest, scale scores were calculated individually. From these two scale scores, a combined score was calculated and these two scores were averaged to provide a single score representing the children’s performance on this subtest. For the normative groups used in both PAL instruments, means and standard deviations are available.

**Phonological memory measures.** The Comprehensive Test of Phonological Processing (CTOPP; Wagner, Togesen, & Rashotte, 1999) Nonword Repetition subtest measures phonological memory. Participants are played fictitious words from a recording and asked to repeat them. Scaled scores are then generated from raw scores according to the
published age-based normative tables. Raw scores are transformed into a scaled score using age-based normative tables. Coefficient alphas reported for the Nonword Repetition subtest range from .80 at age 6 to .79 at age 9 (Wagner et al. 1999). During the development of the test, content validity was evaluated using item discrimination and difficulty statistics. Confirmatory factor analysis of the CTOPP shows the Nonword Repetition subtest loads on the same factor as the Memory for Digits subtest and provides evidence for the validity of using the Nonword Repetition subtest as a measure of auditory memory (Wagner et al., 1999).

After cohort one was assessed in first grade the school district requested the CTOPP Nonword Repetition measure be removed from the battery due to potential practice effects as it was a common measure used by school psychologists within the district. As evidence exists demonstrating the similarities between the two tasks (Wagner et al., 1999), the CTOPP Nonword Repetition measure was replaced by the Wechsler Intelligence Scale for Children-IV Processing Instrument (WISC-IV-PI, Wechsler, 2004) Digit Span Backwards and Digit Span Forward subtest. The WISC-IV-PI is a companion measure to the WISC-IV, which provides an appraisal of a child’s intelligence. It provides age-based scaled scores and is adequately normed. The WISC-IV-PI offers different options for administration of the WISC-IV tasks to assist in examining a child’s information processing in more detail. The two subtests used in this study, the Digit Span Forward and Digit Span Backward, have reported internal consistency reliability coefficients ranging from $r = .68$ (Digit Span backward, age 8) to $r = .88$ (Digit Span forward, age 9). Validity evidence is reported for this measure, including content, construct, and criterion. A study comparing working memory measures (including Digit Span) of the WISC-IV-I and the Clinical Evaluation of Language
Fundamentals, 4th edition (CELF-4) found a correlation of .69 (WISC-IV-I; Wechsler, Kaplan, Fein, Kramer, Morris, Delis, & Maerlender, 2004). The WISC-IV-I is the name of a previous edition of the WISC-IV-PI. The Digit Span Forward and Digit Span Backward subtests are identical in each edition.

**Receptive vocabulary measures.** The Peabody Picture Vocabulary Test – Forth Edition (PPVT-IV; Dunn & Dunn, 2007) was administered to assess receptive vocabulary skills. Children are shown a sheet with four pictures arranged in a grid and asked to point to the picture matching the word the examiner says. Standard scores were generated from the raw scores according to published normative tables. Coefficient alpha reliability for the PPVT-IV is .97 for all age ranges involved in this study and split-half reliability ranges from .97 for ages 6:0 to 6:5 to .94 for examinees age 9. (Dunn & Dunn, 2007). Convergent evidence of construct validity was reported through the results of correlation studies with other tests of expressive vocabulary, language ability, and reading achievement. The correlation of scores on the PPVT-4 with the Expressive Vocabulary Test, 2nd edition (EVT-2) is .84 for ages 5-6 and .80 for ages 7-10, with. The correlation between the PPVT-4 and the Clinical Evaluation of Language Fundamentals, 4th edition (CELF-4) is .67 for receptive language in the age range 5-8 (Dunn & Dunn, 2007).

The Comprehensive Receptive and Expressive Vocabulary Test – Second Edition (CREVT-2; Wallace & Hammill, 2002) Receptive Vocabulary subtest was added to the assessment battery in the second year of the study to replace the PPVT-IV at the request of the school district in which the original study took place. The CREVT-2 Receptive Vocabulary subtest was chosen to replace the PPVT-IV as it measures receptive vocabulary in a similar manner: participants are presented with pages of six pictures, with each page
representing different categories (i.e., vehicles, careers) and asked to point to the picture that matches words orally presented by the examiner. Coefficient alphas were reported to range from .88 at age six to .91 at age 9. Evidence for concurrent validity exists in the form of correlations between scores on the CREVT-2 Receptive Vocabulary subtest with scores on the PPVT-IV of .59, with scores on the WISC-III Vocabulary subtest of .66, and with the CELF-R of .74. (Wallace & Hammill, 2002).

In all, five variables were identified through a review of the research literature as being predictive of reading outcome measures in early elementary school. One variable was phonological awareness which was measured through the CTOPP Elision subtest the PAL Syllables and Phonemes subtests, and the PAL-II Syllables and Phoneme subtests. Another identified variable was orthographic processing which was measured through the WIAT-II spelling subtest. Rapid automatized naming was identified as being important, and was measured through the PAL Letters and PAL-II Letters subtests. Phonological memory was also identified as being predictive and was measured by the CTOPP Nonword Repetition and WISC-IV-PI Digit Span Forward and Digit Span Backward subtests. Finally, receptive vocabulary has been shown to be predictive of reading outcomes and was measured by the PPVT-IV and CREVT-2 Receptive Vocabulary subtest.

**Procedures**

A data set from a previous study was used for this analysis. This initial data set included approximately 130 students from a single North Carolina school district recruited over two years in two cohorts. All students were assessed during their first, second, and third grade year with a battery of neuropsychological assessments including the aforementioned subtests. Annual assessments subsequent to the initial assessment were conducted as close as
possible to the initial assessment date, generally within two weeks in either direction. The assessment battery was split into two blocks of measures with each child being administered the blocks in a counterbalanced fashion from year to year in order to control for order effects. Each instrument in the assessment battery was administered, scored, and standardized in accordance with published test administration instructions. Each battery was second-scored by a researcher or graduate student other than the initial battery administrator. The Franklin Porter Graham Child Development Institute (FPG) Data Management and Analysis Center double-entered the raw scores and standardized the scores according to each measure’s published norms.

Relevant measures in block A in the first year were the WIAT-II Spelling subtest, the PAL RAN Letters subtest, and the PPVT-IV. Block B included the CTOPP Elision and Nonword Repetition subtests. After the first year of data collection, The CTOPP Elision subtest was replaced by the PAL Syllable and Phoneme subtests, the PPVT-IV was replaced by the CREVT-2 Receptive Vocabulary subtest, and the CTOPP Nonword Repetition subtest was replaced by the WISC-IV-PI Digit Span Forward and Digit Span Backward subtest. The battery was also changed for the third year of assessments. During the third year of assessments relevant measures in block A were the WIAT-II Spelling subtest and the CREVT-2 Receptive Vocabulary subtest. Relevant measures in block B were the WISC-IV-PI Digit Span Forward and Backward subtest, and the PAL RAN Letters, Syllables, and Phonemes subtests. During cohort two’s third assessment, the PAL-II was substituted for the PAL at the request of the test developer. Reliability and validity information for these subtests can be found in Table 3.2.
<table>
<thead>
<tr>
<th>Test</th>
<th>Reliability</th>
<th>Summary of Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Carolina 3rd Grade Reading End-of-Grade Test</td>
<td>$\alpha = .93$</td>
<td>Item discrimination and difficulty statistics, item analysis through Mantel-Haenszel differential item functioning, alignment with test content guidelines, correlation between test results and teacher judgment of achievement ($r = .63$) and with expected course grade ($r = .67$), item construction guided by alignment to North Carolina Standard Course of Study for 3rd Grade English Language Arts</td>
</tr>
<tr>
<td>Comprehensive Test of Phonological Processing (CTOPP) Elision</td>
<td>age 6 $r = 0.92$, age 7 $r = 0.91$, age 8 $r = 0.89$</td>
<td>Item discrimination and difficulty statistics, parameters in Item Response Theory models, correlations with other tests of language ($0.74$ with WRMT-R Word Analysis, $0.73$ with WRMT-R Word Identification, $0.67$ with TOWRE Sight Word Efficiency, $0.68$ with TOWRE Phonetic Decoding)</td>
</tr>
<tr>
<td>Process Assessment of the Learner (PAL) Syllables &amp; Phonemes</td>
<td>syllable $r = 0.80$, phoneme $r = 0.92$</td>
<td>Expert judges, empirical item analysis, correlations with other tests (Syllables correlated $0.49$ with WIAT-II Pseudoword Decoding, Phonemes correlated $0.56$ with WIAT-II Pseudoword Decoding)</td>
</tr>
<tr>
<td>Wechsler Individual Achievement Test – Second Edition (WIAT-II) Spelling</td>
<td>$r = 0.91$</td>
<td>Expert judges, Item Response Theory methods (item-total correlations $&lt;20$ removed), correlations with other tests of achievement ($0.73$ with WRAT3 Reading, $0.78$ with WRAT3 Spelling)</td>
</tr>
<tr>
<td>Process Assessment of the Learner (PAL) Word Choice</td>
<td>grade one $r = 0.83$, grade two $r = 0.88$, grade 3 $r = 0.58$</td>
<td>Expert judges, empirical item analysis, correlations with other tests ($0.80$ with WIAT-II Pseudoword Decoding)</td>
</tr>
<tr>
<td>Process Assessment of the Learner (PAL) Rapid Automatized Naming</td>
<td>letters $r = 0.92$, digits $r = 0.84$</td>
<td>Expert judges, empirical item analysis, correlations with other tests ($-0.72$ with WIAT-II Pseudoword Decoding)</td>
</tr>
<tr>
<td>Comprehensive Test of Phonological Processing (CTOPP) Nonword Repetition</td>
<td>age 6 $\alpha = 0.80$, age 7 $\alpha = 0.80$, age 8 $\alpha = 0.80$</td>
<td>Item discrimination and difficulty statistics, parameters in Item Response Theory models, confirmatory factor analysis of CTOPP shows Nonword Repetition loads on the Phonological Memory factor</td>
</tr>
</tbody>
</table>
Table 3.2 continued

