A Comparison of Two Reading Programs on Third Grade Reading Achievement

Tawana D. Miller
Seton Hall University

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

Part of the Curriculum and Instruction Commons, and the Elementary Education and Teaching Commons

Recommended Citation
Miller, Tawana D., "A Comparison of Two Reading Programs on Third Grade Reading Achievement" (2008). Seton Hall University Dissertations and Theses (ETDs). 75.
https://scholarship.shu.edu/dissertations/75
A COMPARISON OF TWO READING PROGRAMS ON THIRD GRADE READING ACHIEVEMENT

BY

TAWANA D. MILLER

Dissertation Committee

Charles Achilles, Ed.D., Mentor
Patricia Guillory, Ph.D.
Barbara Strobert, Ed.D.
Samuel Taylor, Ph.D.

Submitted in Partial fulfillment of the Requirements for the Degree Doctor of Education Seton Hall University 2007
SETON HALL UNIVERSITY
COLLEGE OF EDUCATION AND HUMAN SERVICES
OFFICE OF GRADUATE STUDIES

APPROVAL FOR SUCCESSFUL DEFENSE

Doctoral Candidate, Tawana Miller, has successfully defended and made the required modifications to the text of the doctoral dissertation for the Ed.D. during this Fall Semester 2007.

DISSERTATION COMMITTEE
(please sign and date beside your name)

Mentor:
Dr. Charles Achilles

Committee Member:
Dr. Barbara Strobert

Committee Member:
Dr. Patricia Guillery

Committee Member:
Dr. Samuel Taylor

External Reader:

The mentor and any other committee members who wish to review revisions will sign and date this document only when revisions have been completed. Please return this form to the Office of Graduate Studies, where it will be placed in the candidate’s file and submit a copy with your final dissertation to be bound as page number two.
ABSTRACT

A COMPARISON OF TWO READING PROGRAMS ON THIRD GRADE READING ACHIEVEMENT

Background: The No Child Left Behind Act legislation requires both higher standards and higher success rates at the same time. Failure of students to learn to read adequately as shown by national and state test results and the need for continued school success had prompted education officials to implement two reading programs - Voyager Universal Literacy (VUL) and Harcourt Trophies (HT).

Purpose: The purpose for this study was to determine if any significant \((p \leq 0.05)\) differences existed between the reading comprehension performances of third-grade students using VUL and third-grade students using HT.

Setting: One large suburban school system in the Southeast United States.

Subjects: 465 third-grade students in five schools and 601 third-grade students in five schools.
**Treatment:** Students in five schools participated in VUL, and students in five schools participated in HT.

**Design:** Non-experimental cross-sectional quantitative study.

**Method:** Archived data normally collected at the district were organized and analyzed.

**Data Collection and Analysis:** State and national test data, data on gender, data on schools matched on socioeconomic status (SES) and students with disabilities (SWD) were used. Independent samples t-tests, ANOVA, ANCOVA and effect sizes were obtained using SPSS software.

**Results:** An effect size of .21 for CRCT indicated that students who received VUL improved about .2 standard deviation (SD) units more than did students who received HT.

**Implications:** Study results provided information to assist school administrators in choosing programs shown to be effective in developing proficient readers and increasing reading proficiency for students in NCLB SES and SWD subgroups.
Keywords: Achievement gap, alphabetic principle, SES, direct instruction, Dynamic Indicators of Basic Early Literacy Skills (DIBELS), fluency, reading comprehension and phonological awareness.
Acknowledgements

The task of actually completing a doctoral dissertation cannot be accurately described in words. This task has been my personal and professional goal for more than 20 years. It would be impossible to complete this task without the love, support, and encouragement of my immediate family. When one is writing a dissertation and working endlessly night after night on literature reviews and edits; it is easy to understand how God orders our steps. There were so many days that I existed on the sheer faith that this too shall pass. My four children, Mynxia, Adria, Celeste, and Craig were either getting married, starting careers, applying for medical school, entering college or having my first two grandsons during the course of my pursuit of a doctorate.

I must formally thank them for their patience, love, and support. I hope that I will always be an exemplary role model for them and they know that I will always be there to support them in any way, shape, or form necessary.

I extend a formal acknowledgement to my colleagues in Cohort IX. I cannot begin to say how proud I am to be
associated with such a fine collection of great minds and educators from across the nation and the world. If only we could exert enough effort to influence the policy makers and institutions of higher learning to try some of our solutions to the educational problems of the world. Maybe we could close that achievement gap, one school system at a time.

I retrace my steps to try to find the exact moment when I decided that Seton Hall University would be the right place for me. My decision was much easier due to the efforts and struggles of those who preceded me in life. I shall acknowledge my great-grandmothers, Inez Smallwood [Granny] and Savannah Winfrey [Mama Winfrey] who hold the distinction of being great, strong, caring African-American women who helped all of us make it through the stormy trials of life.

It is certainly the highlight of a lifetime to be in the same classroom with the great minds of Dr. Charles Achilles, Dr. James Caulfield, Dr. Anthony Colella, Dr. John Collins, Dr. Daniel Gutmore, Dr. Charles Mitchell, Dr. Mary Ruzieck, Dr. Joseph Stetar, Dr. Barbara Strobert, Dr.
Elaine Walker and Dr. Kevin Walsh. It was a pleasure as well as a delightful intellectual journey to be able to benefit and grow from their guidance, advice, and encouragement during my short tenure at Seton Hall University.

The support and encouragement of my colleagues, Dr. Annie Duvali, Mrs. Iris Moran, Dr. Patricia Guillory, Dr. William Hammond, Dr. Antoinette McGlasker, Dr. Demetria Taylor, and Dr. Samuel Taylor were immeasurable. Their words of encouragement and gentle pushes helped me to begin and complete the final journey towards the attainment of the doctorate degree. I value their opinions highly and will be forever grateful for their faith in my abilities to stretch my mind to the farthest reaches of the educational arena. My life will be changed dramatically for the duration of my time here on earth.

I must acknowledge the gentle yet firm pushing and nudging that I received from my mentor Dr. Charles Achilles. I would not have completed this study had it not been for his persistence and dedication to excellence and the pursuit of meaningful research. I have learned so much
as his "Southern apprentice" as I would jokingly refer to myself whenever he would knock the wind out of my sails to go back to the drawing board to consider the findings from another study or to add additional data to strengthen my findings. I only wish that some day I could hold in my brain a fraction of all of the research and information that he effortlessly recalls at any given moment. I am truly humbled to have Dr. Achilles as my mentor and I treasure the opportunity to have studied under his watchful eye.

I am eternally grateful for the love, prayers and thoughts of Jerald, Justin, Eric, Velda and everyone else who touched my life in a meaningful way.

Lastly, I wish to acknowledge my husband of 30 years, Kenny Miller. You have been so patient, loving, and always there for me. How many degrees, courses and papers did you wait patiently for me to complete? I must say to you with all of my heart that it is finally over, and you will not have to wait for me any longer. Thank you for waiting on me all of those years and loving me during this time.
DEDICATION

This is dedicated to the memory of my maternal grandmother, Ella Pendergraft, who never had the opportunity to go to college. I was her first grandchild and the first college graduate in the family. "Wyla", as she was affectionately called, was a constant source of encouragement and faith for me. She passed in September, 2003, but her spiritual presence helped me to complete this task. She always believed in me, and I was determined to complete this degree for her.

This dedication also extends to my parents, George Muse Derricotte Jr, and Ann Winfrey Derricotte, who have been the guiding light in my life, steadfast supporters and ever present to step in at a moment's notice to push me forward. Their steadfast love has been my strength whenever life was getting the best of me. I was blessed to be their child. I hope they realize that it is an honor to have them for my parents. I will never be able to love them enough for all they have sacrificed and given me.

With all of my love,

Tawana Myngtria Derricotte Miller, Atlanta, Georgia
TABLE OF CONTENTS

LIST OF TABLES.................................................................xv
LIST OF FIGURES..............................................................xviii

I  INTRODUCTION AND BACKGROUND......................................1

Background.................................................................4
Statement of the Problem...............................................5
Purpose.............................................................................7
Significance of the Study...............................................8
Overview of the Study and Setting..................................9
Reading Research Overview.............................................10
Research Questions and Hypotheses.................................14
  Research Questions......................................................14
  Null Hypotheses..........................................................15
Variables............................................................................15
Delimitations.....................................................................16
Limitations.........................................................................17
Definition of Terms..........................................................18
Summary of Research Design and Methods..........................20
Summary.............................................................................21
  Subjects...........................................................................21
  Treatment.........................................................................21
Relevance of the Study.....................................................22
Organization of Study.......................................................22

II  REVIEW OF RELEVANT RESEARCH, THEORY AND LITERATURE.24

Congressional Charge........................................................26
  Reading First and the National Reading Panel....................27
  Findings and Critiques of the NRP....................................30
  Scientifically Based Reading Research (SBRR)....................40
Definitions of SBRR..........................................................40
Scientifically Based Reading Studies...................................44
A Framework for Reading Comprehension Instruction............49
  Psycholinguistic Theory..................................................50
  Neuroscience and reading instruction.............................57
Reading Comprehension Approaches...................................59
  Phonemic approach and Direct Instruction.......................64
  Literacy-based approach...............................................68
  Reading Instruction Research.........................................72
Review of Reading research and struggling readers.............80
Fluency and Phonological Awareness Instruction................................. 83
Purpose of the two Reading Programs as Related to Research................... 92
Voyager Universal Literacy - Supplemental or Core.................................. 94
Harcourt Trophies - Comprehensive Core Reading Program.......................... 95
Goals of Reading Research........................................................................ 96
Why is Reading research needed to generate or validate a reading model?........ 98
Why is Reading research necessary to make instructional decisions?.............. 98

III RESEARCH DESIGN AND METHODOLOGY.............................................. 101
Participants................................................................. 104
Instrumentation............................................................. 106
Georgia Criterion Referenced Competency Tests (CRCT)................................ 107
Iowa Tests of Basic Skills (ITBS)................................................. 108
Procedures.............................................................................. 109
Methodology............................................................................ 112
Voyager Universal Literacy and Harcourt Trophies strategies........................ 114
Rationale for Study and Data Selected..................................................... 115
Validity and Reliability of the Instruments................................................ 116
Summary..................................................................................... 116

IV DATA AND RESULTS OF DATA ANALYSIS.......................................... 118
Setting for the Study.......................................................................... 119
Data Treatment Presented and Analyzed.................................................... 124
Analyses of Variance (ANOVA)............................................................. 132
Interview with Voyager District Coach...................................................... 146
Summary......................................................................................... 148

V INTRODUCTION, SUMMARY, DISCUSSIONS, CONCLUSIONS AND RECOMMENDATIONS................................................................. 150
Summary of findings............................................................................. 156
Pre-test Scores.................................................................................. 156
Post-test Scores................................................................................ 157
ANOVA Results for post-test 2005-06 CRCT......................................... 162
Limitations of the Present Study........................................................... 166
Implications for policies, practice and future research.............................. 167
Conclusions and/or Discussions...............................169
References...................................................175

Appendix A - Approval Letters
    Superintendent..........................................224
    Test Coordinator......................................227
Appendix B - Institutional Review Board Non-Review Letter.................................229
Appendix C - Glossary.....................................231
Appendix D - Standards Setting, Cutscores and Scale Scores for the Georgia CRCT...........241
Appendix E - Standard Scores for Iowa Tests of Basic Skills (ITBS)............................245
Appendix F - Profile Plots for reading programs, gender, Socioeconomic status (SES) and students with disabilities (SWD).........................246
Appendix G - Interview with Voyager District Coach.......................................................249
Appendix H - Critique of VUL/HT results using Crane’s (1998) Criteria......................258
LIST OF TABLES

1. Summary of critiques of the National Reading Panel...33

2. Partial list of Georgia State-approved Reading and Language Arts Textbooks........................................39

3. Summary of Historic Reading Research reports
   (1884-2007)..................................................................63

4. Authors and Findings for Six U.S. Reading Research Syntheses Studies......................................................74

5. 3rd grade students - Frequency and Percentage of Students in System and two study groups.............121

6. Student Percentages of Free/Reduced, Mobility (Mobil) and Students with Disabilities (SWD) being compared in Georgia study.........................................................123

7. Comparison of Pre- and Post-test CRCT Assessment Means and Differences for HT and VUL...............125

7.1 Estimated Marginal and Pairwise Comparison of HT and VUL Means for CRCT 2006.............................128

8. Comparison of Pre- and Post-test CRCT and ITBS Means and Differences for HT and VUL......................129

8.1 Estimated Marginal and Pairwise Comparison of HT and VUL Means for ITBS 2005.........................131

9. Between Subject Factors (1066 students, 2005-06) for Analysis of Variance of CRCT.........................133
10. Summary of 2006 Means and Standard Deviations........134
11. Tests of Between-Subjects Effects - Gender............136
12. Tests of Between-Subjects Effects - SES..............137
13. Tests of Between-Subjects Effects - SWD.............138
14. Comparison of Progress between HT and VUL cohorts on CRCT 2006 means and differences.............140
15. Comparison of Reading Programs by schools, 2005
   CRCT Reading Means, ITBS Reading Means, 2006 CRCT
   Reading Means and Mean Differences from sample
   Means......................................................142
16. Comparison of Reading Programs by schools and Gender,
   2006 CRCT Reading Means and Mean Differences from
   sample means............................................143
17. Comparison of Reading Programs by schools and
   Free/Reduced Lunch - SES, 2006 CRCT Reading Means and
   Mean Differences from sample means....................144
18. Comparison of Reading Programs by schools and
   Students with Disabilities (SWD), 2006 CRCT Reading
   Means and Mean Differences from sample means........145
19. Comparison of Reading Programs by Mean Differences
   and 2005 CRCT pre-test and 2006 CRCT post-test........158
20. Comparison of Reading Programs by Mean Differences
    and Percentage of Level 1 students....................159
21 Comparison of progress between HT and VUL cohorts on 2006 CRCT Means and Differences ..................163
22 Influence of VUL on reading performance on GA CRCT.164
23 Cutscores for the 2006 CRCT for GPS-based Tests in grades 1-4..................................................243
24 Performance levels and Scale Scores for the CRCT...244
LIST OF FIGURES

1. Modeling the reading system: Four processes........53
2. Domains of information and reading...............55
3. Work Process flowchart............................111
Chapter I

Introduction

Huey (1908/1968) realized almost a century ago that...to completely analyze what we do when we read would almost be the acme of a psychologist's dream for it would be to describe very many of the most intricate workings of the human mind, as well as to unravel the tangled story of the most remarkable specific performance that civilization has learned in all its history. (p. 8 as cited in Anderson & Pearson, 2002, p. 255)

The Elementary and Secondary Education Act of 1965 (ESEA,) as amended by the No Child Left Behind (NCLB) Act signed by President Bush on January 8, 2002, required that all children be proficient in reading and mathematics by the 2013-2014 school year. The NCLB amendment established the Reading First (RF) program in an effort to meet the goal of reading proficiency for all children and closing the achievement gap. To help meet this lofty goal, Congress charged the Director of the National Institute of Child Health and Human Development (NICHD), in consultation with
the Secretary of Education, to convene a National Panel. The NICHD work exemplified by the National Reading Panel (NRF) was to assess the status of research-based knowledge, including the effectiveness of various approaches to teaching children to read (NICHD, 2000). Complaints from small publishers who were prevented from selling products to RF schools to complaints of important large researchers whose programs were completely shut out of RF schools led to a series of six reports from the Inspector General at the U. S. Education Department.

The first report (2006) offered a scathing critique of the Education Department's oversight of the RF grant-application process. It concluded that federal officials seemed biased toward a particular instructional approach, direct instruction, and did not screen consultants for potential conflicts of interest (Manzo, 2007a). The Inspector General’s report also concluded that a federal contractor overseeing states’ implementation of the RF program may have inappropriately guided at least two states toward adopting a specific reading assessment [DIBELS] (Manzo, 2007b). The Inspector General’s final audit concluded that 43 states indicated in their RF
applications that they intended to use DIBELS as one of their assessment test instruments. Out of this quagmire rose the business of developing resources to serve the needs of districts working toward meeting the requirements of NCLB.

School personnel were inundated with claims from large and small publishers, vendors, and test companies that their products were rooted in scientifically based research (SBR) and would close the achievement gap while helping systems meet Adequate Yearly Progress (AYP) according to NCLB legislation. One question district leaders had to answer was: "What criteria should be used when considering selection of reading programs?"

Since the implementation of NCLB, schools have failed to make AYP goals largely due to low scores of students within NCLB subgroups of low SES, students with disabilities (SWD), and English language learners (ELL). Nationally, the percent of rated schools making AYP, based on data released in 2006, dropped from 75% to 71%. The percent of rated schools in need of improvement increased from 13% to 17% (Olson, 2006). These data are alarming in
light of the requirement to have both higher standards and higher success rates at the same time (Jenkins, 2005).

Background of the Problem

Children who struggle in vain with reading in the first grade soon decide that they neither like nor want to read (Juel, 1988). Early childhood educators realize the critical importance of making sure that as many children as possible learn to read in the primary grades. Achieving this goal is important because the more children who become proficient readers at an early age, the fewer children would be retained or require special education services/remediation, perhaps resulting in cost savings. Equally important, school systems would reduce their exposure to consequences for students failing to make AYP.

In spite of the implementation of initiatives to answer NCLB requirements, achievement gaps remain large and are persistently associated with such factors as ethnicity, poverty, gender and students with disabilities (SWD).

Substantial gaps are evident for disadvantaged children in measures of reading and mathematics proficiency, in pro-social behaviors and behavior problems, and in readiness to learn. As a result, many children from
disadvantaged backgrounds fail to meet grade-level expectations on core subjects. (Rand, 2005)

Clearly, the situation described above raises the need to know what approach to elementary reading instruction will develop proficient readers as determined by the required state and national assessments.

Statement of the Problem

Some education critics perceive that a national reading crisis exists in the United States. The percentage of the population that reads literature regularly is an important measure of adult literacy. The percentage of adults age 25 or older who reported reading any literature in the past 12 months declined between 1982 and 2002 from 56% to 47%, with most of the decrease occurring between 1992 and 2002 (National Center For Education Statistics [NCES], 2005). According to one source, 68% of American eighth graders read below the proficient level and more than 50% of African American and Hispanic eighth graders are not proficient; 82% of prison inmates are high school dropouts and a high proportion cannot read (NCES).

In Georgia, 42% of fourth graders scored below basic in reading on the National Assessment of Educational
Progress (NAEP) as compared to a national average of 38% of fourth graders scoring below basic in reading. Fifty-seven percent of Georgia’s economically disadvantaged fourth graders scored below basic in reading on the NAEP as well. Fulton County, which contains the schools in this study, did not report NAEP data for fourth-grade students, but those schools continue to report scores significantly lower than the system’s averages.

In 2002, in an effort to address the issue of low reading performances of elementary students in selected high poverty schools, leaders in the Fulton County School System (FCSS) supplemented the district-wide Harcourt Trophies (HT) Reading Program with the Voyager Universal Literacy (VUL) Program. Students in VUL participated in a small group, primarily scripted, phonics-based reading program. Students in HT participated in a small group literacy-based, integrated reading program. Subsequently, the district conducted system-wide program evaluations on VUL and HT to answer the following question: “Should FCSS leaders continue to invest in VUL or should the system continue to use HT exclusively in all of its elementary schools?”
Purpose

The purpose of this study was to determine if any statistically significant \((p \leq 0.05)\) differences existed between the reading comprehension performances of third-grade students using VUL and third-grade students using HT in relationship to students' SES, gender, and SWD to help the system administrators make important financial decisions.

Study results should provide information to help school administrators choose programs that are shown to be effective in developing proficient readers and increasing reading proficiency for students in the No Child Left Behind (NCLB) (2002) subgroups of low socioeconomic status (SES) [also referred to as economically disadvantaged (ED)] and students with disabilities (SWD). The superintendent and test coordinator of the system provided written authorizations to use the test data (see Appendix A). Letters and Institutional Review Board (IRB) Non-Review Certification approvals are in Appendix B.
Significance of the Study

Reading is the foundation of success in life and success in society. Yet, after several years of scientifically based research (SBR) in reading, a disturbing proportion of children still fail to read in the United States. Major issues such as the ones listed below are evident at the national, state, and local levels:

1. Large numbers of school-age children, including children from all social classes, have significant difficulties in learning to read.

2. Failure to learn to read adequately for continued school success is likely among poor children.

3. An increasing proportion of children in American schools are learning disabled (LD) with most of the children identified as such because of difficulties in learning to read.

4. Even as federal and state governments and local communities invest at higher levels in early childhood education for children with special needs and for those from families living in poverty, these investments are often made without specific planning to address early
literacy needs and sustain the investment. (Snow, Burns, & Griffin, 1998)

The issues in developing proficient readers early are wide-ranging and further studies were necessary to assist educators in identifying SBR reading programs and instructional practices.

Overview of the Study and Setting

This investigator seeks to add to the base of knowledge relating to the efficacy of two reading programs on improving reading comprehension performances of third graders. The results of two state assessments and one nationally mandated assessment were analyzed to ascertain whether statistically significant (p ≤ .05) differences existed between the reading performances of students who received instruction via the VUL program as compared to instruction delivered via the HT reading program.

This study was conducted in 10 elementary schools in a single school system in Georgia during the 2005-2006 school year. One of the 10 elementary schools used the treatment reading program in two third-grade class sections and the control reading program in two additional third-grade class sections. Each elementary school had between four and eight
sections of third graders with a grade level enrollment of 61 to 167 students. Teachers of these students were trained to implement either the VUL or HT program, which they taught daily for 2½ hour blocks.

Reading Research Overview

The NCLB Act required all approaches to academic achievement meet standards of scientific based research (SBR). The National Reading Panel (NRP) used the following guidelines to determine which studies met the standards of evidence:

1. Research must address achievement in one or more skills in reading.

2. It must be generalizable to the larger population of students.

3. The research needs to examine the effectiveness of an approach by comparison with other types of instruction.

4. Other scholars from the field must review the research and consider it high quality. (Reading First Support, 2004)

Scientifically-based reading research (SBRR) uses rigorous, systematic, and objective procedures to obtain knowledge about reading difficulties. This type of research
involves controlled experiments with data analysis and a thorough peer-review process. This study examined student test-score outcomes from two reading programs that advertise to schools that their programs are based upon SBRR.

VUL is referred to as a comprehensive reading system that includes an integrated curriculum for reading, writing, intervention, enrichment, strategies for English language learners (ELL), extended day, home study, technology, DIBELS-like assessment, and initial and ongoing professional development. Daily lessons begin with whole-group teacher-directed instruction and systematic instruction of writing, spelling, or phonemic awareness. In two of the three daily reading stations, students work collaboratively while in the third station, the teacher follows a detailed lesson plan for small-group instruction in phonemic awareness, phonics, fluency, vocabulary and comprehension.

The research that supports VUL's skills-based design noted that children who recognize words more readily are able to focus more attention to the meaning of the words (Chall, 1996; Dowhower, 1987; Ehri, 1977; Ehri, 1995; Ehrlich,
Curtz-Costes, & Loridant, 1993; Goodman, Haith, Guttentag, & Rao, 1985; Guttentag, 1984; Guttentag & Haith, 1978; Guttentag & Haith, 1980; Craut & Smothergill, 1980; Rosinski, 1977; Perfetti, 1985; Samuels, Scherner & Reinking, 1992). However, no independent research was found that compared VUL to HT.

HT is a comprehensive reading program that emphasizes oral language, comprehension, vocabulary, reading, writing, grammar, spelling, phonemic awareness, systematic phonics, fluency, assessment, listening, speaking, an Intervention Resource Kit, and ELL Resource Kit. Daily lessons begin with a read-aloud/question of the day, followed by comprehension/decoding phonics, guided comprehension, independent reading and writing/grammar/spelling. HT does not have a specific professional development component.

The data collected from studies examined by the NRP suggest that text comprehension is enhanced when readers actively relate ideas presented in print to their own knowledge and experiences and construct mental representations in memory (NRP, 2000). Therefore, a relationship exists between the thinking process and the reading text. This relationship is often referred to as
text comprehension. Several reading researchers have developed theories related to text comprehension.

Comprehension is defined as "intentional thinking during which meaning is constructed through interactions between text and reader" (Harris & Hodges, 1995, p. 207). Morrow and Gambrell (2002) stated that literature-based reading instruction is perhaps more closely associated with reader-response theory which explains how readers interpret literature (McGee, 1992). Meaning in the text is constructed by the readers' own interpretation of their experiences while they are reading (Rosenblatt, 1978). The reader uses prior experiences to select images and feelings that allow the reader to shape the text at the same time that the text shapes the reader by creating new experiences (McGee, 1992; Rosenblatt, 1978, 1991).

Mullis, Campbell, and Farstrup (1993) stated that the 1992 NAEP findings revealed that teachers reported a "heavy" emphasis on literature-based instruction, and that the students of these teachers displayed higher levels of reading proficiency. Further studies revealed that children in literacy-based (whole language) classrooms typically do as well or better on standardized tests than students in

The research discussed above is just a sample of that which supports the design of the HT reading program. The research base found in the 2000 Report of the NRP also supports these two programs’ claims.

Research Questions and Hypotheses

The research questions and null hypotheses of this study are listed below. Any differences were tested for significance at $p \leq 0.05$. Effect sizes were computed to show educational importance. Two research questions were derived from the conceptual framework and the research design of the study.

Research Questions

1. What difference does VUL or HT reading instruction have on reading achievement of third-grade students as measured by state and national assessments?

2. What difference does VUL or HT reading instruction have on reading achievement of third-grade students in Fulton County, Georgia using (a) socioeconomic status (SES), (b) gender, or (c) students with disabilities (SWD) as the main effects measured by state assessments?
Null Hypotheses

1. There is no difference in the reading comprehension performances of third-grade VUL students when compared to the performances of third-grade HT students as measured by the Iowa Tests of Basic Skills (ITBS).

2. There is no difference in the reading comprehension performances of third-grade VUL students when compared to the performances of third-grade HT students as measured on the Georgia Criterion Referenced Competency Tests (CRCT).

3. There is no difference in the reading comprehension performances of third-grade VUL students when compared to the performances of third-grade HT students using (a) socioeconomic status (SES), (b) gender, or (c) students with disabilities (SWD) as the main effects measured by the CRCT.

Variables

The two categories for the independent variable of reading programs are 1) Voyager Universal Literacy (VUL) and 2) Harcourt Trophies (HT). The three categories of student tests outcomes within the independent variables, each coded 1 or 2, are: (1) gender - male or female; (2)
low SES - Free/Reduced lunch, yes or no; and (3) students with disabilities (SWD), yes or no.

The dependent variable for the research questions and null hypotheses is reading comprehension as shown by two indices: (1) percentage of students meeting or exceeding state proficiency (Level 2 or 3) performance as determined by scale scores from the CRCT; and (2) reading comprehension as determined by standard scores meeting or exceeding the 50th percentile from the ITBS.

Delimitations

1. The study was delimited to third-grade students in 10 schools in South Fulton County, GA in regular education classrooms.

2. The investigator did not include third-grade students in self-contained special education classes whose Individualized Education Plans (IEP) prescribe the use of state-mandated alternate assessments.

3. Some special education students who received interrelated services were on the third-grade class rosters and their test scores on state criterion-referenced tests (CRT) and norm-referenced tests (NRT) were included in the class averages.
Limitations

1. The study was limited by ethnicity because the majority of students were African-American; therefore, ethnicity was not used as an independent variable.

2. The small school sample size [n=10], student mobility rates, pre-assignment of student groups, and pre-established teacher assignments limit the ability to generalize results.

3. HT has been used since 2003-2004 compared to VUL which has been used since 2004-2005 with third-?grade students. The difference in length of usage between the two reading programs may influence the results.

4. The ITBS test was administered in September, 2005, and fewer students tested at this time when compared to the number tested on the CRCT in April, 2006.

5. Implementation with fidelity and integrity was beyond the researcher’s control and may impact findings.

6. Using group data will not provide as valid an assessment as a design that would embed students into class units, recognizing the influence of teacher, peers, and setting (etc.) on student achievement.

7. Another limitation is the cross-sectional design
because it becomes difficult to establish time order (condition 2 of the necessary conditions for causality). If data are collected from research participants at a single point only, there can be no direct measurable changes that occur over time. (University Of South Alabama; Sociology Index, 2002)

8. Students are not randomly assigned to either VUL or HT but the groups are available and accessible in closely matched schools in the same district.

9. The researcher was an administrator in the district at the time of the study which may indicate the potential for researcher bias.

Definition of Terms

Definitions of terms and precise use of the terms are key for understanding research. For the purposes of this study, terms are defined in two groups: a. terms specific to reading are included in a Glossary (see Appendix C) and b. terms directly related to this study as presented here.

Key Definitions:

Georgia Criterion Referenced Competency Tests (CRCT) are state-mandated achievement tests for students in grades
one through eight that cover the curricular areas of reading, English/language arts, and mathematics.

_Harcourt Trophies (HT)_ is a developmental reading/language arts program. Explicit phonics instruction, direct reading instruction, guided reading strategies; phonemic awareness instruction; systematic intervention strategies; integrated language arts components; and assessment tools used to ensure student success.

_Iowa Tests of Basic Skills (ITBS)_ are national norm-referenced tests in reading, mathematics, social studies, and science. The ITBS are standardized tests given annually to many school students in the United States. These tests are given to students beginning in kindergarten and progressing until Grade 8 to assess educational development. The ITBS covers numerous areas of educational knowledge, including English language use, math concepts and problem solving, social studies, science, geography and map use, and source use (Wikipedia, 2006).

_Voyager Universal Literacy (VUL)_ is a comprehensive K-3 reading system that includes an integrated curriculum for the following components: reading, writing and language
arts, intervention, home study, strategies for English language learners, technology, and initial and ongoing professional development (Florida Center for Reading Research, 2002).

Summary of Research Design and Methods

This research was a non-experimental cross-sectional quantitative study (Johnson, 2001). The design contains, to the extent possible, data regarding matched schools, gender, socioeconomic status (SES), students with disabilities (SWD), state and national test data. The researcher used test data normally collected at the district level that were available at the time of the study. Independent samples t-tests, ANOVA, ANCOVA and calculations for effect sizes were included using SPSS software. Archived data were organized and analyzed around hypothesis testing. To determine whether VUL or HT was more effective, the CRCT scale reading scores and ITBS percentage reading scores of both samples were compared for statistical significance ($p \leq .05$).
Summary

Subjects

The test results from third-grade students from 10 different K-5 FCSS elementary schools were analyzed. There were 465 third-grade VUL students and 601 third-grade HT students in this study. Students in both groups were selected based upon their reading performance on the CRCT and ITBS scores.

Treatment

Distinctions between the two reading programs are listed below:

1. Scripted, skill-based program (VUL) versus meaning based program (HT);
2. Vital Indicators of progress (VIP) which are DIBELS-like assessments used in VUL;
3. Direct Instruction Model used in VUL;
4. Balanced approach to reading instruction in HT
5. Minimum of 80 hours professional development for teachers in VUL;

Research supports reading programs using direct instruction, balanced approaches to reading, and meaning-based reading programs. However, other than the program
evaluations completed by FCSS, the researcher could find no independent research studies comparing HT and VUL.