<table>
<thead>
<tr>
<th>Test</th>
<th>Reliability</th>
<th>Summary of Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CTOPP) Elision</td>
<td>age 6 $\alpha = 0.92$, age 9 $\alpha = 0.89$</td>
<td>Concurrent validity with Woodcock Reading Mastery Test, Revised Word Identification subtest ($r = .73$) and Word Analysis subtest ($r = .74$), and with Test of Word Reading Efficiency Sight Word Efficiency subtest ($r = .67$) and Phonetic Decoding Efficiency ($r = .68$)</td>
</tr>
<tr>
<td>Wechsler Intelligence Scale for Children-IV-Integrated (WISC-IV-I) Digit Span</td>
<td>age 6 DSpF $r = 0.83$, DSpB $r = 0.83$, age 7 DSpF $r = 0.79$, DSpB $r = 0.69$, age 8 DSpF $r = 0.82$, DSpB $r = 0.68$</td>
<td>Research studies, review of theoretical literature, expert reviews, correlations with other tests (Working Memory scale including Digit Span correlated 0.69 with CELF-4)</td>
</tr>
<tr>
<td>Peabody Picture Vocabulary Test-4 (PPVT-4)</td>
<td>age 6:0-6:5 $\alpha = 0.97$, age 6:6-6:11 $\alpha = 0.94$, age 7 $\alpha = 0.94$, age 8 $\alpha = 0.99$</td>
<td>Correlations with other tests of language (0.84 with EVT-2 for ages 5-6, 0.80 with EVT-2 for ages 7-10, 0.67 with CELF-4 for ages 5-8)</td>
</tr>
<tr>
<td>The Comprehensive Receptive and Expressive Vocabulary Test – Second Edition (CREVT-2)</td>
<td>age 6 $\alpha = .88$, age 7 $\alpha = .91$, age 8 $\alpha = .91$</td>
<td>Correlations with other tests of language (0.59 with PPVT-IV, 0.66 with WISC-III Vocabulary, 0.74 with CELF-R)</td>
</tr>
</tbody>
</table>

**Data Analysis**

**Preliminary data analysis.** Each child was measured on the predictor variables at three intervals; once in first grade, once in second grade, and once in third grade. The difference in mean age in months between assessment one and two was 12.098 and the difference in mean age in months between assessment two and three was 12.016 indicating these intervals were essentially evenly spaced. All predictor variable scores were then transformed into a standardized metric. Additionally, each participant’s scores on the North Carolina third grade EOG Test of Reading Comprehension were collected and used as the outcome criterion. The scores from the North Carolina third grade EOG Test of Reading
Comprehension were reported in several formats, one of which took the form of scaled scores (mean = 338.65, standard deviation = 12.57). Scaled scores were used as the outcome measure.

Preliminary analyses examined several demographic variables in order to determine if they were significantly related to the outcome measure. The variables examined were At Risk vs. Typically Developing status (AR vs. TD), IQ, gender, and race classification. Any of these variables found to be significantly related to the outcome measure were considered to be entered into the regression model as covariates. If no significant relationship were observed between these variables and the outcome measure they were not included in the analysis.

The initial data set included 205 children. Of these 205 children sixty-nine were classified AR and received an intervention designed to improve their writing skills. Part of this writing skill intervention included explicit instruction on skills related to the predictor variables such as phonological awareness and orthographic processing. Children receiving the intervention were also instructed on other skills that might affect their outcome scores. A portion of this intervention instruction took place after assessment time points but prior to administration of the North Carolina third grade EOG Test of Reading Comprehension. These children were removed from the data set prior to analysis. Of the remaining 136 children twenty-five either did not have a North Carolina third grade EOG Test of Reading Comprehension score reported or had only a retest score reported. As it is not possible to determine to what extent these twenty-five students were provided intensive reading intervention between the date of the first EOG test and the EOG retest it was decided these
students should be removed from the data set as well. This left 111 children in the data set for final analysis.

**Research Questions, Hypotheses, and Data Analyses**

**Question one.** Do the variables identified through a review of the literature as being predictors of reading achievement in first, second, and third grade contribute to scores on the North Carolina third grade EOG Test of Reading Comprehension?

**Hypothesis one.** It was hypothesized that all five predictor variables, measured at all three ages, would significantly contribute to a model designed to predict scores on the North Carolina third grade EOG Test of Reading Comprehension. The predictor variables in question are measures of phonological awareness, orthographic processing, rapid automatized naming, phonological memory, and receptive vocabulary.

**Data analysis one.** The analysis was conducted utilizing multiple linear regression and was conducted for each of the three time points at which the predictor variables were measured. Scores on the North Carolina third grade EOG Test of Reading Comprehension were used as the criterion variable in all cases. Analysis was conducted with PASW, version 18.0.

The output for each predictor variable at each time point included standardized regression coefficients (β), the standard error for each regression coefficient, and a measure of statistical significance (p). The standardized regression coefficient for each predictor variable denoted the relative contribution of that predictor to the outcome measure after controlling for the contribution of the other four predictor variables and any covariates entered into the model. Standard error was the variability of the values for a particular predictor variable over repeated samplings. The measure of statistical significance indicated
whether or not a given predictor variable contributed significantly to the outcome measure in
the presence of other predictor variables. Each variable determined to be significant at the \( p < .05 \) level was judged to contribute significantly to the prediction of scores on the North Carolina third grade EOG Test of Reading Comprehension for that particular year.

Additionally, for each multiple linear regression analysis a multiple correlation coefficient, \( R \), was reported. As \( R \) is a correlation, \( R^2 \) represented the proportion of the variation in the outcome measure accounted for by the linear combination of the predictors. The output yielded a test of significance on this value as well, indicating whether or not the predictor variables together predicted the outcome measure at a better-than or worse-than chance level. Predictor-time point combinations found not to significantly contribute to outcome measure were considered for elimination from secondary analysis.

**Question two.** Does the growth on measures of these variables, identified through a review of the literature as being predictors of reading achievement in first, second, and third grade contribute to scores on the North Carolina third grade EOG Test of Reading Comprehension?

**Hypothesis two.** It was hypothesized that change over time in receptive vocabulary, orthographic processing, rapid automatized naming, phonological memory, and receptive vocabulary will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

**Data analysis two.** When this study was initially proposed, question two was to be addressed through latent growth curve (LGC) modeling. It was proposed that five growth curves would be created through aggregating each of the five predictor variables at each time point and then these five growth curves would be assessed for their contribution to the
outcome measure, the North Carolina third grade EOG Test of Reading Comprehension. This analysis was proposed in order to determine to what extent changes in skills over time contributed to the outcome measure. Upon further research into growth curve modeling it was determined the proposed model was unfeasible. A growth curve model makes use of several outcome measures aggregated into a growth curve rather than several predictor variables aggregated into a growth curve, as originally proposed.