The CRCT and ITBS were administered in the same manner to all students in the schools during the program delivery. These test results were the data used to determine reading comprehension success. ITBS and CRCT are valid and reliable measures of reading comprehension. Means of ITBS and CRCT were calculated and compared using independent t-tests to determine if any differences between gains for the two samples were statistically significant ($p \leq 0.05$).

Relevance of the Study

The basic purpose of program evaluation research is to test (1) whether education programs help the students they are designed to serve and (2) whether new ideas for education programs are worthy of extension to a wider selection of schools and settings.

Organization of the Study

Chapter I - Included an introduction to the problem of low reading performance of third-grade students in high poverty schools. The study will provide information to help school administrators choose programs that are shown to be effective in developing proficient readers and increasing
reading proficiency for students in the NCLB (2002) subgroups of low SES or Economically Disadvantaged (ED) and SWD.

Chapter II - Discusses the research, literature and theoretical framework related to reading acquisition specifically related to low SES and SWD subgroups.

Chapter III - Discusses the methodology used in the program evaluation. The researcher used test data normally collected at the district level that were available at the time of the study.

Chapter IV - Describes the data and statistical analyses used on these data.

Chapter V - Provides a summary of study findings, discussions, limitations, conclusions and recommendations for policy practice and additional research.
Chapter II

Review of Relevant Research, Theory and Literature

The review of important research, theory and literature related to this investigation encompasses studies and articles from 1783 to 2006. The review is divided into six sections: (1) Congressional Charge, (2) A Framework for Reading Comprehension Instruction, (3) Reading Comprehension Approaches, (4) Reading Research and struggling readers, (5) Purpose of two reading programs as related to the research, and (6) Goals of Reading Research. The review culminates in the theoretic framework to guide the present study.

Journal searches were completed at Seton Hall University (SHU) or from major research institutions located in Atlanta, Georgia. Reading achievement in primary grades, reading comprehension and DIBELS, DIBELS, reading programs, reading comprehension and struggling readers and reading comprehension and third grade were the successful descriptors. Use of the reference sections of prior dissertations, various publications and suggestions from dissertation committee members provided studies and
information of relevance to the research topic. The genesis of this research is a derivative of the No Child Left Behind (NCLB) Act signed January 8, 2002.

In September 2001, President Bush expressed the goal that “no child should be left behind” because he or she cannot read (Sweet, 2004). The lofty goal was an adaptation of the long-time motto of the Children’s Defense Fund (CDF) directed by M. W. Edelman: “Leave No Child Behind” (Children's Defense Fund, 2006). In 2001, CDF launched a five-year campaign to make children a national priority. At its core is the landmark Act to Leave No Child Behind, which Senator Chris Dodd and Representative George Miller introduced in Congress on that same day, cosponsored by 95 of their House and Senate colleagues. This Act was likely the antecedent to NCLB.

Reading is...the most basic educational skill, and the most basic obligation of a school is to teach reading. Yet, tests [2001] showed that almost two-thirds of African-American children in the fourth grade cannot read at a basic level, and reading performance overall is basically unimproved over the past 10 years. (The White House, 2002)
President Bush stated that based upon sound research, instruction would be supported with a central role for phonics.

Congressional Charge

In 1997, Congress charged the Director of the National Institute of Child Health and Human Development (NICHD), in consultation with the Secretary of Education, to convene a national panel to assess the status of research-based knowledge (NICHD, 2000). The panel, referred to as the National Reading Panel (NRP) developed guidelines to determine which studies met the scientific standard for evidence of instructional efficacy.

The Congressional charge was never intended to imply that the NRP would endorse, approve, or sanction any particular reading curriculum. As stated by Wilhoit, executive director of the Council of Chief State School Officers:

The law [NCLB] said nothing about picking specific programs, it just indicated scientifically based programs. But when we looked at the other programs that were being approved, we saw very little evidence that those were more scientific than the ones we were
trying to use. (as cited in Manzo, 2007, p. 19)

However, the Elementary and Secondary Act, P. L. 89-10, 1965, (ESEA) as amended by NCLB established the RF program to assist state and local agencies in establishing reading programs that were based on scientifically based reading research (SBRR) (Office Of Inspector General, 2006).

Reading First and the National Reading Panel

In 1993, The National Right to Read Foundation (NRRF) was established to promote evidence-based instruction in reading (Sweet, 2004). NRRF focused its efforts on advancing the cause of evidence-based reading instruction. Under President Clinton's watch and in conjunction with heightened attention on the reading deficit, the Reading Excellence Act passed the House in late fall of 1998 (Sweet, 2004). The Reading Excellence Act introduced and defined the term scientifically-based reading instruction (SBRI). The definition of SBRI as presented in the Reading Excellence Act is included in the glossary in Appendix C. The Reading Excellence Act existed for only three years but laid a solid foundation for Reading First (RF), SBRR, and instruction.
Prior to the formation of RF, the National Research Council (NRC) released a consensus report *Preventing Reading Difficulties in Young Children* (PRD; Snow, Burns, & Griffin, 1998), which signaled an attempt to end the so-called "reading wars" that had been raging for decades. The "reading wars" existed between researchers and practitioners using a phonemic approach to reading and researchers and practitioners using a whole language approach to reading.

The conflicts among reading researchers was moderated, if not eliminated, by the realization that all participants were primarily interested in ensuring the well-being of young children and promoting optimal literacy instruction (Snow, et al., 1998, pp. v-vi). The conclusion was:

All members agree that reading should be defined as a process of getting meaning from print, using knowledge about the written alphabet and about the sound structure of oral language for the purpose of achieving understanding. All thus also agreed that early reading instruction should include direct teaching of information about sound-symbol relationships to children.
who do not know about them. And that it must also maintain a focus on the communicative purposes and personal value of reading. (Snow, et al., 1998, pp. v-vi)

Just when the smoke had seemingly settled on the "reading wars" battlefield, the behind-the-scenes battle was pushed to the forefront as nationally known researchers and government officials were asked to resign their positions. A report released May 9, 2007 by Senator Edward M. Kennedy, D-Mass., charged that Edward Kame'enui, a central figure in the SBRI research, while serving as a high-level federal advisor to states, was also under contract to the publishers of his early-reading program to promote the product (as reported by Cavanaugh, 2007). Kame'enui received consulting fees from his work designing Voyager Expanded Learning, one of the reading programs often funded through RF grants (Manzo, 2007c). Kame'enui appeared before the House Education committee on April 20, 2007 regarding involvement in possible conflicts of interest procedures as noted by the Inspector General and the Government Accountability Office (Kame'enui, 2007).
The Office of Inspector General (OIG) released six RF reports from September 2006 to March 2007 in response to complaints about mismanagement of the federal RF program. Specifically the Department of Education:

1. Intervened to influence a state’s selection of reading programs; and

2. Intervened to influence reading programs being used by local educational agencies (LEAs) after the application process was completed (OIG, 2006).

Some programs approved in RF applications were more often the same programs [Voyager Expanded Learning] that had been developed in part by Kame’enui. From all indications, educators were strongly encouraged to use direct instruction and phonemic reading programs rather than risk losing RF funds. All of this to say that the debate regarding which method of reading instruction, and/or reading program was preferable continued. Several researchers have praised or admonished the work of the NRP. Findings and Critiques of the National Reading Panel

The NRP report was written at the request of the U.S. Congress and supported by the NICHD in cooperation with the U. S. Department of Education (Shanahan, 2004). The report
was to provide a summary of reading research that could be used to support federal education policy.

The teaching of reading is complex and the panel decided that topics had to be limited. The NRP limited its analysis to studies that reported some form of an experiment (Shanahan, 2004). The NRP examined the instructional practices of phonemic awareness, phonics, oral reading fluency, encouraging children to read, vocabulary, comprehension, teacher education, and technology. Analyses of research indicated that instruction in the areas of phonemic awareness, phonics, fluency, vocabulary and comprehension increased reading achievement for students. The NRP effort was not without numerous critics. Potentially the most significant criticisms of the NRP are those that challenge the basic premises of the panel, as these comments could expose reasons that the NRP should not influence practice or policy (Shanahan, 2004). Major criticisms of the NRP findings are listed in Table 1. Some criticisms were aimed directly at the NRP panelists while other criticisms involved the noted exclusion by the
NRP of specific instructional strategies, methodologies, or qualitative reading research studies.
### Table 1

**Summary of critiques of the National Reading Panel**

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Year</th>
<th>Criticism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yatzin; Coles;</td>
<td>2000</td>
<td>Some important reading topics such as language and literature have been neglected.</td>
</tr>
<tr>
<td>Edmondson &amp;</td>
<td>2001;2002</td>
<td>A second widespread complaint has concerned the panel's decision to examine only experimental or quasi-experimental research when evaluating the effectiveness of instruction. (Note that only 38 out of 438 studies were used for the phonics subgroup of the NRP report)</td>
</tr>
<tr>
<td>Shannon; Garan;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newkirk; Pressley;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shaker; Heilman</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krashan</td>
<td>2001</td>
<td>The panel should not have opposed encouraging children to read. Just letting kids read can be better than instruction; a finding of no difference is actually evidence for free reading in classrooms.</td>
</tr>
<tr>
<td>Garan</td>
<td>2002</td>
<td>Most of the panel &quot;had never been directly involved in actually teaching children to read but came from an unusual assortment of occupations&quot; (2002, p.3).</td>
</tr>
<tr>
<td>Shanahan</td>
<td>2004</td>
<td>A surprising amount of criticism has taken the form of saying bad things about panelists, making erroneous claims about what the panel did, and complaining about how the NRP findings will be misused.</td>
</tr>
<tr>
<td>Allington</td>
<td>2002</td>
<td>Allington (2002) challenged the entire enterprise of applying research findings of this type to instructional practice. Allington attacked the very idea of promoting instructional approaches that have been found to be effective in research. He also asserted that using an &quot;external expert intervention model to research how to teach reading is an enormously flawed approach&quot; (pp. 90-91).</td>
</tr>
<tr>
<td>Kohn</td>
<td>2006</td>
<td>Kohn defined selective reliance as ... &quot;the tendency to invoke or ignore research selectively, depending on whether it supports ideas one happens to like... by making use of research only on certain topics or by citing only certain studies for a given topic.&quot; Kohn stated that &quot;There is ample reason to doubt, for example, whether an approach consisting mostly of direct systematic instruction in phonic skills accurately reflects the best available scientific findings - unless the term &quot;scientific&quot; is recast so as to exclude all data except those that support this position (pp.8-22, Kohn, 2006).</td>
</tr>
</tbody>
</table>
Allington (2005) stated that overall

I think I've been less critical of the full report of the National Reading Panel than Garan, Krashen, and other critics. In my view, the NRP was under-funded, understaffed, and given a timeline that was far too short. Consequently, the panel members rushed a report that still arrived more than a year after the initial deadline. (p. 462)

Garan stated that there was a substantial mismatch between what the NRP actually found and what the NRP report summary said the panel found as repeated by Allington, (2005). Most educators probably read the summary rather than the full report which could be the cause of some misinformation or misinterpretations of the NRP report. Errors in reporting the findings reflect a simple ideological bias in favor of a particular sort of reading instruction for beginning readers and for struggling readers -- the sort of reading instruction that the full NRP report doggedly avoided recommending (Allington, p. 462).
Yatzin was the only elementary teacher member of the NRP panel. Yatzin's most critical complaint, according to Cowen (2003), was that the NRP did not fully respond to Congress's original charge to research the basic process by which children learn to read. Yatzin stated that Congress gave the NRP an unrealistic timeline and the Panel's research findings and conclusions were not critiqued by teachers prior to being released to the public. Some of the claims are philosophical [e.g., Yatzin's claim as stated above]...while other claims are largely methodological and are easier to refute on the basis of widely accepted standards and norms for synthesizing research data (Shanahan, 2004, p. 261).

The NRP reported that systematic phonics instruction was most effective when provided within the context of a comprehensive reading program that also addressed phonemic awareness, fluency, vocabulary, and comprehension strategies. These conclusions were based on a collection of similar findings replicated across several research studies conducted over several years (Lyon & Chhabra, 2004).

The analysis of data from the phonics portion of the NRP study is the most controversial portion of the NRP
report according to Shanahan (2004). Camilli, Vargas, and Yurecko (2003) took the studies identified by the NRP and reviewed them to consider which ones actually met the criteria and which did not. Their conclusion was identical to the NRP’s original conclusion on phonics: “Systematic phonics instruction did outperform treatment conditions in which a more typical or moderate level of phonics instruction was provided” (Camilli et al., 2003, p. 34).

Nothing in the NRP report indicates that add-on programs [e.g. phonic workbooks, phonics additives] are less effective than more integrated phonics lessons. Indeed, Shanahan wrote that “the authors of this summary have strayed into the realm of belief and myth rather than that of research or research combined with logic (by claiming that adding a supplementary phonics program to a basal or literature program is necessarily ineffective). NRP did not find that and, given the nature of the research findings we reported on phonics, I would be surprised if the statement were true” (Shanahan, 2003, p.647). Nonetheless, integrated phonics curriculum materials - with decodable texts, of course - have become the standard for schools wishing
to purchase 'scientific' curriculum materials with federal Reading First dollars. (Allington, 2005, p. 462)

The NRP findings have had profound effects upon the selection of textbooks and reading programs in U. S. schools in addition to the teaching of phonics. In many school systems reading programs are approved based on whether they meet RF criteria determined by the findings of the NRP (2000).

Unlike some states, Georgia has not published a list of approved reading programs aligned with the findings of NRP. Georgia did have several school systems that qualified for the RF grant. However, Georgia is home to Dr C. Cupp (retired reading director for the Georgia Department of Education) of Savannah who developed effective reading materials prior to the advent of RF. Several Georgia districts included Cupp's materials in their RF applications and all were rejected (Bracey, 2005). As can be seen in the partial listing of Georgia State-approved K-8 Reading and Language Arts textbooks in Table 2, Voyager Universal Literacy (VUL) was not listed whereas Cupp's Readers were listed. The publishers of VUL made inquiries
at the state level to consider making VUL, one of the Georgia state-adopted reading textbooks.

In March 2005, Cupp filed a complaint alleging that, as a small-time publisher, she was largely prevented from selling Dr. Cupp Readers to RF schools (Gutierrez, 2007).
Table 2

Partial List of Georgia State-approved Reading and Language Arts Textbooks

<table>
<thead>
<tr>
<th>Publisher</th>
<th>Title</th>
<th>Grade Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cupp Publishers, Inc.</td>
<td>Dr. Cupp Readers</td>
<td>K-5</td>
</tr>
<tr>
<td>Cupp Publishers, Inc.</td>
<td>Ten Minute Phonics</td>
<td>K-8</td>
</tr>
<tr>
<td>Division of McGraw-Hill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Companies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Houghton Mifflin Company</td>
<td>Houghton Mifflin Georgia</td>
<td>K-6</td>
</tr>
<tr>
<td>Harcourt School Publishers</td>
<td>Harcourt Trophies</td>
<td>K-6</td>
</tr>
<tr>
<td>Rigby</td>
<td>Rigby Literacy</td>
<td>K-5</td>
</tr>
</tbody>
</table>

Source: Georgia Department of Education (2002)
Scientifically Based Reading Research (SBRR)

One claimed purpose of NCLB is to "close the achievement gap" so that no child is left behind. The Congressional charge to the NRP was to undertake comprehensive, evidence-based analyses of the experimental and quasi-experimental research literature relevant to a set of selected topics judged to be of central importance to teaching young students to read (NICHD, 2000, p. 1).