In order to conduct analysis that would address the original purpose of question two as closely as possible it was determined the best way to make use of the longitudinal nature of the data set would be to run additional regressions. For each predictor variable determined to be significantly predictive of the outcome measure a multiple linear regression would be run inputing the predictor variable at all three time points.

As with the output from question one, the output for each of these regressions included standardized regression coefficients ($\beta$), the standard error for each regression coefficient, and a measure of statistical significance ($p$). The standardized regression coefficient for each time point denoted the relative contribution of that predictor variable time point to the outcome measure after controlling for the contribution of the other two times points and any covariates entered into the model. Standard error was the variability of the values for a particular predictor variable over repeated samplings. The measure of statistical significance indicated whether or not a given predictor variable time point contributed significantly to the outcome measure in the presence of other predictor variables time points. Each variable time point determined to be significant at the $p < .05$ level was judged to contribute significantly to the prediction of scores on the North Carolina third grade EOG Test of Reading Comprehension for that particular predictor variable.
Additionally, for each multiple linear regression analysis a multiple correlation coefficient, $R$, was reported. As $R$ is a correlation, $R^2$ represented the proportion of the variation in the outcome measure accounted for by the linear combination of the predictors. The output yielded a test of significance on this value as well, indicating whether or not the predictor variables together predicted the outcome measure at a better-than or worse-than chance level.
CHAPTER IV
RESULTS

Preliminary Analyses

The data set was examined for missing items and the distribution of these missing items was tested for randomness. Little’s chi-square statistic for testing if values are missing completely at random (MCAR), whose null hypothesis is that items are MCAR, was significant, $\chi^2(120.19, N = 96) = 0.48$, precluding the assumption that missing data points were MCAR. As the data set consisted of 111 cases with 23 observed variables, there are 2553 data points represented in this study. Out of those 2553 data points, fourteen values were missing. This number of missing cases is small (approximately 0.55% of the data set) and so EM estimation was used to impute missing data (IBM, 2010).

Prior to engaging in statistical analyses, variables of interest were transformed in order to make their examination more straightforward. The At Risk (AR) vs. Typically Developing (TD) variable was recoded into a dichotomous state, with TD being set to 0 and AR values being set to 1. Additionally, all subtest scores used to establish predictor variable values were transformed into z scores in order to facilitate further transformations.

Next, data were examined for outliers. This was done through examination of Mahalanobis distances calculated for each case and for each time point under analysis (1st grade, 2nd grade, 3rd grade). No cases exceeded the critical value required to be considered an outlier at the $p < .001$ significance level, considered a “very conservative probability estimate for a case being an outlier.” (Tabachnick and Fidell, 2001, pg 74).
**Phonological awareness.** Values for the phonological awareness predictor variable at time point one were derived from cohort one’s scores on the CTOPP Elision subtest at time point one and from cohort two’s scores on the PAL Phoneme subtest at time point one. Phonological Awareness at time points two and three were derived from the sample’s scores on the PAL Phoneme subtest at time point two and three, respectively. In order to test the assumption that both the CTOPP Elision subtest and PAL Phoneme subtest measured similar abilities a correlation was calculated between each individual’s score on either of these subtests and their PAL Phoneme subtest score at time point two. Cohort one’s CTOPP Elision subtest correlated significantly with their PAL Phoneme time point two subtest ($r = .712, p < .001$), as did cohort two’s PAL Phoneme time point one score with their PAL Phoneme time point two score ($r = .766, p < .001$). Since both of these time point one measures had similar values for $r$, it was assumed they were measuring the same ability and a single value was calculated for time point one phonological awareness.

**Orthographic processing.** Values for the orthographic processing predictor variable at all three time points were established through each child’s score on the WIAT-II Spelling subtest and their score on the PAL Word Choice Accuracy subtest at their respective time points. Scores from each were significantly correlated with each other at time points one ($r = .469, p < .001$), two ($r = .561, p < .001$), and three ($r = .626, p < .001$). The arithmetic mean of each child’s score was calculated for each time point.

**Rapid automatized naming.** Values for the Rapid Automatized Naming (RAN) predictor variable were calculated from the PAL Letters and PAL Letter Group subtest scores at their respective time points. Scores on these two subtests were found to correlate significantly with each other at time point one ($r = .714, p < .001$), time point two ($r = .760, p < .001$), and
< .001), and time point three \((r = .778, p < .001)\), but not to an extent that collinearity was of concern. Therefore a single value for the RAN predictor variable was established by calculating the arithmetic mean of the PAL Letters and PAL Letter Groups subtests at each time point.

**Phonological memory.** Values for the phonological memory predictor variable at time point one were derived from cohort one’s scores on the CTOPP Nonword Repetition subtest at time point one and from an arithmetic mean calculated from cohort two’s scores on the WISC-IV Digit Span Forward (DSF) and Digit Span Backward (DSB) subtests at time point one. Phonological Memory at time points two and three were derived from an arithmetic mean of the sample’s scores on the WISC-IV DSF and DSB subtests. In order to test the assumption that both the CTOPP Nonword Repetition subtest and WISC-IV DSF and DSB subtests measured similar abilities a correlation was calculated between each individual’s score on either of these subtests and an average of their WISC-IV DSF and DSB subtest score at time point two. While cohort two’s combined WISC-IV DSF and DSB scores correlated significantly from time point one to time point two \((r = .700, p < .001)\), cohort one’s CTOPP Nonword Repetition scores did not correlate significantly with their combined WISC-IV DSF and DSB scores \((r = .245, p < .072)\). These values were still used as cohort one’s time point one predictor variable values, however, as the CTOPP Nonword Repetition subtest is considered a valid measure of an individual’s phonological memory (Wagner et al. 1999).

**Receptive vocabulary.** Values for the receptive vocabulary predictor variable at time point one were derived from cohort one’s scores on the PPVT-4 at time point one and from cohort two’s scores on the CREVT-2 Receptive Vocabulary subtest at time point one.
Receptive Vocabulary at time points two and three were derived from the sample’s scores on the CREVT-2 Receptive Vocabulary subtest at time points two and three, respectively. In order to test the assumption that both the PPVT-4 and CREVT-2 Receptive Vocabulary subtest measured similar abilities a correlation was calculated between each individual’s score on either of these subtests and their CREVT-2 Receptive Vocabulary subtest score at time point two. Cohort one’s PPVT-4 subtest correlated significantly with their CREVT-2 Receptive Vocabulary time point two subtest ($r = .648$, $p < .001$), as did cohort two’s CREVT-2 Receptive Vocabulary time point one score with their CREVT-2 Receptive Vocabulary time point two score ($r = .670$, $p < .001$). Since both of these time point one measures had similar values for $r$, it was assumed they were measuring the same ability and a single value was calculated for time point one phonological awareness. See Table 4.1 for a list of which subtest scores were used to define predictor variables.

Table 4.1
*Table of Predictor Variable Composition*

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Time point 1</th>
<th>Time point 2</th>
<th>Time point 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonological Awareness</strong></td>
<td>CTOPP Elision for cohort one or PAL Phoneme for cohort</td>
<td>PAL Phoneme</td>
<td>PAL Phoneme</td>
</tr>
<tr>
<td><strong>Orthographic Processing</strong></td>
<td>Mean of WIAT-II Spelling and PAL Word Choice Accuracy</td>
<td>Mean of WIAT-II Spelling and PAL Word Choice Accuracy</td>
<td>Mean of WIAT-II Spelling and PAL Word Choice Accuracy</td>
</tr>
<tr>
<td><strong>Rapid Automatized Naming</strong></td>
<td>Mean of PAL Letters and PAL Letter Groups</td>
<td>Mean of PAL Letters and PAL Letter Groups</td>
<td>Mean of PAL Letters and PAL Letter Groups</td>
</tr>
<tr>
<td><strong>Phonological Memory</strong></td>
<td>CTOPP Nonword Repetition or WISC-IV Digit Span Forward and Digit Span Backward</td>
<td>WISC-IV Digit Span Forward and Digit Span Backward</td>
<td>WISC-IV Digit Span Forward and Digit Span Backward</td>
</tr>
<tr>
<td><strong>Receptive Vocabulary</strong></td>
<td>PPVT-4 or CREVT-2 Receptive Vocabulary</td>
<td>CREVT-2 Receptive Vocabulary</td>
<td>CREVT-2 Receptive Vocabulary</td>
</tr>
</tbody>
</table>
Data Screening

Initial screening of the data was conducted to determine if the AR vs. TD variable was significantly related to the outcome measure. This was done in order to evaluate whether or not it should be set as a covariate in subsequent analyses. A t-test comparing the AR group’s scores on the outcome measure with the TD group’s scores on the outcome measure was significant, \( t(109) = -3.80, p < .001 \). This information, combined with the prevalence of literature connecting reading difficulties with writing difficulties (e.g., Berninger et al., 2002), resulted in the decision being made to set the AR vs. TD variable as a covariate in all regression analyses.