Definitions of SBRR

The introduction of the term scientifically-based reading research in the Reading Excellence Act and the subsequent work of the NRP started a movement that led to the inclusion of more than 110 references to the term scientifically based research (SBR) in the NCLB Act (PL 107-110, 20 U.S.C § 7801, Subchapter IX, "General Provisions"; The Reauthorization of the Office of Educational Research and Improvement (2002). The definition presented in the Reading Excellence Act as well as Title I, Part B, Section 1208 (6) of the Elementary and Secondary Education Act (ESEA) of 1965 as amended defines SBRR as research that:
(A) applies rigorous, systematic, and objective procedures to obtain valid knowledge relevant to reading development, reading instruction, and reading difficulties; and

(B) includes research that—

(i) employs systematic, empirical methods that draw on observation or experiment;

(ii) involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;

(iii) relies on measurements or observational methods that provide valid data across evaluators and observers and across multiple measurements and observations; and

(iv) has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and
scientific review (Office of Inspector General, 2006).

This definition used in NCLB's RF program is less rigorous and essentially leaves out two of the six prongs of SBR. The two missing prongs relate to research design and the replication of results. The two prongs are:

(v) is evaluated using experimental or quasi-experimental designs in which individuals, entities, programs, or activities are assigned to different conditions and with appropriate controls to evaluate the effects of the condition of interest, with a preference for random-assignment experiments or other designs to the extent that those designs contain within-condition or across-condition controls;

(vi) ensures that experimental studies are presented in sufficient detail and clarity to allow for replication or, at a minimum, offer the opportunity to build systematically on their findings (No Child Left Behind Compliance Manual, 2004).
In the actual NCLB regulation, the last two prongs are numbers four and five. It would behoove educators to follow the leaders in the fields of medicine and welfare policy who demonstrate that rigorous evidence can produce advances. The NRP work was a large-scale effort to replicate in education the research methodology work done in the medical and welfare policy fields.

The debate concerning effective reading instruction combined with SBRR has spanned several decades. An examination of a variety of public data bases by NRP staff revealed that approximately 100,000 research studies on reading had been published since 1966, with perhaps another 15,000 appearing before that time (NICHD, 2000).

The research base [used by NRP] was actually much smaller as reported by Yatvin et al. (2003). Thousands [of studies] were never considered by the NRP because they dealt with topics NRP chose not to investigate or that did not meet the criteria for inclusion. The total number of studies actually examined by all the NRP subgroups combined was 438. Conclusions of the phonics subgroup were based on only 38 studies (Subgroups, pp. 2-91, 2-131). When researchers analyzed the 38 phonics studies the NRP
selected, using a more appropriate and complex statistical approach, any advantage of systematic phonics over embedded phonics vanished (Allington, 2007). The details of what constitutes SBRI and SBRR continue to be discussed and contested, and some assessment instruments, instruction practices and monitoring methods touted as SBRI and SBRR vary in the extent that they are truly scientifically supported.

Scientifically Based Reading Studies

Several studies have been conducted regarding the effect of student progress monitoring and its influence on students' progress in reading achievement (Baker & Good, 1995; Fuchs & Fuchs, 2005; Safer & Fleischman, 2005) on the design of reading instruction. Other studies show curriculum-based measurement to be a reliable indicator for making instructional decisions (Fuchs, Deno & Mirkin, 1984; Good & Jefferson, 1998; Deno, 2003). Differentiated instruction (Powell & Napoletello, 2005); phonics instruction (Lyon & Chhabra, 2004), classroom assessment (Black & Williams, 1998; Stiggins, 2002); Dynamic Indicators of Beginning Early Literacy Skills (DIBELS) assessments (Good, Kaminski, & Simmons 2001); continue to
be reliable indicators and predictors of success or failure in reading. No other assessment is used as widely as DIBELS in Reading First (RF) schools (Manzo, 2007a).

Not all researchers favor the use of DIBELS at the classroom level. DIBELS has drawn recent criticism from those who cite the tendency of some educators to teach to the tests, and to give DIBELS too much weight in gauging reading ability. Another concern is the aggressive promotion of DIBELS by federal employees and consultants to the RF program (Manzo, 2005). DIBELS do not tell the teacher what to teach or how to teach. They simply tell the teacher how well the instruction is working within the context of foundational literacy skills for each student who may require close and accurate monitoring (Langdon, 2004). Some critics charge that DIBELS got a competitive edge not because of its superiority, but because its developers and their colleagues at the University of Oregon, were key consultants to the U. S. Department of Education for RF (Manzo).

Good, a former Oregon University researcher and the developer of DIBELS, was on the assessment committee that evaluated 29 early-literacy tests, including DIBELS,
(Manzo, 2005). DIBELS was "one of many screening tools on the market that could have been used to perform Reading First assessments," the inspector general writes, but "only DIBELS was featured in" the academy [Secretary’s Reading Leadership Academies] materials" (Manzo, 2007a, p. 13).

Research has demonstrated that when teachers use student progress monitoring, students learn more, teacher decision making improves, and students become more aware of their own performance (Safer & Fleischman, 2005). Fuchs and Fuchs (2005) analyzed research on student progress monitoring considering only experimental studies. They concluded that when teachers use systematic progress monitoring to track their students' progress in reading, mathematics, or spelling, they are better able to identify students in need of additional or different forms of instruction, they design stronger instructional programs and their students achieve better.

As noted by Baker and Good (1995), some evidence shows the validity and reliability of student progress monitoring procedures in evaluating progress of English Language Learners. Extensive research conducted by Good, Kaminski, and Simmons (2001) attests to the validity of DIBELS and
curriculum-based measurement. Elliott, Lee, and Tollefson (2001) studied the reliability of DIBELS and curriculum-based measurement for identification of kindergarten students at risk for reading failure and found that curriculum-based measurement (CBM) is one of the few alternative forms of assessment where an impressive body of data supports the technical adequacy and the practical application of these techniques in assessing young children. The researchers claimed that DIBELS measures are practical, easily repeated and adapted to the curriculum and represent many of the best features of alternative assessments according to Elliott et al., (2001).

Conversely, Hintze, Ryan, and Stoner (2003) found that DIBELS "benchmark" or cut-scores may be set too high, from a diagnostic accuracy point of view. When DIBELS are used district-wide to classify children as in need of early intervention services, the potential for costly mistakes (i.e., large numbers of false positives = high costs) suggests that further research on benchmark or cut-scores is necessary:

The research demonstrated that using the DIBELS recommended cut-scores for ISF [initial sound fluency]
and PSF [phoneme segmentation fluency] resulted in high sensitivity to identifying children with low phonological awareness skills as indicated by the CTOPP [Comprehensive Test of Phonological Processing]. (Hintze et al., 2003, p. 554)

Allington (2007) claimed that no studies demonstrated that using DIBELS improves instruction or reading achievement. The DIBELS developers provided several positive studies of its predictive validity, but independent researchers have not been able to replicate those findings (Carlisle, Shilling, Scott, & Zeng, 2004). Pressley, Hilden, and Shankland (2005) concluded that, based on available data, the fairest conclusion is that DIBELS mis-predicts reading performance on other assessments much of the time, and at best is a measure of who reads quickly without regard to whether the reader comprehends what is read. (p. 1)

In the primary grades, such an assessment system in schools at minimum must reliably (a) measure growth on foundational reading skills on a frequent and ongoing basis, (b) predict success or failure on criterion measures of performance (i.e., high-stakes tests), and (c) provide
an instructional goal that if met, will prevent reading
collapse and promote reading success (Good et al., 2001). A
significant body of research (e.g., Deno, 2003; Good &
Jefferson, 1998; Fuchs, Deno, & Mirkin, 1984) conducted
over the past 30 years has shown CBM to be a reliable and
valid predictor of subsequent performance on a variety of
outcome measures, and CBM’s utility for a wide range of
instructional decisions (Good, Jefferson, & Shinn, 1998).
In light of the Congressional charge to educators and
education researchers to identify SBR resources to
ameliorate the reading deficit in the United States, it is
reasonable to assume that the result of such identification
would be the development of an instructional structure that
supported and enhanced reading comprehension.

A Framework for Reading Comprehension Instruction

By the time they reach third grade, average American
students can read most of the words in their spoken
vocabulary (Spear-Swerling & Sternberg, 1996). Yet, as
students progress through school, the achievement gap
widens between “good” and “poor” readers resulting in a
phenomenon referred to as the “Matthew Effect.” The
“Matthew Effect” is when good readers become better readers
and poor readers become more frustrated and fall further behind. Skilled readers are good comprehenders (Snow et al., 1998), and possess core reading skills including the ability to identify individual words quickly and accurately (Adams, 1990; Ehri, 1998; Perfetti, 1985; Rayner & Pollatsek, 1989).

Research to date shows that there is no one best approach to teaching reading, a conclusion Chall (1967) drew long ago. Reading instruction strategies which include skill-based and literacy-based instruction embedded within a balanced approach to literacy to include components of fluency and phonemic awareness support the idea of reading as a product and a process. Reading for some students is an extensive, sequential, bottom-up process producing a product (reading comprehension) that is an essential component for further learning. How reading process and product are achieved is the topic of several theories, including the widely accepted psycholinguistic theory.

Psycholinguistic Theory

The psycholinguistic theory of reading proposes that real reading occurs when the three kinds of information listed below are used in concert.
1. Semantic-associational information includes our knowledge of what words refer to in the real world.

2. Syntactic information refers to the ordering relationships among words in sentences.

3. Grapho-phonemic information is what novice readers learn in the phonics component of their reading program. (Pearson, 1976)

Other researchers (Goodman, 1967; Smith, 1971; Samuels, 1970) have discussed language-based models of the reading process to provide a richer understanding of how children read and learn to read (Samuels, 1978). Goodman, Smith and Samuels provided psycholinguist models that can be referenced to provide a framework for making instructional decisions. According to Ehri (1998a), Goodman had proposed that reading is a psycholinguistic guessing game in which readers learn to use context to predict words in text.

According to Cowen (2003) "Adams's research is based on models of psycholinguistics that develop an understanding of the reading process by examining and explaining the importance of a reading system that is interconnected by four processors" (p. 46). Adams (1990a/1990b) described a reading system that depicts four
key components or processors at work in a synchronized work, which showed how the research supported a balanced approach to reading instruction. These ideas are depicted in Figure 1.
• The Orthographic processor is responsible for perceiving the sequences of letters in the text.

• The Phonological processor is responsible for mapping the letters onto their spoken equivalents.

• The Meaning processor contains our knowledge of word meanings.

• The Context Processor is in charge of constructing an ongoing understanding of text. (Adams, 1990b, p. 21)

Whitehurst and Lonigan (1998) proposed that emergent and conventional literacy derive from individuals' ability to use information from two interdependent domains, outside-in and inside-out as represented in Figure 2. According to Whitehurst and Lonigan (1998) fluent reading involves a number of components skills and processes. A reader must decode units of print into units of sound and units of sounds into units of language. This is an inside-out process. The fluent reader must understand those auditory derivations, which involves placing them in the correct conceptual and contextual framework. This is an
outside-in process. The bidirectional arrows in the figure illustrate that there is a cross-talk between different components of reading. For example, the sentence context affects the phonological rendering of the italicized letters in those two phrases: "a lead balloon," "lead" me there" (p. 13).

FIGURE 2. Domains of Information and Reading
Learning to decode (i.e., to smoothly, effectively, and effortlessly process inside-out information to print) is a critical step in learning to read for meaning (Whitehurst & Lonigan, 2002).

Many reading experts agree that phonological processing ability plays an important role in the reading acquisition of young learners (Blachman, 1994; Goswami & Bryant, 1990; Snow, et al., 1998; Stanovich, 1992). Memory for sounds and the rate at which children can retrieve them from short-term memory also contribute to individual differences in children’s reading performance (Mann & Lieberman, 1984; Torgesen, 1985). The connection between phonological processing and reading performance is unquestionable. To be able to read, children must show sensitivity to sounds, develop a correspondence between these sounds and letters of the alphabet, and then retrieve and blend these sounds into a sequence called reading (Hasselbring & Goin, 2004).

The psycholinguistic model can be contrasted with views of reading that assign primary activity to decoding processes and relegate the semantic and syntactic processes to the role of verifying the phonological message (Pearson,
1976, p. 87). Recent models of reading include neuroscience. Some education policymakers have used the conclusions of neuroscience research to claim that neuroscience proves the necessity of intensive phonics instruction for students who struggle with reading (Willis, 2007).

Neuroscience and reading instruction

Neuroscience is a blend of psychology and biology or the convergence of cognitive psychology with neuroanatomy and neurophysiology. Neuroimaging is an important technique for studying the neurobiological underpinnings of reading, reading disorders, and reading interventions (McCardle & Chhabra, 2004). Brain imaging has demonstrated differences in activation patterns between good and struggling readers of all ages (Shaywitz & Shaywitz, 2007). More than a century ago, studies of adults who suffered a stroke with subsequent loss of the ability to read revealed neural systems that influenced reading. Three neural systems that influence reading were apparent from these studies. The neural systems are:

- an anterior system which serves articulation and word analysis;
- a posterior system which serves word analysis;
- a second posterior system which serves rapid, automatic, fluent identification of words (Shaywitz & Shaywitz, 2007).

Studies from large and well-studied populations of students with reading disabilities confirm that in young children (Fletcher et al. 1994; Stanovich & Siegle, 1994) as well as in adolescents (Shaywitz et al., 1999) a deficit in phonology represents the most robust and specific correlate of reading disabilities (Morris et al. 1998; Ramus et al. 2003).

These findings form the basis for the most successful and evidence-based approaches to reading instruction and interventions for struggling readers (NICHD, 2000). Increased activation in the neural systems for reading were noted in struggling readers who received an intervention focused on evidence-based application of the alphabetic principle when compared to the neural systems of struggling readers who received other types of interventions (Shaywitz & Shaywitz, 2007). A goal of neuroimaging studies is to determine:
1. differences in brain function that accompany reading disability versus reading proficiency, and
2. the effects of well-designed instruction on both reading and brain development (McCandless & Chhabra, 2004, p. 383).

The research and work in neuroscience and neural systems are relatively new but much progress has been made identifying neural systems for reading in good readers and identifying disruption in these systems in struggling readers (Shaywitz & Shaywitz, 2004).

Reading Comprehension Approaches

There has been disagreement within the reading profession with regard to effective strategies for reading acquisition. The most common dispute pits the use of phonics against the use of whole language—meaning-based—approaches (Cowen, 2003). For more than two decades, most intensely from the early 1980s through the 1990s, a “war” has been waged between these two camps over how best to teach students to read. The struggle opposed advocates of skills-based, phonics-based approaches against partisans of naturalistic, “whole-language” strategies (Moats, 2007). “The focus of attention has shifted from the researchers’
theories and data back to the teacher, ... with a heterogeneous group of children, all awaiting their passports to literacy" (Snow, Burns, & Griffin, 1998, p. vi).

When making decisions about reading programs, educators look at various instructional models whose implementations result in different levels of proficiency and sustainability in reading. Critical data relevant to the present study and to the performances of elementary students in Georgia show that:

1. Approximately 50% of the students placed in special education are diagnosed with a learning disability, and of those, 60% have problems in reading (U. S. Department of Education, 1997).

2. Educators in Georgia have made considerable progress in decreasing the gap between the state assessments (Criterion Referenced Competency Tests or CRCT) and national assessments - National Assessment of Educational Progress (NAEP) in reading from 2003 to 2006. The revised English Language Arts Quality Core Curriculum uses the Primary Literacy Standards for Kindergarten Through Third Grade published by the National Center for
Education and the Economy and the University of Pittsburgh in the strands for reading, writing, and conventions for those grade levels. In kindergarten through grade 3, the reading standards focus on the results of the NRP, emphasizing phonemic awareness, phonics, fluency, vocabulary, and comprehension (Georgia Department of Education, 2004).

The state does promote and encourage the use of standards-based education and a balanced approach to reading instruction, though implementation of reading instruction is not consistent throughout the state or within school systems.