In order to evaluate additional variables as potential covariates the relationship between the AR vs. TD variable and several other demographic variables was examined. AR and TD groups were not found to be significantly different from each other on measures of gender, \( \chi^2(1, N = 111) = 0.86, p = .35 \); ethnicity, \( \chi^2(4, N = 111) = 4.14, p = .39 \); or Hispanic/Caucasian identification, \( \chi^2(1, N = 111) = 0.05, p = .83 \). The two groups were found to be significantly different from each other on measures of IQ and Age at all three time points. As a result of these analyses it was decided the results of the regression analyses would be most appropriate to the research question if Age and IQ were entered into the regression analyses as a pair of second level covariates. See Table 4.2 for additional information on Age and IQ as they related to the AR vs. TD variable. Means and standard deviations for the AR group, for the TD group, and for the sample as a whole are presented in Table 4.3.
Table 4.2  
*Table of AR vs. TD variable as related to age and IQ*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Time Point</th>
<th>Relationship with AR vs. TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ</td>
<td>1</td>
<td>$t(109) = -3.46 = , p = .001$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$t(109) = -3.62 = , p &lt; .000$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$t(109) = -4.16 = , p &lt; .000$</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>$t(109) = 2.55 = , p = .012$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$t(108) = 2.63 = , p = .010$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$t(108) = 2.44 = , p = .016$</td>
</tr>
</tbody>
</table>

Table 4.3  
*Table of Means and Standard Deviations*

<table>
<thead>
<tr>
<th>Time</th>
<th>Variable</th>
<th>Total Sample</th>
<th>At Risk</th>
<th>Typically Developing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>Age</td>
<td>82.30</td>
<td>5.42</td>
<td>83.9</td>
</tr>
<tr>
<td></td>
<td>IQ</td>
<td>99.51</td>
<td>13.66</td>
<td>94.14</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>-0.14</td>
<td>0.97</td>
<td>-0.65</td>
</tr>
<tr>
<td></td>
<td>OP</td>
<td>-0.05</td>
<td>0.99</td>
<td>-0.61</td>
</tr>
<tr>
<td></td>
<td>RAN</td>
<td>0.00</td>
<td>0.98</td>
<td>-0.47</td>
</tr>
<tr>
<td></td>
<td>PM</td>
<td>-0.12</td>
<td>0.74</td>
<td>-0.37</td>
</tr>
<tr>
<td></td>
<td>RV</td>
<td>0.15</td>
<td>0.92</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>94.40</td>
<td>5.44</td>
<td>96.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100.04</td>
<td>15.16</td>
<td>93.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.13</td>
<td>1.10</td>
<td>-0.63</td>
</tr>
<tr>
<td>2</td>
<td>OP</td>
<td>0.18</td>
<td>0.90</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>RAN</td>
<td>0.31</td>
<td>0.93</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>PM</td>
<td>-0.22</td>
<td>0.72</td>
<td>-0.53</td>
</tr>
<tr>
<td></td>
<td>RV</td>
<td>-0.07</td>
<td>0.78</td>
<td>-0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>106.39</td>
<td>5.50</td>
<td>107.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>106.65</td>
<td>15.00</td>
<td>99.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.16</td>
<td>1.29</td>
<td>-0.65</td>
</tr>
<tr>
<td>3</td>
<td>OP</td>
<td>0.01</td>
<td>0.77</td>
<td>-0.36</td>
</tr>
<tr>
<td></td>
<td>RAN</td>
<td>0.17</td>
<td>1.01</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>PM</td>
<td>-0.20</td>
<td>0.69</td>
<td>-0.43</td>
</tr>
<tr>
<td></td>
<td>RV</td>
<td>-0.04</td>
<td>0.84</td>
<td>-0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>337.95</td>
<td>23.60</td>
<td>327.84</td>
</tr>
</tbody>
</table>

Note: Age is presented in months, IQ is a standard score (mean = 100, SD = 15), Reading EOG is a modified scaled score. All other variables are reported as z-scores.
**Question one.** Do the variables identified through a review of the literature as being predictors of reading achievement in first, second, and third grade contribute to scores on the North Carolina third grade EOG Test of Reading Comprehension?

**Question one analysis.** Three multiple linear regressions were conducted using data from each of the three time points in order to determine the predictive contribution of each predictor variable to the outcome measure in all three grades. For each multiple linear regression AR vs. TD status was entered as a covariate in a first block and IQ and age were entered as covariates in a second block in order to control for between-group differences and to minimize the effects of age and IQ on the estimated predictive qualities of the five predictor variables. Setting these variables as covariates helped to ensure that the end results of the analysis estimated the contribution of the five predictor variables were measured isolated from as many confounding variables as possible and to account for limitations inherent in the data set.

Assumptions necessary for multiple linear regression were evaluated prior to running these analyses. The data were first examined for the existence of bivariate collinearity and multicollinearity. Examination of bivariate correlation was conducted using Pearson Product-moment Correlations between all independent variables (IV) for each of the three measurement points. Though many of the IV pairings were significantly correlated, none of these correlations were .9 or higher, the value recommended by Tabachnick and Fidell as a critical value for establishing the existence of bivariate collinearity (2001, pg 88).

Investigation for potential multicollinearity was conducted through examination of tolerance and variance inflation factor (VIF) values. Specifically, variables were examined at each of the three time points for tolerance values less than .10 or VIF values greater than 10.
No variables were found to have values exceeding these amounts set forth by Kline (2005) as being indicative of problematic levels of multicollinearity. As such, analysis proceeded under the assumption that problematic levels of bivariate collinearity and multicollinearity were not present in the data set.

The results of the multiple linear regression indicates the model was able to predict the outcome measure at a rate significantly better than chance at all three time points, and accounted for approximately 50% of outcome measure variance at time point one, 67% of outcome measure variance at time point two, and 64% of outcome measure variance at time point three. The contribution of each predictor variable at each time point varied after controlling for the contributions of other variables. Receptive vocabulary accounted for a significant portion of variance at time point one. Phonological awareness, RAN, and receptive vocabulary accounted for a significant portion of variance at time point two. Orthographic processing and RAN accounted for a significant portion of variance at time point three. See Table 4.4 for output data, the columns of which represent each regression and the rows of which represent covariates and posited predictor variables. Power estimates for all three time points were determined to be greater than .999 (Soper, 2011). Following the regressions, plots of residuals versus predicted value were examined and the data were determined to be homoscedastic. Additionally, examination of residual histograms showed the data to be sufficiently normally distributed and standardized residual plots revealed slightly kurtotic but otherwise normal distribution of errors.