In addition to inconsistent reading instruction implementation, one must also consider the changing views of reading over time which drastically impact instructional delivery. Reading instruction has changed according to the varied purposes and definitions that existed and developed. Three distinct time periods show various views of the purpose of reading. An early nineteenth century view of reading (prior to 1826) presented reading as a facilitator and caretaker of memory; a mid-nineteenth century view of reading (1826-1882) proposed it as an expressive art; and a

Dates vary as to when the first reading research was actually conducted. The reported dates ranged from 1879 to 1884. Robinson, Faraone, Hittleman, and Unruh (1990) stated that prior to 1884 there were no research reports (published in English) focused on reading. Samuels and Kamil (1984) reported that research on reading dates as far back as 1879, when a paper was published on eye movements in reading (Walker, 2006). Table 3 summarizes some findings from reading research reports at various times.
Table 3

Summary of Historic Reading Research Reports (1884-2007)

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Number</th>
<th>Type</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1884-1910</td>
<td>35</td>
<td>Lab investigations unrelated to pedagogy</td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td>1- Romanes (1884)</td>
<td>Reading</td>
<td>Material must be read twice</td>
</tr>
<tr>
<td>Pre-1910</td>
<td></td>
<td>Distinctions between oral and silent reading</td>
<td>Silent reading unseated oral reading</td>
</tr>
<tr>
<td>1910-1950</td>
<td></td>
<td>Studies due to poor standardized test scores</td>
<td>Investigated product of reading not process</td>
</tr>
<tr>
<td>1965-present</td>
<td></td>
<td>Multiple perspectives, background knowledge of reader, context of reading act</td>
<td>Development of models and theories based on concepts from other disciplines</td>
</tr>
</tbody>
</table>

Reading evolved into two major views of reading instruction - the skills-based approach (which emphasizes the use of phonics) and the meaning-based or whole-language approach (which emphasizes reading comprehension and enrichment) (Walker, 2006). According to Allington (2002), "I admit it, I am not a strident anti-basalist. But over the years I have become more concerned that the purchase of basal reading series has the potential for producing
unintended negative effects" (p. 6). Why is it that federal and state agencies are now so interested in whether a reading series is based on "scientific research"?

Are teachers really prepared to truly teach reading especially to children "at-risk" of being successful or has scientific research clearly identified the superiority of a particular instructional approach?

*Phonemic Approach and Direct Instruction*

Complaints have been made that the United States Department of Education (DOE) exercised undue influence in the administration of the RF program, by promoting or endorsing specific reading programs, materials, assessment instruments, and models of instruction (Office Of Inspector General, 2007, p. 5). The report concluded that federal officials seemed biased toward a particular instructional approach, direct instruction (Manzo, 2007a). Most RF grants approved included a phonemic approach and direction instruction.

The skills-based approach was heavily influenced by the work of Chall, a Harvard reading researcher and professor. In skills-based learning, phonics skills are
taught separately with the expectation that learned reading comprehension skills will follow. With automatic word recognition, the child does not have to concentrate on the words and can concentrate fully on the meaning of the text (Chall, 1996; Dowhower, 1987; Ehri, 1995; LaBerge & Samuels, 1974; Perfetti, 1985; Samuels, Scherner & Reinking, 1992). Several studies demonstrated that phoneme awareness is one of the best predictors of reading success (Bowey, 1995; Bradley & Bryant, 1983; Bradley & Bryant, 1985; Fox & Routh, 1975; Juel, 1988; Langenberg, 2000; Lundberg, Olofsson, & Wall, 1980; Mann, 1984; Muter, Hulme, Snowling, & Taylor, 1997; Naslund & Schneider, 1996; Share, Jorm, Maclean & Matthews, 1984; Stanovich, Cunningham & Cramer, 1984; Stanovich, Cunningham & Freeman, 1984; Stuart & Coltheart, 1988; Stuart & Masterson, 1992; Tunmer & Nesdale, 1985; Williams, 1984).

In 1967, the federal government began Project Follow Through (FT), one purpose of which was to compare the long-term effects of reading instructional methods. According to Lindsey (2004), a study of FT completed in the 1970s, was the largest education study done at that time, costing over
$600 million, and covering 79,000 children in 180 communities. Researchers examined various programs and education philosophies to learn how to improve the education of disadvantaged children in grades K-3 (Lindsey, 2004). FT was perceived a failure because all of the tested reading instruction models, except one, did not produce the desired results. The only model that brought children close to the 50th percentile in all subject areas was the Direct Instruction (DI) model (Grossen, 1995).

In an influential essay, Rosenshine and Stevens (1986) used DI to describe systematic teaching, or explicit teaching, or active teaching. Interpretation and development in particular applications has been provided in a second line of scholarship associated with the work of Siegfried Engelmann and his colleagues (Schug, Tarver, & Western, 2001). DI is not the same as phonics but it is often associated with phonemic awareness due to its foundation in close analyses of the comprehension and reasoning skills needed for successful performance (Schug et al.).
Other research supports the efficacy of the DI model. O'Neill (1988) reviewed 150 studies and concluded DI was correlated with improved learning among primary children from working and middle-class backgrounds - a conclusion supported by almost every relevant study on the topic.

The Coleman and Vaughn (2000) descriptive study found significant effect sizes for direct instruction and peer-tutoring reading interventions for elementary students with Emotional Behavioral Disorders (EBD). The same study (as cited in Kamps et al., 2003) showed that the DI reading curriculum demonstrated higher growth in reading fluency than did literature-based instruction.

DI emphasized the systematic teaching of phonemic awareness and language skills among other techniques. Follow up studies (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977; Bock, Stebbins, & Proper, 1977; Meyer, 1984) showed that the effects of DI are sustainable. Further studies (as cited in e.g. Gersten & Keating, 1987; Gersten, Keating, & Becker, 1988; Adams, 1995) showed the long-term benefits of DI.
Literacy-based approach

The literacy-based or meaning-based approach to reading was heavily influenced by the work of Goodman, who developed the psycholinguistic perspective, which asserts that readers rely more on the structure and meaning of language rather than graphic information from text. Goodman's reading model is better known as the whole-language approach. The term "whole language" does not refer only to providing interesting comprehensible texts and helping children understand less comprehensible texts. It involves instilling a love of literature, problem-solving and critical thinking, collaboration, authenticity, personalized learning, and much more (Goodman, Bird, & Goodman, 1991; Krashan, 2002). Children focus on the wholeness of words, sentences, paragraphs, and entire books to derive meaning from context (Walker, 2006). This review of literature supports the current thinking that a blended or balanced approach to reading instruction should be considered rather than the use of one reading model exclusively.

A guiding principle of the literature-based perspective is that literacy acquisition occurs in a book-
rich context with an abundance of purposeful communication where meaning is socially constructed (Cullinan, 1987). Several theoretical orientations support the literature-based perspective, but it is more closely associated with reader-response theory (Morrow & Gambrell, 2002). Reader-response theorists posit that literature is not an object to be studied, nor does it have one correct interpretation (Iser, 1978); rather, meaning in the text is constructed by the readers' own interpretation of their experiences while they are reading (Rosenblatt, 1978). Researchers have explored literature-based instruction and provided insights about new ways of conceptualizing literacy development in literature-based classrooms (Allington, Guice, Michelson, Baker & Li, 1996; McGee, 1992).

The first literature-based beginning reading programs were created in the late 1980's for textbook selection in California, which mandated literature programs based on whole language for beginning reading instruction (California English Language Arts Committee, 1987). To support the implementation of literature in reading programs, the state textbook guidelines in California and Texas (California English Language Arts Committee, 1987;
Texas Education Agency, 1990) called for the elimination of contrived texts and the use of text with literary merit. The most compelling evidence for wide-spread implementation of literature-based reading instruction in the United States came from the findings of the 1992 NAEP (Mullis, Campbell, & Farstrup, 1993). This study revealed that teachers reported a "heavy" emphasis on literature-based instruction, and that the students of these teachers displayed higher levels of reading proficiency.

Although some research on literature-based instruction has shown increased reading achievement levels, policymakers and educators question the use of literature-based instruction for beginning readers (Stahl, McKenna, & Pagnucco, 1994). Hiebert (1999), who revisited the question of whether literature should be the sole material for reading instruction, concluded that although authentic literature is vitally important, young children also need texts that are more systematic with skill development to learn how to read; and the skills must be explicitly taught.

Reutzel, Oda, and Moore (1989) reported the positive effects of literature-based programs on the print-awareness
and word-reading acquisition of kindergarten students. Others suggest that decoding and comprehension can be enhanced through literature (Dahl & Freppon, 1995). The literature also revealed that children in the literature-based program were more strategic readers than those in a skills-based program (Dahl & Freppon).

Whole language advocates argue that when whole language is defined correctly, when it includes real reading, students in these classes do better on tests of reading comprehension, [than students in skill-based classes] with no difference on skills tests (Krashen, 1999). A study by Block (1993) on the effects of a literature-based program revealed that the literature-based group outperformed the control group on the reading comprehension, vocabulary, and total battery sections of the Iowa Tests of Basic Skills.

Whole language advocates claim that the rules of phonics are complex and have numerous exceptions. So many are unteachable (Smith, 1994). Skill-building advocates claim that this is not the case. Krashan (2002) noted that Johnson's (2001) re-examination of Clymer's original review of 45 phonics generalizations of words revealed that Clymer's conclusions stand: the rules of phonics are
enormously complex. Johnson's analysis confirmed that extensive phonics teaching is a hopeless endeavor.

Jeynes and Littell (2000) reviewed 14 studies and concluded that overall, low SES children do not benefit from whole language instruction, but "there may be some advantages to the whole language approach in its purist form" (p. 21). Krashan (2002) commented that A close analysis of the actual studies [(Harris & Serwer, 1966a; Harris & Serwer, 1966b; Manning, Manning, & Long, 1989; Morrow, O'Connor, & Smith, 1990] reviewed by Jeynes and Littell (2000) showed that when tests of reading comprehension are considered, when real reading is considered as the core element of whole language, and when details of studies are examined closely, whole language does very well in method comparison studies. Although the authors [Jeynes, Littell] conclude otherwise, these studies actually provide evidence for the limits of phonics instruction and the efficacy of whole language. (p. 6)

Reading Instruction Research

Discussions and review of research for or against one particular method of reading instruction did little to
reduce the student achievement gap in reading. The next few pages review some major studies that have influenced reading instruction to a large degree.

A summary of six U. S. research syntheses (Table 4) on beginning reading instruction highlights the major findings to support current (2007) knowledge about beginning reading instruction. The synthesis of current research reveals that the early research was mainly concerned with predictors of reading ability combined with reliance upon the skills-based or the whole-language based approach to instruction. Some mid-1980s research revealed signs that a balanced approach to reading instruction should address the influence of home and social factors.

Six major U.S. reading research studies (see Table 4) provide converging evidence of a balanced approach to reading instruction. Each of the six studies provided a U.S. national synthesis of current knowledge of reading instructional practices for beginning readers. Each study was supported by nationally recognized research organizations, commissions or academies.
Table 4

Authors and Findings for Six U.S. Reading Research Syntheses

<table>
<thead>
<tr>
<th>Title</th>
<th>Year</th>
<th>Author(s)</th>
<th>Findings</th>
</tr>
</thead>
</table>
| The Cooperative Research program in First-Grade Reading Instruction | 1967 | Guy L. Bond, Robert Dykstra | Convergent research over 30 years verifies that the two most important predictors for beginning students’ success in learning to read are  
1. Knowledge of Letter names (Alphabetic principle) and  
2. Ability to discriminate between word sounds (Phonemic awareness) (Bond & Dykstra, 1997) |
| Learning to Read: The Great Debate                 | 1967 | Jeanne S. Chall             | Research shows overwhelming support for programs that “included systematic phonics” (1967, p.79). Chall’s investigation related to lower socioeconomic children and to children with lower level abilities showed that early code emphasis proved most beneficial to this growing at-risk population. Her view advocated that children reading challenging literature raises vocabulary achievement, and subsequently, reading comprehension achievement (Cowen, 2005). “No program can do all things for all children, and no program can be all things for all teachers” (Chall 1967, p.167). |
| Becoming a Nation of Readers: The Report of the Commission on Reading | 1985 | Richard C. Anderson, Efrida H. Heibert, Judith A. Scott, Ian A.G. Wilkinson | First major report to champion a common ground for phonics and a constructivist approach, which allows phonics and reading for meaning to co-exist (Cowen). Anderson et al. (1985) stress the importance of a balanced reading approach by stating rather boldly that phonics should be taught early, that teachers should keep it simple, and phonics need not be taught beyond the end of second grade for most children. |
| Beginning to Read: Thinking and Learning About Print | 1990 | Marilyn J. Adams            | In contrast to the first three national studies listed, Adams’s (1990) findings seem to more fully recognize the importance of the home and community on beginning reading |
Preventing Reading Difficulties in Young Children

The Committee on the Prevention of Reading Difficulties in Young Children - Catherine Snow, Susan Burns, and Peg Griffin (Eds.)

Prevention of Reading Difficulties (PRD) is the first report to promote an extensive list of guidelines in literacy instruction and diversity issues related to reading instruction and achievement among minority groups of African American and Latino children, English Language learners, and children with specific reading disabilities (Cowen). PRD recommends that the best course of action in dealing with...risk factors is to provide early intervention as soon as possible. PRD's research basically supports the findings of its predecessors,...however it places greater importance on phonological awareness and the need to provide direct instruction in this area.

Report of the national Reading Panel: Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction

| Members of the National Reading Panel with Donald N. Langenberg, chair, NRP subgroup chairs serving as editors: Alphabetic, Linnea Ehri; Comprehension Michael L. Kamil; Fluency, S. J. Samuels; Timothy Shanahan, Sally Shaywitz, |
|---|---|
| 1. Because the structure of English spelling is alphabetic, it is extremely important for beginning readers to discover and develop phonemic units, but to do so requires direct instruction in the manipulation of sounds and letters in individual words. |
| 2. Phonics instruction is most effective when introduced as early as kindergarten. |
| 3. Fluency of reading is highly correlated with improved comprehension and that fluency is taught best through guided interaction with direct feedback provided by a teacher or knowledgeable facilitator. |
| 4. Vocabulary instruction is |
Teacher Education, Gloria Correro, Michael L. Kamil; Technology/Next Steps, Michael L. Kamil

crucial to the comprehension processes of a skilled reader (NICHD, p.4-3). There are 8 specific comprehension strategies that appear effective for classroom instructional use. Appropriate teacher education “does produce higher achievement in students” (p.5-2). Comprehensive literacy software programs exist that are beneficial in teaching vocabulary and phonemic skills.

Source: Cowen (2005); & Tawana D. Miller (2007)

The studies in the syntheses cover more than 30 years and much has changed since Chall’s (1967) Learning to Read: The Great Debate was written (Cowen, 2005). Researchers continue to find that there is no one best approach to reading instruction. Phonics currently is seen as a means to an end that can be taught while using the best children’s literature to help beginning readers learn to read fluently for meaning (Cowen, 2005). Findings for whole language approaches are gleaned from the discussions of reading comprehension posted in all six of the research studies. Conclusions from several of the studies show that by developing vocabulary knowledge, reading fluency, specific comprehension strategies, while using multiple reading activities together with positive teacher feedback,
children improve their overall reading comprehension (Cowen, 2005).

Unlike in some states, Georgia did not promote any specific comprehensive school reform model in 2005. The Northwest Regional Educational Laboratory (NWREL, 2007) listed several comprehensive school reform models specifically aimed at the improvement of reading in students identified as “at risk”. Some of those programs were Success for All, America’s Choice, Accelerated Schools and DI.

A recent evaluation completed by the Comprehensive School Reform Quality Center at the American Institutes for Research (CSRQCAIR) revealed that only 10 of the 18 most widely used school improvement programs for middle and high schools have ‘moderate’ or ‘limited’ evidence to show they work, and none deserves a top rating. (Viadero, 2006, p. 9)

Two models mentioned above, Success for All and America’s Choice School Design were given a “moderate” rating though none had a research track record robust
enough to earn one CSRQCAIR's two highest ratings- "very strong" or "strong" (Viadero, 2006).