As phonological memory was not found to be significantly predictive of the outcome measure, it was not included in analysis for question two.
Table 4.4
Table of Regression Data, Analyses for Question One

<table>
<thead>
<tr>
<th></th>
<th>Regression for Time Point 1</th>
<th>Regression for Time Point 2</th>
<th>Regression for Time Point 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Risk status</td>
<td>$\Delta R^2 = .117^{**}$</td>
<td>$\Delta R^2 = .117^{**}$</td>
<td>$\Delta R^2 = .104^{*}$</td>
</tr>
<tr>
<td>IQ and Age</td>
<td>$\Delta R^2 = .246^{**}$</td>
<td>$\Delta R^2 = .326^{**}$</td>
<td>$\Delta R^2 = .332^{**}$</td>
</tr>
<tr>
<td>Predictors</td>
<td>$\Delta R^2 = .142^{**}$</td>
<td>$\Delta R^2 = .233^{**}$</td>
<td>$\Delta R^2 = .206^{**}$</td>
</tr>
<tr>
<td>Total $R^2$</td>
<td>$R^2 = .506^{**}$</td>
<td>$R^2 = .676^{**}$</td>
<td>$R^2 = .643^{**}$</td>
</tr>
<tr>
<td>Covariates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Risk</td>
<td>$\beta = -.342^{**}$</td>
<td>$\beta = -.342^{**}$</td>
<td>$\beta = -.323^{*}$</td>
</tr>
<tr>
<td>Std Error</td>
<td>4.34</td>
<td>4.34</td>
<td>4.12</td>
</tr>
<tr>
<td>Covariates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>$\beta = -.225^{*}$</td>
<td>$\beta = -.200^{*}$</td>
<td>$\beta = -.143$</td>
</tr>
<tr>
<td>Std Error</td>
<td>0.346</td>
<td>0.324</td>
<td>0.309</td>
</tr>
<tr>
<td>IQ</td>
<td>$\beta = .435^{**}$</td>
<td>$\beta = .552^{**}$</td>
<td>$\beta = .572^{**}$</td>
</tr>
<tr>
<td>Std Error</td>
<td>0.141</td>
<td>0.119</td>
<td>0.119</td>
</tr>
<tr>
<td>Individual Predictors***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phon. Awareness</td>
<td>$\beta = .110$</td>
<td>$\beta = .249^{*}$</td>
<td>$\beta = .039$</td>
</tr>
<tr>
<td></td>
<td>2.68 (2.56)</td>
<td>5.31 (1.82)</td>
<td>0.69 (1.35)</td>
</tr>
<tr>
<td>Orth. Processing</td>
<td>$\beta = .126$</td>
<td>$\beta = -.013$</td>
<td>$\beta = .302^{*}$</td>
</tr>
<tr>
<td></td>
<td>3.01 (2.49)</td>
<td>-0.318 (2.29)</td>
<td>.925 (2.59)</td>
</tr>
<tr>
<td>RAN</td>
<td>$\beta = .024$</td>
<td>$\beta = .306^{**}$</td>
<td>$\beta = .242^{**}$</td>
</tr>
<tr>
<td></td>
<td>0.571 (2.14)</td>
<td>7.79 (1.61)</td>
<td>5.49 (1.48)</td>
</tr>
<tr>
<td>Phon. Memory</td>
<td>$\beta = .156$</td>
<td>$\beta = .090$</td>
<td>$\beta = .079$</td>
</tr>
<tr>
<td></td>
<td>4.98 (2.63)</td>
<td>2.96 (2.44)</td>
<td>2.63 (2.54)</td>
</tr>
<tr>
<td>Recep. Vocabulary</td>
<td>$\beta = .212^{*}$</td>
<td>$\beta = .249^{*}$</td>
<td>$\beta = .092$</td>
</tr>
<tr>
<td></td>
<td>5.45 (2.16)</td>
<td>7.54 (2.34)</td>
<td>2.45 (2.23)</td>
</tr>
</tbody>
</table>

Note: *indicates statistically significant at $p < .05$, **indicates statistically significant at $p < .001$ ***for each predictor variable a beta weight is listed, followed by the unstandardized coefficient (standard error).

**Question two.** Does the growth on measures of these variables, identified through a review of the literature as being predictors of reading achievement in first, second, and third grade contribute to scores on the North Carolina third grade EOG Test of Reading Comprehension?

**Question two analysis.** The results from the analyses conducted as part of question one indicated phonological memory was not a significant predictor of the outcome measure at any time point. As such phonological memory was omitted from subsequent analysis and...
four additional multiple linear regressions were conducted at all three time points: phonological awareness, orthographic processing, RAN, and receptive vocabulary. For each of these four multiple linear regressions AR vs. TD status was entered as a covariate in a first block and IQ and age as estimated at grade three were entered as covariates in a second block in order to control for between-group differences and to minimize the effects of age and IQ on the estimated predictive qualities of the five predictor variables. Setting these variables as covariates helped to ensure that the end results of the analysis estimated the contribution of the predictor variables at each time point were isolated from as many confounding variables as possible and to account for limitations inherent in the data set.

The results of the multiple linear regression indicate each set of predictor variables were able to predict the outcome measure at a rate significantly better than chance at all three time points. Examination of the beta weights for each predictor variable indicated phonological awareness and RAN were most predictive at time point two and orthographic processing was most predictive at time point three. Receptive vocabulary was never significantly predictive at any given time point, but it contributed the most to outcome measure prediction at time point two. See Table 4.5 for output data, the columns of which represent each regression and the rows of which represent covariates and posited predictor variables. Power estimates for all three time points were determined to be greater than .999 (Soper, 2011). Following the regressions, plots of residuals versus predicted value were examined and the data were determined to be homoscedastic. Additionally, examination of residual histograms showed the data to be sufficiently normally distributed and standardized residual plots revealed slightly kurtotic but otherwise normal distribution of errors.
Table 4.5

Table of Regression Data for Question Two

<table>
<thead>
<tr>
<th>Overall Model</th>
<th>Regression for Phonological Awareness</th>
<th>Regression for Orthographic Processing</th>
<th>Regression for RAN</th>
<th>Regression for Receptive Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta R^2 = .117^{**}$</td>
<td>$\Delta R^2 = .117^{**}$</td>
<td>$\Delta R^2 = .117^{**}$</td>
<td>$\Delta R^2 = .117^{**}$</td>
</tr>
<tr>
<td>At Risk status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQ and Age</td>
<td>$\Delta R^2 = .346^{**}$</td>
<td>$\Delta R^2 = .346^{**}$</td>
<td>$\Delta R^2 = .346^{**}$</td>
<td>$\Delta R^2 = .346^{**}$</td>
</tr>
<tr>
<td>Predictors</td>
<td>$\Delta R^2 = .112^{**}$</td>
<td>$\Delta R^2 = .174^{**}$</td>
<td>$\Delta R^2 = .148^{**}$</td>
<td>$\Delta R^2 = .052^{*}$</td>
</tr>
<tr>
<td>Total $R^2$</td>
<td>$R^2 = .575$</td>
<td>$R^2 = .637$</td>
<td>$R^2 = .611$</td>
<td>$R^2 = .515$</td>
</tr>
<tr>
<td>Covariates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Risk</td>
<td>$\beta = -.342^{**}$</td>
<td>$\beta = -.342$</td>
<td>$\beta = -.342^{**}$</td>
<td>$\beta = -.342^{**}$</td>
</tr>
<tr>
<td>Std Error</td>
<td>4.34</td>
<td>4.34</td>
<td>4.34</td>
<td>4.34</td>
</tr>
<tr>
<td>Covariates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>$\beta = -.127$</td>
<td>$\beta = -.127$</td>
<td>$\beta = -.127$</td>
<td>$\beta = -.127$</td>
</tr>
<tr>
<td>Std Error</td>
<td>0.332</td>
<td>0.322</td>
<td>.322</td>
<td>0.322</td>
</tr>
<tr>
<td>IQ</td>
<td>$\beta = .596^{**}$</td>
<td>$\beta = .596^{**}$</td>
<td>$\beta = .596^{**}$</td>
<td>$\beta = .596^{**}$</td>
</tr>
<tr>
<td>Std Error</td>
<td>0.122</td>
<td>0.122</td>
<td>.122</td>
<td>0.122</td>
</tr>
<tr>
<td>Time Points***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Point 1</td>
<td>$\beta = .036$</td>
<td>$\beta = .009$</td>
<td>$\beta = -.066$</td>
<td>$\beta = .075$</td>
</tr>
<tr>
<td></td>
<td>0.87 (2.51)</td>
<td>0.20 (2.27)</td>
<td>-1.59 (2.09)</td>
<td>1.93 (2.43)</td>
</tr>
<tr>
<td>Time Point 2</td>
<td>$\beta = .342^{*}$</td>
<td>$\beta = .170$</td>
<td>$\beta = .294^{*}$</td>
<td>$\beta = .215$</td>
</tr>
<tr>
<td></td>
<td>7.31 (2.69)</td>
<td>-4.48 (3.23)</td>
<td>7.50 (2.92)</td>
<td>6.51 (3.51)</td>
</tr>
<tr>
<td>Time Point 3</td>
<td>$\beta = .038$</td>
<td>$\beta = .599^{**}$</td>
<td>$\beta = .165$</td>
<td>$\beta = .066$</td>
</tr>
<tr>
<td></td>
<td>0.70 (1.98)</td>
<td>18.36 (3.35)</td>
<td>3.86 (2.53)</td>
<td>1.84 (3.33)</td>
</tr>
</tbody>
</table>

*indicates statistically significant at $p < .05$, **indicates statistically significant at $p < .001$ ***for each predictor variable a beta weight is listed, followed by the unstandardized coefficient (standard error).
CHAPTER V

DISCUSSION

The current study examined to what extent measures of reading skills known to contribute to early literacy development also contributed to scores on the North Carolina third grade End of Grade (EOG) Test of Reading Comprehension. The EOG Test of Reading Comprehension is a standardized measure used in the state of North Carolina to estimate a child’s reading achievement. Many decisions are made based at least in part on the results of mass administrations of this instrument, ranging from child retention to decisions regarding teacher efficacy to evaluations of school and district performance. It is important to have an understanding of the extent to which the results of this outcome measure align with skills identified through decades of literacy research as being important to the development of early literacy, as it is this literacy research that guides curriculum development and implementation and subsequent assessment of a student’s growth on said curriculum.