Rosenshine (2002) examined the effectiveness of DI and Reading Mastery (RM) as instructional methods to help students read at grade level. Rosenshine noted that in future years, reading researchers should report standardized test scores, means and standard deviations for third and fourth grade students grouped by years in a program. In addition, researchers should report effect sizes to help determine how much of a difference an intervention made. Standardized tests in reading begin with a focus on decoding in first grade and shift to a focus on reading comprehension in third grade and higher. Therefore, success for a program with students from poor families, must involve nothing less than helping students read [with comprehension] at grade level, on standardized tests, at the end of third grade and higher (Rosenshine, 2002).

The International Reading Association (IRA) listed 10 research-based practices in its position statement on
evidence-based reading instruction (IRA, 2002). These practices included:

1. reading for authentic meaning;
2. use of high-quality literature;
3. a comprehensive word study/phonics program;
4. use of multiple texts;
5. a balance of teacher-and-student-led discussions;
6. a whole-class community for building background knowledge;
7. working with students in small groups;
8. allowing time for students to read in class;
9. use of direct instruction in teaching decoding and comprehension strategies;
10. use of a variety of assessment techniques to inform instruction. (IRA, 2002, p. 8)

The IRA position statement posited that no single method or single combination of methods can successfully teach all children to read (IRA, 2002). The two programs being compared in the present study focused on methods and the needs of struggling readers.
Review of reading research and struggling readers

The United States was well-known and recognized as a "Global Superpower". However, in the year 2005, more than 40 years after the passage of the Civil Rights Act of 1964 and President Lyndon Johnson's War on Poverty, minority and poor students consistently lag behind others (White, Asian, high SES students) in America in basic math and reading skills (Wilkins, 2005). Several large national studies have been conducted to address the concern regarding the reading achievement gap.

NAEP has been used as the "Nation's Report Card" since the early 1970's to assess reading progress of the nation's 4th and 8th - graders. Few national studies have assessed the reading skills of children when they enter kindergarten and have documented the development of these skills through 5th grade (NCES et al., 2003). In the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K), sponsored by the NCES researchers seek to collect data that may fill the knowledge gap. The ECLS-K assessed children's reading skills, children's home literacy environment and
the reading instruction from their teachers and schools in grades kindergarten through fifth grade (NCES, 2003).

According to the NCES -ECLS-K study, children from a “literacy-rich” home environment demonstrated higher reading skills than did other children and this relationship was more prevalent in the high versus low SES homes. The work of Bernstein in the 1960’s at University College, London involving elaborated and restricted speech codes substantiates the findings in the ECLS-K study. Differences in reading skills and knowledge usually seen in later grades are noticed when children first enter school and persist for the next one to two years. Thus for students to be proficient readers by the third grade as the NCLB legislation requires, educators must develop environments in their school settings and use instructional reading practices starting at the kindergarten or prekindergarten levels and beyond.

Despite some evidence of success and strong progress by Black, Hispanic, and socioeconomically (SES) disadvantaged students from 1970 to 1988, the education reform efforts of the last decade have not enabled
significant numbers of students to become educationally competitive or to close the gaps in achievement (D'Amico, 2001; Lee, 2002; NCES, 2001; Olson, 1996). Struggling readers need more than effective short-term interventions; they also need effective reading instruction in their regular classroom programs (Hiebert & Taylor, 1994). Many students struggling in reading do not qualify for special or compensatory education support services because of the differing criteria used in various school districts for student entry into these programs or the lack of sufficient funding in some school districts for support programs (Spear-Swerling & Sternberg, 1996; Wang, Reynolds, & Walberg, 1988).

Even if struggling readers do receive reading support through special or compensatory education programs, most of their teacher-directed time still occurs in the regular classroom reading programs (Allington & McGill-Franzen, 1989; Cunningham & Allington, 1994; Haynes & Jenkins, 1986; O'Sullivan, Vsseldyke, Christenson, & Thurlow, 1990). Most remedial and special education support programs have not proven to be effective in accelerating the reading growth of struggling readers (Johnston & Allington, 1991; Walmsley
& Allington, 1995). Support programs that are effective in accelerating the reading growth of struggling readers, such as Reading Recovery (Pinnell, Lyons, Deford, Bryk, & Seltzer, 1994) or small-group interventions (Hiebert, 1994), are not available for all children as they are designed to support limited numbers of students (Baumann & Ivey, 1997) and may not be sufficiently effective to accelerate and maintain the reading growth of struggling readers over time (Center, Wheldall, Freeman, Outhred, & McNaught, 1995; Hiebert, 1994; Shanahan & Barr, 1995).

Findings from a recent study on tutoring intervention model called Reading Rescue (Ehri, Dreyer, Flugman, & Gross, 2007) for language-minority students revealed that the small-group intervention was not effective. Struggling readers who received this intervention [Reading Rescue] performed no better than struggling readers who did not receive it (Ehri et al. 2007).

Fluency and Phonological Awareness Instruction

The VUL program has an extensive phonological awareness component that, according to Smith et al. (2001), is beneficial for all students prior to formal reading instruction. Phonological awareness instruction is
suggested for learners at risk for reading delays. In a study of four basal reading programs, Smith et al. (2001) found that initially, core instruction in a basal program is desirable for all students, and then additional, more intensive phonological awareness instruction may be needed by learners at risk for reading disabilities and delay.

Noted reading educator Gates (2002), stated that we must give children interesting material of suitable difficulty, provide them with simple but sound guidance, and give ample opportunity to read on their own and to learn to read better in the course of reading. Gates articulated the two real goals of reading: a) to teach children to read well and b) to love to read. In several classrooms additional guidance is provided by paraprofessionals. Struggling readers are the children most likely to spend time with the least well-trained personnel - teacher aides (Allington & Cunningham, 2007, p. 14).

When considering these two reading goals as they relate to student performances by gender and SES, the present researcher also contemplated the earlier discussion regarding class size. The well known and conclusive results
of the Tennessee's Student Teacher Achievement Ratio (STAR) experiment have empirically shown that small class size makes for a better learning environment (Word, et al., 1990; The Revitalizing Education Project, 2005). Class size has proximal influence on students, which Eccles and Midgley (1999) referred to a stage-environment perspective and argued that changes are often associated with declines in many adolescents' motivation and behavior.

Declines in motivation and behavior can directly impact student performance. Some concerns about student performance have been raised over behaviors [teacher] that create a self-fulfilling prophecy by undermining the learning and well-being of those students for whom the teachers hold the lowest expectations.

Another important factor that impacts third graders in Georgia is the state law regarding retention in grades. Third graders must pass a "gateway" test to be promoted to the fourth grade. Achilles and Mitchel (2001) noted that unfortunately, by grade 3, many children have already been retained in either kindergarten or grade one. From a school perspective (reported in Chard & Kame‘enui, 2000), children who displayed poor reading skills in first grade had a 90%
chance of continuing to have poor reading skills three
years later. Students showing reading problems at the end
of third grade are not likely to improve significantly by
the end of eighth grade (Felton & Wood, 1992, as cited in
Chard & Kame'enui). As students fall behind, a "one size
fits all" approach regarding curriculum choices, continued
reliance on whole-class instruction, and groups that are
too large promote higher rates of failure than necessary
(Elbaum, Vaughn, Hughes, & Moody, 1999).

Interventions selected for young readers should be
targeted for certain populations who have historically
lagged behind the national achievement levels in reading.
Student achievement and conduct are enhanced when teachers
establish smoothly running and efficient procedures for
monitoring student progress, providing feedback, enforcing
accountability for work completion, and organizing group
activities (Eccles et al., 1998; Pintrich & Schunk, 1996).

Educators are faced with the problem of how to teach
reading in the primary grades effectively while
simultaneously raising reading achievement scores using
available resources. One main reason why students are
referred for special education services is that they are
deficient in reading (Snow et al., 1998). The majority of students identified as learning disabled have deficits in reading (Lyon, 1996). Consider that in Georgia, most students identified for special education services are disproportionately African American and male.

If students could become more proficient readers in the early years, special education referrals and grade retention could be reduced, making students more powerful learners and saving public school systems costs. Assigning students to special education classes is an inappropriate and ineffective response because NCLB requirements mandate a standardized assessment participation percentage, as well as a required progress percentage for subgroups that include special education students.

Class size reduction is one approach to, among other things, improve academic achievement for all students and particularly for low-income and minority students (Achilles, Finn, Gerber, & Pannozzo, 2000). Summers and Wolfe (1977) found that low-achieving students perform worse when in larger classes. Thus the literature suggests that disadvantaged students - minorities, low-SES, and low-achievers - gain the most from smaller classes.
For example, research has shown that attending small classes in the elementary grades (K-3) is accompanied by important long-term advantages, namely taking advanced course work in high school, taking college entrance examinations, and graduating from high school (Krueger & Whitmore, 2001; Finn, Gerber, & Boyd-Zaharias, 2005; Finn, Fox, McClellan, Achilles, & Boyd-Zaharias (2006). This is a SBR response. As stated by Word et al., (1990) The consistency of the finding of the small-class effect across all measures is important. . . . A comparison of results for grades K, 1, 2 and 3 provides a picture of routine consistency. The classes of inner-city students consistently score lower on achievement measures than classes in the other three locations. (Note that most of the free-lunch students and a majority of the minority students were in the inner-city classes). The small-class effect is extremely strong (significant p < .001) in all contrasts. (p. 11) Students benefit from small classes wherever the small classes are located. In reading at each grade level, effect sizes for low SES students exceeded those for
high SES students. At Grade 2 the difference was substantial” (Word et al., 1990, p. 19). Krueger and Whitmore (2001) estimated that “if all students were in a small class in grades K-3 for one to four years... the black-white test-score gap would fall by 38% in grades K-3, and by 15% thereafter (Executive Summary).

Similarly, Webb and Meyer (2004), authors of one Wisconsin Student Achievement Guarantee in Education (SAGE) Report, noted in an introductory summary that students in small classes showed statistically higher achievement, by about one-fifth of a standard deviation, than did students in larger classes. The data do raise the possibility that the SAGE program has a more positive effect on African American students than on White students.

In summary, the preponderance of studies on small class size indicate positive effects particularly at lower grades and for minority students. However, in most research in this area, except in the STAR experiment it has been difficult to tease out small class size effects from other variables.

Learning to read and reading proficiency are critical areas of concern. Millions of students may be
improperly identified as "learning disabled" when they simply have not been properly taught to read at an early age. Blacks, particularly males, are disproportionately identified as having attention deficit hyperactivity disorder (ADHD). Standards-based reforms are prompting some school leaders to assign more students to special education classes so as to exclude them from participating in proficiency tests (Darling-Hammond, 1997; Finn, Rotherman, & Hokanson, 2001).

If educators are working extremely hard to hit the standards of accountability, what conclusions can be drawn regarding those students who are struggling to reach standards? What aside from human kindness will compel educators to continue until they reach the remaining 30 percent? The standards movement, in sum, has formally institutionalized the very cycle of failure that it intended to break (Furr, 2001).

In some states standards-based education has been used for several years. In Georgia, the standards-based movement was in the initial stages of implementation in 2004-2005. Recent developments, such
as class size reduction (CSR) and the implementation of performance-based instruction, suggest the growing importance of teacher knowledge of learning outcomes according to the NRP Report (2000) as identified by the NICHD (2000). Among 4th graders in Georgia, for example, 80% of students achieved proficiency on the state reading test, while only 27% scored proficient on the NAEP, 2005 (Institute for America's Future, Rocha, & Sharkey, 2005). This large discrepancy could be explained by some published comments regarding "flaws" in the NAEP achievement levels. For example:

The most recent congressional mandated evaluation conducted by the National Academy of Science (NAS) relied on prior studies of achievement levels. . .

The panel (NAS) concluded NAEP's current achievement-level-setting-procedures remain fundamentally flawed. The judgment tasks are difficult and confusing; raters' judgments of different item types are internally inconsistent; appropriate validity evidence for cut scores is lacking, and the process has produced unreasonable
results. (USDOE, p. 15, 2003; and as reported by Bracey, 2003; in Tienken & Achilles, 2005)

Overall, the national trend in reading shows improvement across most metrics at age 9 between 1999 and 2004 as well as between 1971 and 2004 (NCES, 2005).

Purpose of the two Reading Programs as Related to Research

Several studies have been conducted by Voyager nationally to highlight the results of VUL on reading comprehension. VUL claims that it has 13 scientific research studies and 25 independent impact studies in the core reading area.

In an independent study published in 2006, researchers evaluated the impact of the VUL program on student literacy learning with n=447 economically disadvantaged kindergarten students in 8 months from 2002 and 2003 in eight schools from Cleveland, Ohio and Washington, DC (Frechtling, Zhang & Silverstein, 2006). Overall, the students using VUL showed significantly greater gains than non-VUL students in letter identification, phonological awareness, and other emerging reading skills, with effect sizes ranging from 0.23 to 1.32 ((Frechtling et al., p. 93).
In an independent study conducted in 2002, researchers evaluated the effectiveness of the VUL program with n=58 economically disadvantaged kindergarten classes over a three-month period in Orange County, Florida (Hecht & Torgesen, 2002). The researchers found that VUL students performed above average when compared to the mean performance of U. S. children on nationally standardized measures of phonemic awareness. At post-test, VUL students significantly outperformed the control students in phonemic decoding ability, letter sound knowledge, print concepts, phonemic segmenting, and phonemic blending skills (Florida Center for Reading Research, 2002).

A second evaluation directed by Hecht (2003) documented that VUL students generally outperformed control children on the timed DIBELS measures administered by Orange County Public Schools staff. Both gain scores and standardized effect sizes revealed that children in both the current (2002-03) and the 2001-2002 cohorts made substantial improvements in reading related skills (Hecht, 2003, p. 15). The reading related skills referred to word identification, word analysis, spelling, letter name knowledge, letter sound knowledge, print concepts,
phonological awareness - elision, phonological awareness - segmenting, phonological awareness - blending, expressive vocabulary and DIBELS Nonsense Word Fluency.

This researcher was unable to locate independent current studies or evaluations comparing test results or effectiveness of VUL and HT other than Fulton County School System Evaluations. The survey information and results are noted in this study.

Voyager Universal Literacy - Supplemental or Core

The VUL producers suggest using DI, scripted lessons and a skills-based approach as the basis of its intensive aggregated reading program designed for struggling readers.

In the matter of reading achievement, the important question to raise is not whether achievement levels . . . seem to be impressive or unimpressive. . . .; it is whether improvements in instruction could enable the state [Wisconsin] to do better, and at a lower cost... Children who struggle with reading make up a large share of all those who get slotted into special programs in the K-12 schools. And the special programs are expensive. If more children learned to read well
in the early grades, the need for special, remedial programs would diminish, freeing up resources for other uses. (Schug, Tarver, & Western, 2001, p. 6)

Harcourt Trophies - Comprehensive Core Reading Program

The HT is a comprehensive basal reading program that emphasizes phonemic awareness, phonics instruction in the early grades but features a literacy-based approach to reading instruction. In literature-based teaching, the teacher takes the children’s interests as the crucial starting point and mobilizes those interests via imaginative, age-appropriate activities that will be more effective than deliberate, teacher-centered instruction in helping children learn to read (Chall, 2000). Many state-adopted textbooks and reading programs in Georgia are literacy-based.

Harcourt Trophies (HT) producers claim a wealth of research studies but none are posted on the publisher’s website. The researcher contacted the publisher to acquire copies of the research. Harcourt School Publishers contracted with Educational Research Institute of America (ERIA) to conduct studies of the effectiveness of HT. Results of the studies were compared to national norm
groups. The studies showed that HT third grade classes made more than twice the growth of a national comparison control group on the Reading Comprehension Subtest and on the Reading Score as measured by the Stanford Achievement Tests (SAT), 5th edition 2001-2002.