Such identified skills include phonological awareness (e.g., Schatschneider, Fletcher, Francis, Carlson, and Foorman, 2004), orthographic processing (e.g., Juel, 1998), rapid automatized naming (e.g., McBride-Chang & Manis, 1996), phonological memory (e.g., Cain, Bryant and Oakhill, 2004), and receptive vocabulary (e.g., DeThrone, Petrill, Schatschneider, and Cutting, 2010). Identified by literacy researchers as being predictive of early literacy development, it would be reasonable to assume that these skills would also be predictive of a measure that purports to measure such literacy development, such as the EOG Test of Reading Comprehension.
The development of the EOG Test of Reading Comprehension was conducted in a manner consistent with industry standards for the development of psychometrically sound instruments (Office of Curriculum and School Reform Services, 2004, p. 14). Such standards include the establishment of instrument reliability and validity. Evidence is provided of the EOG’s reliability in the form of internal consistency and evidence of validity is provided in the form of both content and criterion validity. What is not provided, however, is an assessment of the extent to which skills predictive of reading development are also predictive of scores on the outcome measure; that is, to what extent does the instrument align with other instruments purported to measure the same skills?

The current study attempted to assess whether or not scores on the EOG can be predicted by knowledge of skills known to be predictive of reading outcomes in general. This was done by evaluating the predictive contribution of measures of skills of early literacy development to scores from the EOG Test of Reading Comprehension. Strengths of this study include the use of psychometrically sound measures of such early literacy skills, a relatively large sample, and multiple time point measures. The text below describes the result of this study’s analysis and discusses the extent to which the results relate to each of the previously postulated hypotheses. A discussion of the results, the limitations of this study, and potential future research are also discussed in the text below.

Question one. Do the variables identified through a review of the literature as being predictors of reading achievement in first, second, and third grade contribute to scores on the North Carolina third grade EOG Test of Reading Comprehension?
Hypothesis one. It was hypothesized that the variables identified as being predictors of reading achievement, as measured in first, second, and third grade, would significantly contribute to scores on the North Carolina third grade EOG Test of Reading Comprehension.

The proposed hypothesis was partially supported by the obtained results. After controlling for limitations of the data set, and after controlling for the contribution of child Age and IQ, the predictor variables as a whole accounted for a significant amount of the variance in the outcome measure at each of the three time points. The predictor variables as a whole as measured in second grade accounted for the greatest amount of variance in the outcome measure, followed by the predictor variables as measured in third grade, followed by the variables as measured in first grade. At no time point, however, did the predictor variables account for more variance than the Age and IQ variables.

When examined individually, however, each predictor variable contributed to a different extent at different time points after holding constant the effects of the other predictor variables. In first grade only receptive vocabulary was significant by itself. In second grade measures of phonological awareness, RAN, and receptive vocabulary were each significant contributors to the outcome variable. In third grade only orthographic processing and RAN were significant contributors to the outcome variable. At no time point was phonological memory found to be a significant contributor to the outcome variable.

Question two. Does the growth on measures of these variables, identified through a review of the literature as being predictors of reading achievement in first, second, and third grade contribute to scores on the North Carolina third grade EOG Test of Reading Comprehension?
Hypothesis two. It was hypothesized that change over time in receptive vocabulary, orthographic processing, rapid automatized naming, phonological memory, and receptive vocabulary will significantly contribute to scores on the third grade EOG Test of Reading Comprehension.

Due to a flaw in the proposed methodology hypothesis two was not addressed. In its place the question to what extent does the predictive value of each predictor variable change across the three measured time points with respect to scores obtained on the outcome measure was addressed. Therefore the hypothesis investigated was that each predictor variable found significantly predictive from analysis related to question one would contribute to scores on the North Carolina third grade EOG Test of Reading Comprehension to a greater extent in time point two than time point one, and to a greater extent at time point three than at time point two.

This proposed hypothesis was not supported by the results. Orthographic processing followed the hypothesized pattern, increasing in its predictive contribution across time points. Phonological awareness, RAN, and receptive vocabulary did not, with each of these three variables obtaining the highest level of contribution at time point two. Not all variables contributed significantly at all time points, either. While significant contributions were observed at various time points for three of the four predictor variables, at no one point was receptive vocabulary found to be significantly contributory to the outcome measure.

Interpretation for Questions One & Two

Results of the analysis suggest there exists meaningful relationships between the predictor variables and the outcome measure. The predictor variables significantly contributed to prediction of the outcome measure at all three time points when taken as a
whole. This contribution was significant even after accounting for the contribution by the
covariates. All predictors taken together accounted for 50.6% of outcome measure variance
when measured in first grade, 67.6% of outcome measure variance when measured in second
grade, and 64.3% of outcome measure variance when measured in third grade.

At the first time point receptive vocabulary was the only variable that rose to
statistical significance after accounting for the other predictor variables. At the second time
point phonological awareness, RAN, and receptive vocabulary each accounted for a
significant proportion of the variance after accounting for the contribution by the other four
predictor variables. At the third time point orthographic processing and RAN each accounted
for a significant proportion of the variance after accounting for the contribution by the other
four predictor variables. At no point did phonological memory significantly contribute to
model prediction in and of itself. When the examined longitudinally, the significant predictor
variables phonological awareness, RAN, and receptive vocabulary were found to contribute
the greatest at time point two and orthographic processing contributed the greatest at time
point three.

While the overall models were found to be significantly predictive of the outcome
measure at all three time points, individual predictor variables were found to be significantly
predictive at different time points. Though the data set does not allow for investigations of
such, part of the pattern of predictor variable statistical significance is likely related to where
these skill fall on the hierarchy of pre-literacy skills and how they are representative of
exposure to language in general and reading in particular, and classroom literacy instruction
in specific. Several methods for skill remediation and instruction can be found in the reports
from the National Reading Panel (National Reading Panel, 2000) and the National Early Literacy Panel (National Early Literacy Panel, 2008).

For example, receptive vocabulary was found to be significant in first and second grade, but not in third grade. Receptive vocabulary is a measure of an individual’s understanding of spoken words’ meanings. Children with either high or low receptive vocabulary scores have likely been exposed to either large or small amounts of spoken and written words, respectively (Nagy, Herman & Anderson, 1985). The difference between a relatively small vocabulary and a relatively large vocabulary likely becomes smaller as children age and progress through their education. By the time children reach third grade it is likely almost all have a working knowledge of most of words they will experience in the classroom and on standardized measures of reading, and therefore look more similar on a measure of receptive vocabulary (Bowey, 1996).

The same changing pattern of importance can be observed in three other variables found to be significantly predictive at some time points but not others: orthographic processing, RAN, and phonemic awareness. Of the five skills evaluated as part of this study orthographic processing (Burt, 2006), RAN (McBride-Chang & Manis, 1996), and phonemic awareness (Foorman et al., 1998) are most related to the decoding of written words. Orthographic processing and phonemic awareness allow a reader to make sense of unfamiliar letter patterns. Measures of RAN, along with measures of phonemic awareness are an effective estimate of how efficient an individual is able to engage in phonological processing (Scarborough, 1998), the phonemic decoding of written words. The findings from this study suggest children only begin to differentiate on such measures of more advanced engagement with literacy after they have been in school for more than a year.
At no time point was phonological memory found to be significant after accounting for other predictor variables’ contribution to the outcome measure. It is possible that the reduced importance of phonological memory is related to the format of the outcome measure itself. Measuring phonological memory and phonemic awareness involves having a child vocalize words, an activity not required when taking the North Carolina third grade EOG Test of Reading Comprehension. Rather, the test has the examinee read and respond to questions in a written, multiple choice question format. This difference in format could in part explain why phonological memory wasn’t found to be significantly predictive at all time points and why phonemic awareness was only found to be significantly predictive at time point two, and only at the \( p < .05 \) level; their relevance to the modality of the skills assessed by the outcome measure is limited. Phonemic awareness and phonological memory in general, however, likely allow an individual to access the curriculum and learn words, becoming more fluent in other skills observed by this study to more significantly contribute to the outcome measure.