Program evaluations and analyses conducted in FCSS showed no statistically significant differences in the CRCT scores of VUL students and a comparison group of students, except for one cohort of VUL students (3rd graders in 2005-2006) who had participated in VUL for four consecutive years) who showed slightly higher CRCT scores in Reading in 2005-2006 than did the comparison group. Second graders in 2004-2005 showed slightly higher CRCT scores than did the comparison group in the FCSS program evaluation.

Goals of Reading Research

There are certain identifiable goals for current reading research (Pearson, 2002). These goals should be addressed when making decisions regarding data analysis. Some goals of current reading research are:

1. Research to generate a theory of reading.
Attempts to explain the underlying processes in reading
often fall in this category (Montare, Elmon, & Cohen, 1977; or Spilich, Vesonder, Chissi, & Voss, 1979).

2. Research to generate or validate a model of reading. Many basic research studies are often directed at identification of components of the reading process and their relation to each other. The distinction is not clear-cut when trying to distinguish between the terms theory and model. While they have been used interchangeably by many, it has been argued that they should not be (see Kamil, 1978, or Pearson & Kamil, 1977-1978, for a discussion of this issue).

3. Research to collect data. (Often these studies are conducted to obtain data for use in future research studies. At other times, these studies are designed to collect data for use in making instructional decisions, such as how to sequence or construct materials (Canney & Schreiner, 1976-1977; Finn, 1977-1978).

4. Research to make instructional decisions. Program evaluations, materials development, and classroom procedures are all included in this category (Kurth, 1978; Sheehan & Marcus, 1977; Tharp, 1982). The present study’s literature review addressed these goals.
Why is reading research needed to generate or validate a reading model?

1. Educators are faced with teaching reading in the primary grades effectively while simultaneously raising reading achievement scores using available resources.

2. America as a nation has not made significant strides in closing achievement gaps [gender, ethnicity, socioeconomic status] in reading.

Why is Reading research necessary to make instructional decisions?

1. For students to be proficient readers by the third grade as the No Child Left Behind (NCLB) legislation requires, educators must develop learning environments in their school settings and use instructional reading practices starting at the kindergarten or prekindergarten levels and beyond.

2. If students could become more proficient readers in the early years as shown in the Tennessee STAR study Achilles & Finn, 2006; Word et al. (1990), special education referrals and grade retention could be reduced,
helping students become powerful learners and saving public school systems costs.

This researcher posits that the VUL treatment provides more specificity than the literature-based HT comparison for students that should positively influence the reading performances of at-risk students. In this study the researcher investigated the VUL intervention to determine the influence of the direct-instruction DIBELS-based reading program on reading performance as it relates to gender, SES and SWD. The framework for the present study is presented in Chapter III with the methodology.

Chapter II has provided the review of relevant research theory and literature. The review included NRP findings which have profound effects upon textbook selection, reading and phonics instruction in the United States. Nothing in the NRP report indicated that supplemental phonics programs were less effective than integrated phonics lessons. The six RF reports from OIG indicated that educators across the U.S. were strongly encouraged to use DI combined with phonemic reading programs.
The review of relevant research reveals that struggling readers can benefit from small classes, DI, application of the alphabetic principle and a balanced approach to reading to include multiple reading activities in a literature-rich environment. The syntheses of research suggests that combined reading instruction strategies which include skill-based and literacy-based instruction are necessary to help beginning readers improve reading comprehension.

Chapter III provides the research design and methodology used to determine which of two approaches to reading instruction was more effective in increasing reading comprehension performance.
Chapter III

Research Design and Methodology

The purpose for this investigation was to determine which of two approaches to reading instruction, Voyager Universal Literacy (VUL) or Harcourt Trophies (HT) was more effective in increasing reading comprehension performances for third grade students generally and when examined by socioeconomic status (SES), gender, and students with disabilities (SWD) in our district. The key structural differences between VUL and HT was the use of the Direct Instruction model (DI) and Vital Indicators of Progress (VIP) assessments in VUL and the balanced approach to reading instruction in HT. The two programs were delivered in similar environments with similar blocks of instructional time.

A program evaluation is designed to determine if the "intervention is effective in attaining the desired goals and benefits" (Rossi, Lipsey, & Freeman, 2004). If one of the desired goals is sustained gains in reading achievement accompanied by a reduction in the achievement gap as measured by increased graduation rates for low SES students, then research should substantiate the lasting
effects of the instructional programs such as VUL or HT. The hypothesis supporting long-term benefits was based on theory about early childhood interventions that have continuing impact (Ramey & Ramey, 1998). This chapter includes a description of the research design, subjects, instruments, procedure, methodology, rationale for the study and data selected for the analyses. The research questions addressed in this study are as follows:

1. What difference does VUL or HT reading instruction have on reading achievement for third-grade students as measured by state and national assessments?

2. What difference does VUL or HT reading instruction have on reading achievement for third-grade students in Fulton County, Georgia using (a) socioeconomic status (SES), (b) gender, or (c) students with disabilities (SWD) as the main effects measured by state assessments?

The design of the study is a Type 8 (Johnson, 2001) cross-sectional, explanatory research study (see Chapter I). The study was an example of Type 8 (Johnson) because
the research objectives are explanatory and the time-dimension is cross-sectional (supported by several prior years of data) thereby producing a cross-sectional, explanatory research study. This research is non-experimental quantitative research as defined in the glossary. The design includes a blend of a simple ex post facto design and a randomized two-factor design. In the two-factor design, the study has two independent variables (VUL and HT) which were studied to determine whether those variables interact in some way as they influence the dependent variable – reading comprehension.

Program evaluation and post-hoc analysis methods were used in this study. Program evaluation is applied research that can assist with real world implications (Babbie, 1999). As Kerlinger (1986) stated:

It can be said that nonexperimental research is more important than experimental research. That is, of course, not a methodological observation. It means, rather, that most social scientific and educational research problems do not lend themselves to experimentation, although many of them do lend
themselves to controlled inquiry of the nonexperimental kind. (pp. 359-360)

This study was organized around hypothesis testing. The researcher developed a theory to explain how VUL and/or HT operate with certain groups of students.

Participants

Ten samples representing 10 schools having heterogeneous groups of third-grade students were selected to address the research questions and the null hypotheses. The sample for the study consisted of 1066 third-grade students enrolled in five VUL elementary schools \( [n=465] \) and five HT schools \( [n=601] \) in a large suburban school system in the southeast for the 2005-2006 school year. One of the five VUL schools used the treatment reading program in two third-grade classrooms and the control reading program in two additional third-grade classrooms. Five out of a total of seven VUL schools in the district included VUL instruction at the third grade. The five VUL schools were matched with five HT schools having students with similar SES and mobility.
The students were typically 9-11 years old. Access to analysis and progression of test scores could include ages 5-9 years. All classes were heterogeneous (grade, gender, SES, SWD) sections of students that (a) did not meet, (b) meets or (c) exceeds state standards on the Georgia CRCT.

The school system in the study was not eligible for RF grant funds because in 2005-06, all of the Title I schools in the district made AYP and no elementary schools were in "needs improvement" status. The free/reduced lunch percentage for the entire system was 37%. VUL was piloted in a small number of Title I schools having high percentages of poverty and high numbers of students not performing on grade level as measured by state and national assessments.

Because the VUL or HT reading programs were used exclusively in certain schools, teachers did not have any choice regarding which reading program was taught. Each third grade teacher was previously trained in the use of the specific reading program. The VUL teachers were trained and monitored by the school Curriculum Support Teacher (CST) as well as by the VUL District Coach. The HT teachers were trained and monitored by the school CSTs. The district
Curriculum Support Analyst (CSA) provided system training for school CSTs and the VUL District Coach provided training for the CSTs and principals in VUL schools. The VUL District Coach was a certified VUL trainer who had received specific training from Voyager Expanded Learning.

Instrumentation

The following instruments were used to assess third-grade student outcomes: 1) Georgia CRCT for Spring 2005 and 2006; and 2) ITBS, Fall 2006. Any available Spring 2005 CRCT scale reading scores were used as pretests to determine whether the comparison group was equivalent to the intervention group as baseline.

These instruments assessed third grade student reading achievement. These assessment outcomes were examined for comparison of effect sizes (ES).

Student cognitive outcomes of academic performances on the assessment instruments were measured by norm-referenced tests (NRT-ITBS) and criterion-referenced tests (CRT - Georgia CRCT), resulting in a norm-referenced score or a criterion-referenced score, respectively. Norm-referenced scores compare test-takers to a standard or norm, and may also be described as standards-based assessment as they are
aligned with the standards-based education reform movement (American Federation of Teachers, 2006). NRTs are associated with traditional education, which gauge success by rank ordering students, while standards-based assessments are built on the egalitarian belief that all students can succeed if they are assessed against high standards required of all students.

**Georgia Criterion Referenced Competency Tests (CRCT)**

The CRCT are state-mandated achievement tests for students in grades 1 through 8 that cover the subject areas of reading, English/language arts, and mathematics. Students in grades 3 through 8 are also required to take the CRCT in science and social studies. Scores on all reports are expressed as scale scores, which can range from 150 to 450 or 650 to 950 (depending on whether it is for the old Quality Core Curriculum (QCC) or the new Georgia Performance Standards (GPS)) for each grade and content area (GaDOE, 2006).

Scoring of CRCT reading: (a) 650-799 = Level 1; (b) 800-849 = Level 2; and (c) 850-950 = Level 3. Acceptable levels for grades 1 through 8: (a) Level 2 = meets state
grade level expectations and (b) Level 3 — exceeds state grade level expectations.

Iowa Tests of Basic Skills (ITBS) (see Glossary for further explanation)

The ITBS are sets of norm-referenced standardized tests given annually to some school students in the United States. Georgia law (O.C.G.A., Section 20-2-281) mandates that a national NRT be administered, annually, to students in grades three, five, and eight to obtain information about how the performance of Georgia’s students compare with that of students in a national sample, an external reference group (GaDOE, 2005).

ITBS score reports to the school system included the following:

1. an alphabetical list report by classroom teacher, which list students in a class alphabetically and provides normative scores for each subtest and the total battery;

2. score reports to the local systems and to GDOE which included the following:

   a. school, system, and state demographic summaries, which provide normative data for all students and
disaggregated by groups specified in state and federal law for score reporting, and standard and non-standard administrations;

b. software for test data management for each system and school, with accompanying data on CD-ROM(s) or diskettes in an ASCII format (GaDOE, 2007).

Scoring of ITBS reading ranges from the 1st to the 99th percentile, with the average score set at the 50th percentile. Scores also can be reported as “grade equivalents,” “stanines,” and “normal curve equivalents” (National Center for Fair and Open Testing).

The CRCT and ITBS are part of the assessments required by the Georgia Governor’s Office of Student Achievement (OSA) in response to directives of NCLB. Adequate Yearly Progress (AYP) is determined by the success of subgroups of students in the system and schools on the state tests.

Procedures

Subjects for both VUL and HT were samples from the population of third graders using the data provided in Table 5 in Chapter IV. Hypothesis testing was based on
psycholinguistic theories of reading (Goodman, 1967; Smith, 1971; Samuels, 1970, Pearson, 1976) theory about DI (Adams & Engelmann, 1996; Engelmann and Carnine, 1991); and theory about early childhood interventions that have continuing impact (Ramey & Ramey, 1998). Outcomes most likely to endure have “(1) developmental timing, they start early and continue; (2) program intensity, that is, many hours per week and (3) direct provision of learning experiences, rather than relying on intermediary sources” (Achilles, 2007 p. 5). A work process flowchart was developed to graphically represent the researcher’s methodology for the program evaluation. The process is presented in Figure 3.
FIGURE 3
Methodology

The researcher compared VUL and HT mean reading performances to determine if a statistically significant (p<.05) difference, if any, existed between student test outcomes for programs. The CRCT and ITBS reading results obtained from third grade students using VUL were compared to the CRCT and ITBS reading results obtained from the other students using HT. Results from this comparative and evaluative study should assist district administrators in allocating resources to programs that help students with disabilities (SWD), students by gender and low socioeconomic status (SES). Because the populations for the schools were more than 95% African-American, gender and ethnicity of interest could transfer to African-American males. Policies for the school system indicated that because the study used historical data aggregated for schools and avoided identification of students board approval was not required to conduct the study (see letters Appendixes A and B).

The researcher used cognitive data (CRT and NRT data) normally collected centrally at the district level and that have been publicly released, minus individual student IDs. Proprietary equipment was used. VUL program and HT have
been purchased by the school system and are in regular use. No student names or identifiers were used during the study and all data will remain confidential. Data were obtained from data warehouse records of third grade students to obtain all test scores, gender, SWD, and SES.

Primary analyses were of test outcomes by reading program, gender, special education eligibility, SES, state and national test data (used as benchmarks) for the school year 2005-2006. The spring CRCT 2005 results were used to establish a baseline or pre-test for the analyses.

The general linear modeling used two-way ANOVA, ANCOVA and t-tests as appropriate. Independent t-tests were used to treat the main effect of the 2 factor (VUL/HT) independent variable of reading programs with reading proficiency on the CRCT as the dependent variable. Independent t-tests were used to treat the main effect of the 2 factor (VUL/HT) independent variable of reading program with reading comprehension percentages from the ITBS.

ANOVA was used to test for the effect of each 2 factor independent variable (gender - female/male; SES - free/reduced YES/NO; special education eligibility YES/NO) on the dependent variable of reading comprehension scores.
ANCOVA was used to test the homogeneity of slopes. Interactions, measures of effect size and variations of the independent variables were calculated using the SPSS software v. 13.0. Data were analyzed to determine if a statistical significance ($p \leq 0.05$) existed in reading comprehension among the independent variables.

The researcher had full access to all student information due to the scope of the job duties and responsibilities as a school improvement central office administrator. Data are on the school system database. Permission to conduct this research was granted by the Superintendent of FCSS and the Director of Assessment and Evaluation as state earlier in Chapter I (letters are in Appendixes A and B). The study did not require review by the Institutional Review Board (IRB) due to the lack of involvement of human participants (Appendix C).

Voyager Universal Literacy and Harcourt Trophies Strategies

Third-grade students in VUL participated in two-hour reading periods that included a 45-minute large group lesson, a 60-minute lesson for reading stations (three stations - small groups), and then a 15-minute writing, vocabulary, or spelling connection lesson. Students in HT participated in two-hour reading periods that included
flexible grouping to include a 30-minute large group lesson (question of the day, vocabulary, comprehension), followed by 90-minute lessons for small groups, reading centers, writing, grammar and spelling.

Rationale for Study and Data Selected

One important question to be answered when selecting the data to be analyzed is: "Should school system leaders continue to invest in VUL or should the system continue to use HT exclusively in all of its elementary schools?"

Additionally the researcher would address the rationales presented below.

1. The investigator seeks to add to the base of knowledge relating to the efficacy of two reading programs on improving reading comprehension performance of third graders.

2. Study results would provide information to help school administrators choose programs shown to be effective in developing proficient readers and increasing reading proficiency for students in the No Child Left Behind (NCLB) subgroups of low SES and SWD.
Validity and Reliability of the Instruments

The Encarta Dictionary defines validity as "effective" or bringing about the results or ends intended. No test measures performance with perfect consistency (reliability), therefore it is important to note the standard error of measurement (SEM) when interpreting test scores. For the Georgia CRCT, the SEM is computed for each scale score and an error band (plus/minus one SEM unit) is constructed for each content area and grade. The scale scores at the test level are comparable across forms and test administrations for the same content area and grade level. Starting with the spring of 2006 - students taking the CRCT will have a reported Lexile score linked to the reading score. The Lexile scores use the Rasch method for calculation of validity. Georgia educators employed a modified Angoff methodology in making their recommendations for standards setting and the cut score that was used for 2005-2006 CRCT. Appendix D includes the Angoff methodology and cut scores for the CRCT.