An examination of documents relevant to the North Carolina third grade EOG Test of Reading Comprehension verify these assertions regarding the activities required by the instrument. This outcome measure was designed to measure concepts from the North Carolina Standard Course of Study (NCSCS, Office of Curriculum and School Reform Services, 2004) and does so through the assessment of three different goals. See Table 5.1 for more information.

Though goal three mentions the use of oral language, this is in the form of receptive oral language. All student responses are in the form of responses to written multiple choice
questions (Office of Curriculum and School Reform Services, 2004). This precludes the test’s results being reflective of skills related to word calling, such as phonological memory.

Table 5.1
North Carolina Third Grade EOG Test of Reading Comprehension Goals

<table>
<thead>
<tr>
<th>Goal</th>
<th>Description of Goal</th>
<th>Percent Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The learner will apply enabling strategies and skills to read and write.</td>
<td>5-10</td>
</tr>
<tr>
<td>2</td>
<td>The learner will apply strategies and skills to comprehend text that is read, heard, and viewed.</td>
<td>62-68</td>
</tr>
<tr>
<td>3</td>
<td>The learner will make connections through the use of oral language, written language, and media and technology.</td>
<td>23-27</td>
</tr>
</tbody>
</table>

Additional Exploratory Analyses for Questions One & Two

Covarying out IQ and Age allows for a better estimate of the contribution of literacy-related variables to the score outcome. Having this information informs the appropriateness of its use as a reading outcome measure. However, a secondary purpose of this study was to determine whether or not the available data allowed for prediction of outcome measure performance. Establishing whether or not the data allowed for such prediction would suggest a protocol for screening students prior to the administration of the EOG Test of Reading Comprehension. This screening protocol would allow educators to estimate whether or not individual students are at risk for obtaining low scores on the EOG, which would in turn allow them to better target intervention resources.

By administering the afore-described measures of writing AR vs. TD status, IQ, phonological awareness, orthographic processing, RAN, phonological memory, and receptive vocabulary, in addition to knowing the children’s age, we can account for approximately 50.6% of outcome measure variance with first grade predictor measures, 67.6% of outcome measure variance with second grade predictor measures, and 64.3% of outcome measure
variance with third grade predictor measures. By making use of the provided constant and
beta weights (β) a predictive regression equation can be constructed in order to estimate an
individual child’s predicted outcome measure score. The correlation between the sample’s
actual scores and predicted scores were measured to be .711 in first grade, .822 in second
grade, and .802 in third grade. See Table 5.2 for the predictive regression equations derived
from previous analyses.

Table 5.2
Predictive Regression Equations with All Covariates

<table>
<thead>
<tr>
<th>Time Point</th>
<th>Regression Equation for estimated outcome measure</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>369 + -.580(AR vs. TD) + -.925(Age) + .458(IQ) + .110(PA) + .126(OP) + .024(RAN) + .156(PM) + .212(RV)</td>
<td>.711</td>
</tr>
<tr>
<td>2</td>
<td>379.168 + 1.33(AR vs. TD) + -.784(Age) + .318(IQ) + .248(PA) + -.013(OP) + .306(RAN) + .090(PM) + .249(RV)</td>
<td>.822</td>
</tr>
<tr>
<td>3</td>
<td>297.35 + 3.13(AR vs. TD) + -.202(Age) + .572(IQ) + .039(PA) + .302(OP) + .242(RAN) + .079(PM) + .092(RV)</td>
<td>.802</td>
</tr>
</tbody>
</table>

Of interest with constructing screening protocols, however, is the time required to
administer each assessment. Removing AR vs. TD and IQ from the analyses results in similar
predictive regression equations and correlations, summarized in Table 5.3.

Table 5.3
Predictive Regression Equations with Age as Covariate

<table>
<thead>
<tr>
<th>Time Point</th>
<th>Regression Equation for estimated outcome measure</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>425.277 + -.1058(Age) + .152(PA) + .128(OP) + .053(RAN) + .190(PM) + .290(RV)</td>
<td>.674</td>
</tr>
<tr>
<td>2</td>
<td>408.835 + -.747(Age) + .297(PA) + -.016(OP) + .286(RAN) + .138(PM) + .344(RV)</td>
<td>.811</td>
</tr>
<tr>
<td>3</td>
<td>363.771 + -.236(Age) + .019(PA) + .306(OP) + .246(RAN) + .205(PM) + .281(RV)</td>
<td>.790</td>
</tr>
</tbody>
</table>
All three time point equations in both conditions produced significant models. Removing AR vs. TD and IQ measures from the proposed screening protocol would eliminate anywhere from forty to ninety minutes from the screening protocol without reducing the accuracy of the regression equation by a substantial amount; indeed, the correlation between actual scores and scores predicted by the time point two and time point three equations were nearly identical.

As with all screening protocols, there is a tradeoff between accuracy and time to administer. Determining which combination of subtests would make for an ideal screening battery would depend on the requirements of each individual agency and a comprehensive evaluation of the possibilities is beyond the scope of this discussion. The data presented in this section, however, provides a model for establishing a screening protocol in order to estimate future performance on standardized outcome measures once the validity of the outcome measure has been established. Constructing such a protocol would have two main benefits.

Firstly, creating a protocol and implementing administration and evaluation procedures would allow for more efficient application of remedial interventions. If an outcome measure is deemed to be a valid measure of academic achievement, and if component skills can be established as being predictive of outcome measure scores, measuring component skill ability can be used to estimate outcome measure scores. By estimating outcome measure scores educators can provide intensive instruction on component skills to children most at risk for poor performance.

The second benefit of creating and implementing such a screening protocol would be to allow educators to instruct students on core educational components rather than “teaching
to the test.” A criticism of the use of standardized outcome measures is that it encourages educators to “teach to the test.” Teaching to the test refers to the practice of training students to accurately respond to the types of items typical to the outcome measure, rather than providing students with generalized instruction in a content area. By providing students with generalized instruction rather than specific training it is hoped they will be able to apply said instruction across a wider variety of settings and increase their overall academic abilities.

A method by which the accuracy of such a screening method might be analyzed is to examine the model’s accuracy of case classification into achievement levels. Achievement levels on North Carolina EOG and EOC tests range from I through IV (See Table 1.1 for additional information). By examining the patterns of classification (see Table 5.4) it can be determined how accurately the screening protocol classifies students as compared to their actual performance. Such data can be reexamined each year for further fine tuning of the screening protocol.

<p>| Table 5.4  |
| Actual Achievement Levels vs. Predicted Achievement Levels |</p>
<table>
<thead>
<tr>
<th>Achievement Level</th>
<th>Actual</th>
<th>Predicted at T1 Full</th>
<th>Predicted at T1 Screen</th>
<th>Predicted at T2 Full</th>
<th>Predicted at T2 Screen</th>
<th>Predicted at T3 Full</th>
<th>Predicted at T3 Screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>23</td>
<td>34</td>
<td>27</td>
<td>29</td>
<td>28</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>II</td>
<td>11</td>
<td>19</td>
<td>23</td>
<td>20</td>
<td>27</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>III</td>
<td>45</td>
<td>30</td>
<td>34</td>
<td>30</td>
<td>36</td>
<td>29</td>
<td>33</td>
</tr>
<tr>
<td>IV</td>
<td>32</td>
<td>28</td>
<td>27</td>
<td>32</td>
<td>29</td>
<td>35</td>
<td>29</td>
</tr>
<tr>
<td>% Correct</td>
<td>45.95</td>
<td>49.55</td>
<td>55.86</td>
<td>51.35</td>
<td>49.55</td>
<td>50.45</td>
<td></td>
</tr>
</tbody>
</table>

Tests of significance indicated the prediction of student achievement levels were significantly different for all three full model predictions (T1: $\chi^2(9, N = 111) = 49.64, p < .001$; T2: $\chi^2(9, N = 111) = 75.26, p < .001$; T3: $\chi^2(9, N = 111) = 78.57, p < .001$) and all three screening protocol predictions (T1: $\chi^2(9, N = 111) = 45.11, p < .001$; T2: $\chi^2(9, N = 111)$
= 64.28, \( p < .001 \); \( \chi^2(9, N = 111) = 52.53, p < .001 \), suggesting further research and refinement is needed prior to engaging in actual implementation.

Inspection of the results reveals the most accurate model for achievement level prediction was the full model at time point two. The full model, including AR vs. TD and IQ as covariates, predicted student achievement levels more accurately in second grade, while the proposed screening protocol predicted student achievement more accurately at time point one and at time point two. The overall accuracy for prediction averaged around 50.45% and appeared to err on the side of under prediction of student achievement levels at all time points. Calculating screening protocol data in this manner will allow for the development of more accurate and efficient screening protocols.

**Limitations**

This study provided insight into the validity of the EOG Test of Reading Comprehension. Limitations, however, need to be addressed as these aspects of the study attenuate the generalizability of the results. This study made use of a preexisting data set that was ascertained in order to meet parameters established for a different research study. The sample was comprised of children classified as either being At Risk or Typically Developing with respect to written expression. Such classification was made based on a child either being below or above a standard score of 85 on a standardized test of written expression. This precludes the assumption that the sample is comprised of a random sampling of students. Given the well established link between reading and writing (Berninger et al., 2002), this dichotomized sample cannot be said to be truly representative of the population from which it was derived with respect to reading ability. The idea that these two groups of children are
potentially fundamentally different is borne out by the previously reported differences in performance on predictor and outcome measures. See Table 4.3 for additional information.

A second limitation of this study are the measures used to estimate each child’s skills in each domain. Due to requirements imposed by the school district in which these data were collected, measures used to estimate phonological awareness, phonological memory, and receptive vocabulary were changed after cohort one was assessed in year one. This introduced the possibility of the measurement of different aspects of each construct between cohorts during year one.

All measures substituted after cohort one year one were psychometrically validated measures of the same construct. There was found, however, a low correlation between cohort two’s year one measure of phonological memory and cohort two’s year two measure of phonological memory. These measures, the CTOPP Nonword Repetition subtest and the WISC-IV DSF and DSB combined subtests, respectively, are both established measures of phonological memory as they require an examinee to repeat verbally presented information. The low correlation between the two scores could potentially be explained by the fact that the WISC-IV DSF and DSB subtests require memory of known words in the form of single digit numbers in sequence, while the CTOPP Nonword Repetition subtest requires the repetition of nonsense words which do not have symbolic or abstract meaning.

A third limitation of this study is that the sample size was potentially too small for conducting multiple linear regression with as many predictor variables as were used. Recommended sample size for multiple linear regression ranges from five cases per predictor to twenty cases per predictor, with an overall N of at least 100 being ideal. Tabachnick and Fidel (2007) recommend 50 + 8(k) or 104+k subjects when testing a full regression model.
As the majority of the analyses as run included eight predictor variables, five main variables and three covariates, the 111 children is short of the 114 or 112 such formulae suggest.

**Future Research**

The results of this study should help to inform future research into the areas of standardized test validation and outcome measure prediction. Findings suggest that it might be possible to increase scores on outcome measures by providing instruction in relevant component skills. Assessed through this study were five such skills determined through a review of relevant literature to be important in the development of early literacy. Though these skills were frequently seen as being important to the development of reading, there were several other variables also mentioned that would have been interesting to measure.

Such variables include visual acuity, a family history of dyslexia, and concepts of print. These variables likely contribute differently at different points in the development of early literacy. Research investigated by the National Early Literacy Panel (2008) suggest measures of concept of print are more predictive of reading outcomes as measured prior to and during kindergarten and less predictive of such reading outcomes when measured in second or third grade. Skills such as concepts of print and phonemic awareness are skills mastered early in the process of learning to read. By grade three most all children will have mastered these skills. As children will not differentiate in their mastery of these skills in second and later grades, measures of these skills taken in second and later grades will not be predictive of reading outcome measures. This does not preclude the use of these measures as predictors of reading outcome measures. It merely suggests that their contribution will likely only be significant when measured in first grade or prior.
Visual acuity could be more accurately conceived of as a covariate as some difficulties in visual acuity are more likely representative of biological deficits rather than learning deficits. The same could be true of a family history of dyslexia; such hereditary conditions could potentially be passed down from parent to child and would play a significant role in an individual’s ability to acquire the skills necessary for the acquisition of literacy.

It is also possible, however, that a family history of dyslexia could be representative of a sociological phenomenon. An individual from a family that does not and has not valued the pursuit of literacy might assess similar to an individual with an organic cause of dyslexia, and vice versa. The difference between innate dyslexia and a lack of exposure to books and printed materials could in part be accessed through a measure of an individual’s understanding of basic print concepts.

Also of interest for future research would be to apply this model of outcome measure evaluation and at-risk screening to other measures. The state of North Carolina administers EOG and End of Course (EOC) tests at many times and across several subjects. As addressed earlier, such extensive use of high-stakes testing encourages “teaching to the test,” an inefficient pedagogical practice. A more efficient use of instructional time would be to instruct students in the skills that underlie an individual’s development of the subject being assessed. By identifying contributive skills and the extent to which they contribute we can better understand what students at risk for low performance look like, and we can better apply our rehabilitative resources.

Conclusions
Early literacy acquisition is not a linear and singular process. Its development is dependent on the development of several sub skills. A review of the literature suggested five skills important to the acquisition of early literacy from grades one through three, each of which contributing to a different extent at different time points. Measures of these variables at time points from 111 children were regressed on scores obtained from the North Carolina third grade End of Grade Test of Reading Comprehension in order to determine the extent to which they were able to predict scores on this outcome measure. The measures as a whole significantly predicted outcome scores even after controlling for IQ, age, and AR vs. TD status with respect to written expression, a feature of the data set used in this analysis.

Predictive contributions of each variable individually were found to be different at each of the three time points. At time point one only receptive vocabulary was significantly predictive after controlling for the effects of other predictor variables. At time point two phonological awareness, RAN, and receptive vocabulary were each predictive after accounting for the contributions of the other four variables. In third grade orthographic processing and RAN were predictive after accounting for the contributions of the other four variables.

Results of this study indicate that it is possible to predict scores on the North Carolina third grade EOG Test of Reading Comprehension through evaluation of a child’s component reading skills, which suggests the North Carolina third grade EOG Test of Reading Comprehension can be a valid tool to estimate reading ability. The results also posit a simple screening protocol that will suggest which students might be at risk for failure on the EOG in time to intervene. This study also suggests a framework for further investigating the validity of other EOG and EOC tests. By making use of the results of this study and through the
application of its methods to other high stakes tests, it is postulated that the application of instructional resources can be better evaluated and students can more effectively be taught.
Appendix A

To: Stephen Hooper
Psychiatry, School Of Education, Developmental
CB:7255 1450 Nc Hwy 54 East

From: Biomedical IRB

Authorized signature on behalf of IRB

Approval Date: 6/20/2008
Expiration Date of Approval: 6/15/2009

RE: Notice of IRB Approval by Expedited Review (under 45 CFR 46.110)
Submission Type: Modification
Expedited Category: Minor Change to Previously Reviewed Research
Study #: 05-2755 (Former IRB Number 05-CDL-1325)

Study Title: Attention, Memory, and Executive Functions in Written Language Expression in Elementary School Children
Sponsors: Dept of Education Institute
This letter serves as revision to the approval letter dated 6/20/08.

This submission has been approved by the above IRB for the period indicated. It has been determined that the risk involved in this modification is no more than minimal. Unless otherwise noted, regulatory and other findings made previously for this study continue to be applicable.

Investigator’s Responsibilities:
When applicable, enclosed are stamped copies of approved consent documents and other recruitment materials. You must copy the stamped consent forms for use with subjects unless you have approval to do otherwise.

This study was reviewed in accordance with federal regulations governing human subjects research, including those found at 45 CFR 46 (Common Rule), 45 CFR 164 (HIPAA), 21 CFR 50 & 56 (FDA), and 40 CFR 26 (EPA), where applicable.

CC: Kathleen Anderson, Developmental & Learning Ctr
REFERENCES


Debra P. v. Turlington, 730 F.2d 1405 (11th Cir. 1984)


88