Summary

Chapter III provided detailed description of the design and methodology used by the researcher to gather information to analyze the differences of the two reading
programs' (VUL and HT). A quantitative design collecting 2005-2006 historical data of students in each reading program at ten schools was used.

This study was conducted to determine if the VUL reading program produced significantly different reading comprehension achievement in relationship to gender, SWD SES in third-grade students as compared to the HT program. Because the VUL program uses a DIBELS-like assessment and DI model, application to similar students using DIBELS and DI would likely have similar results to this study.

Chapter IV includes the data and statistical analyses of data collected to evaluate the reading achievement gains of selected third grade students using VUL or HT reading programs.
CHAPTER IV
Data and Results of Data Analysis

This chapter presents the data and statistical analyses of data collected to compare the reading performance of third-grade students who either were instructed with Voyager Universal Literacy (VUL) or Harcourt Trophies (HT). VUL used Direct Instruction (DI) and a three-reading-station rotation where the teacher followed a detailed lesson plan for small-group instruction in phonemic awareness, phonics, fluency, vocabulary, and comprehension. The VUL teacher in station 3 provided for small-group reading instruction of between 3-7 (struggling or emerging) readers up to a maximum of 10 (on-track readers) for up to 20 minutes of daily DI. The HT process is that teachers use flexible student work groups/stations but the parameters and content of each group are left to the discretion of the teacher. HT lesson delivery and verbiage are less scripted than the VUL lesson delivery and verbiage.

Georgia Criterion Referenced Competency Tests (CRCT) scale reading scores from spring 2005 were used as pretests to determine whether the comparison group was equivalent to the intervention group at baseline.
The researcher sought to determine if data indicated that VUL-taught third-grade students achieved a significant achievement gain as measured by a comparison with HT-taught third graders. This information would be shared with the school system administrators. Study results will provide information to help school administrators choose programs that are shown to be effective in developing proficient readers and increasing reading proficiency for students in the NCLB (2002) subgroups of the SES, gender and SWD (See Appendix F, p 246). The study also includes the theory, research and literature review on early reading instruction.

The first section describes the setting for this study. The second section presents descriptive and context data comparing the differences if any, of the VUL and HT interventions on the students by SES, SWD and gender.

Setting for the Study

This study was conducted in a suburban school system outside of Atlanta, Georgia. At the time of the study, the district was the fourth largest school district in the state of Georgia and the 45th largest in the nation with a population of 80,660 students. In five South Fulton schools, 465 third grade students participated in VUL program for the 2005-2006 school year; 601 students in five
South Fulton schools participated in HT reading program for the 2005-2006 school year. These constituted the groups compared; HT was the comparison and VUL was the treatment group.

The data in Table 5 show available demographic comparisons between all third-grade students in the school system and the students in both study groups on the 2003-04 and 2004-05 CRCT. Student demographics were fairly consistent, and thus similar between the comparison (HT - 2005-2006) and the treatment (VUL 2005-2006) groups in the study. Students eligible for free/reduced meals (SES) ranged from 82 to 88% of the N for the study groups in 2005-2006. During the study, gender, students with disabilities (SWD) and race were stable: the percentage of females in each group was 49%; the percentage of males was 51%; the percentage of SWD ranged from 11% to 13%; the percentage of Black students ranged from 84% to 93%; the percentage of White students was 2%. The percentage of Limited English Proficient (LEP) students ranged from 6% in the comparison group to 0% in the treatment group.
Table 5

3rd Grade Students - Frequency and Percentage of Students in System and two study groups

<table>
<thead>
<tr>
<th>Characteristics of Students</th>
<th>2005-2006 System</th>
<th>Harcourt HT</th>
<th>Voyager VUL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd grade population</td>
<td>6038</td>
<td>601</td>
<td>465</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>%</th>
<th>n</th>
<th>%</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited English Proficient</td>
<td>379</td>
<td>6</td>
<td>37</td>
<td>6</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Eligible for Free/Reduced Meals</td>
<td>2420</td>
<td>40</td>
<td>527</td>
<td>88</td>
<td>380</td>
<td>82</td>
</tr>
<tr>
<td>Students with Disabilities (SWD)</td>
<td>896</td>
<td>15</td>
<td>67</td>
<td>11</td>
<td>59</td>
<td>13</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2935</td>
<td>49</td>
<td>294</td>
<td>49</td>
<td>230</td>
<td>49</td>
</tr>
<tr>
<td>Male</td>
<td>3103</td>
<td>51</td>
<td>307</td>
<td>51</td>
<td>235</td>
<td>51</td>
</tr>
<tr>
<td><strong>Racial/Ethnic Origin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>488</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Black</td>
<td>2400</td>
<td>40</td>
<td>503</td>
<td>84</td>
<td>432</td>
<td>93</td>
</tr>
<tr>
<td>Hispanic</td>
<td>590</td>
<td>10</td>
<td>66</td>
<td>11</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>White</td>
<td>2339</td>
<td>39</td>
<td>15</td>
<td>2</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Multiracial</td>
<td>218</td>
<td>4</td>
<td>14</td>
<td>2</td>
<td>13</td>
<td>3</td>
</tr>
</tbody>
</table>

Unless noted, values for HT and VUL together are based on 1066 participants.
The samples consisted of third-grade students who were enrolled in schools with high free and/or reduced lunch (F/R) populations and high rates of student mobility (Mobil). Table 6 indicates percentages of F/R lunches, student mobility and SWD. The VUL schools were matched with HT schools having similar socioeconomic status (SES) and mobility. The values for free/reduced (F/R) lunch and mobility are percentages of total school enrollment at the time of the study. The values for SWD are the percentages of third-grade samples at the specific schools.
Table 6

Student Percentages of Free/Reduced (F/R), Mobility (Mobil), and Students with Disabilities (SWD) being compared in Georgia study

<table>
<thead>
<tr>
<th>VUL</th>
<th>F/R</th>
<th>Mobil</th>
<th>SWD</th>
<th>HT</th>
<th>F/R</th>
<th>Mobil</th>
<th>SWD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>1</td>
<td>92.2</td>
<td>68.5</td>
<td>15.0</td>
<td>6</td>
<td>86.7</td>
<td>49.4</td>
<td>8.0</td>
</tr>
<tr>
<td>2</td>
<td>95.4</td>
<td>67.9</td>
<td>16.0</td>
<td>7</td>
<td>91.1</td>
<td>50.3</td>
<td>17.0</td>
</tr>
<tr>
<td>3</td>
<td>77.9</td>
<td>68.7</td>
<td>12.0</td>
<td>8</td>
<td>82.5</td>
<td>55.6</td>
<td>5.0</td>
</tr>
<tr>
<td>4</td>
<td>83.1</td>
<td>9.9</td>
<td>14.0</td>
<td>9</td>
<td>83.5</td>
<td>52.6</td>
<td>11.0</td>
</tr>
<tr>
<td>5</td>
<td>78.9</td>
<td>49.9</td>
<td>9.0</td>
<td>10</td>
<td>85.3</td>
<td>64.3</td>
<td>10.0</td>
</tr>
<tr>
<td>AVG</td>
<td>85.5</td>
<td>60.9</td>
<td>13.2</td>
<td>AVG</td>
<td>85.8</td>
<td>54.4</td>
<td>10.2</td>
</tr>
</tbody>
</table>

The analysis compared results of the VUL and HT cohort reading comprehension performances using the Georgia Criterion Referenced Competency Tests (CRCT) and the Iowa Tests of Basic Skills (ITBS). The CRCT was the required state assessment. The ITBS was the national NRT normally administered in September of each school year. Fewer students were tested on the ITBS in September 2005 when compared to the number tested on the CRCT in April, 2006.
Data Treatment Presented and Analyzed

Table 7 presents mean pre-test and post-test reading CRCT data collected from the cohorts (see Glossary) Harcourt Trophies (HT) students (n=442,601) and Voyager Universal Literacy (VUL) students (n=304,465). Pre-test data for both groups were collected in April 2005 and the post-test data were collected in April 2006. The HT and VUL students were compared using a mean CRCT scale score to determine reading comprehension levels.

The reading portion of the CRCT assesses reading skills, vocabulary acquisition, literacy comprehension and reading for information. Vocabulary acquisition refers to skills required to read, interpret and apply difficult text and new vocabulary in a variety of texts (Georgia Department of Education, 2007). Reading comprehension and vocabulary; knowledge are strongly correlated (Baumann & Kame'enui, 1991; Daneman, 1988; Langenberg, 2000; Paul & O'Rourke, 1988; Stanovich, 1986). Vocabulary size is a good predictor of reading comprehension (Langenberg, 2000; Rosenshine, 1980). Table 7 shows April 2005 scale CRCT pre-test scores, April 2006 scale CRCT post-test scores, mean differences for HT and VUL students, t-values (t-test and significance computations are in Appendix E).
Table 7

Comparison of Pre- and Post-test CRCT 2006 Assessment Means and Differences for HT and VUL

<table>
<thead>
<tr>
<th>Test</th>
<th>Rdg.</th>
<th>prog</th>
<th>n</th>
<th>X</th>
<th>SD</th>
<th>Diff.</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>HT</td>
<td>442</td>
<td>335.77</td>
<td>42.7</td>
<td>-7.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRCT 2005</td>
<td>VUL</td>
<td>304</td>
<td>343.26</td>
<td>44.8</td>
<td>-7.49</td>
<td>-2.31</td>
<td></td>
<td>p≤.021</td>
</tr>
<tr>
<td>Post Test</td>
<td>HT</td>
<td>601</td>
<td>818.11</td>
<td>26.6</td>
<td>-5.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRCT 2006</td>
<td>VUL</td>
<td>465</td>
<td>823.67</td>
<td>27.9</td>
<td>-5.56</td>
<td>-3.32</td>
<td></td>
<td>p≤.001</td>
</tr>
<tr>
<td>Mean Diff</td>
<td>HT</td>
<td>482.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of Gains</td>
<td>VUL</td>
<td>480.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

With a mean CRCT pre-test of 335.77 for the HT cohort and mean CRCT pre-test of 343.26 for the VUL cohort a 2-tailed t-test of difference of means ($t = -2.308, df = 744$, $p = .021$) indicated that the difference was significant favoring VUL cohort. Results show that VUL students had a mean-pretest increase of 7.49 greater than did the HT cohort. With a mean CRCT post-test reading scale score of
818.11 for the HT cohort and a mean CRCT post-test reading scale score of 823.67 for the VUL cohort a 2-tailed t-test of difference of means ($t= -3.316, df = 1064, p=.001$) indicated there was a significant difference at $p \leq .05$. The VUL cohort had a significant mean level difference of 5.56 greater than did the HT cohort. This statistical result combined with a visual inspection of pre- and post-means indicates that the VUL cohort started ahead of the HT cohort and that the post-test results may indicate significance at increasing the CRCT reading comprehension performances of third grade students. From pre-test to post test the VUL cohort increased 480 points whereas the HT cohort increased 482 points from pre-test to post-test. The VUL cohort started with a higher mean score than did the HT cohort and maintained most of the difference over the course of the intervention.

In order to clearly establish a baseline for this study, the post-test 2006 CRCT scale reading scores for VUL and HT were adjusted using the pre-test 2005 CRCT scale reading scores as the covariate. The analysis of covariance (ANCOVA) was used for 2 independent samples where the independent samples of HT and VUL present 2 categorical levels of the independent variable (reading program) and the dependent variable is the pre-test 2005 CRCT pretest.
Table 7.1 contains the April 2006 scale CRCT estimated marginal means, standard error, and mean differences for HT and VUL students, F-value, significance and partial eta squared.

The Bonferroni (1998) method was used to adjust for multiple comparisons. Using the Bonferroni method, the alpha level (chance taken by researchers to make a Type I error) is adjusted downward to ensure that the overall (experiment-wise) risk for a number of tests remains $p = 0.05$. Even if more than one test is done, the risk of making Type I errors or finding an effect incorrectly significant continues to be at 0.05. But reducing the chance of Type I errors could increase the chance of Type II errors. If a single hypothesis of no effect is tested using more than one test, and the hypothesis is rejected if one of the tests shows statistical significance, Bonferroni correction should be applied (Bonferroni, 1998). Table 7.1 shows that the adjusted means are significantly different. With an adjusted CRCT mean of 823.5 for the VUL cohort and an adjusted CRCT mean of 819.7 for the HT cohort the univariate tests of difference of means ($F = 6.586, df = 742$) $p = .010$ indicated there was a significant difference at $p \leq .05$. The effect size for the reading programs is small as indicated by the partial eta squared of .009.
Table 7.1

Estimated Marginal and Pairwise Comparison of HT and VUL Means for CRCT 2006

<table>
<thead>
<tr>
<th>Rdg.</th>
<th>Std.</th>
<th>Mean</th>
<th>Part.</th>
<th>Diff/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prog</td>
<td>n</td>
<td>X</td>
<td>Error</td>
<td>Diff.</td>
</tr>
<tr>
<td>HT</td>
<td>441</td>
<td>819.7</td>
<td>.962</td>
<td>-3.871</td>
</tr>
<tr>
<td>VUL</td>
<td>304</td>
<td>823.5</td>
<td>1.159</td>
<td>3.871</td>
</tr>
</tbody>
</table>

p≤0.05

Table 8 contains the September, 2005, reading ITBS standard score data collected from the cohorts of Harcourt Trophies (HT) students (n=527) and Voyager Universal Literacy (VUL) students (n=397). The ITBS assessed vocabulary and reading comprehension. The ITBS scores are reported as percentiles, raw scores, standard scores, grade equivalents and normal curve equivalents (NCE). A standard score describes a student's performance along a standard scale (standard scores are in Appendix E).
Table 8

Comparison of Pre- and Post-test CRCT and ITBS Means and Differences for HT and VUL

<table>
<thead>
<tr>
<th>Test</th>
<th>Rdg.</th>
<th>n</th>
<th>X</th>
<th>SD</th>
<th>Diff.</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>HT</td>
<td>442</td>
<td>335.77</td>
<td>42.7</td>
<td>-7.490</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRCT 2005</td>
<td>VUL</td>
<td>304</td>
<td>343.26</td>
<td>44.8</td>
<td>-7.490</td>
<td>-2.308</td>
<td></td>
</tr>
<tr>
<td>Post-Test</td>
<td>HT</td>
<td>527</td>
<td>168.22</td>
<td>19.9</td>
<td>-2.352</td>
<td></td>
<td>p≤.021</td>
</tr>
<tr>
<td>ITBS 2005</td>
<td>VUL</td>
<td>397</td>
<td>170.57</td>
<td>19.3</td>
<td>-2.352</td>
<td>-1.801</td>
<td>p≤.072</td>
</tr>
</tbody>
</table>

With a mean ITBS standard score of 168.22 for the HT cohort and a mean ITBS standard score of 170.57 for the VUL cohort a 2-tailed t-test of difference of means ($t = -1.801$, $df=922$) $p ≤ .072$ indicated that the difference was not significant at $p ≤ .05$. This statistical result of testing in the beginning of the school year indicates that although the VUL group surpassed the HT group, the results were not significantly different at $p ≤ .05$. Absolute comparison of the standard deviation is problematic since
the standard deviations can be influenced by the size of
the mean. Any difference between HT and VUL, ITBS standard
scores could be attributed to chance.

Table 8.1 contains the Fall 2005 ITBS standard
estimated marginal means, standard error, mean difference
for HT and VUL students, F-value, significance and partial
eta squared. The table shows that the adjusted means are
not significantly different. With an adjusted ITBS mean of
170.8 for the VUL cohort and an adjusted ITBS mean of 168.4
for the HT cohort, the univariate tests of differences of
means ($F = 3.761$, $df = 730$), $p = .053$ indicate a marginal
significant difference at $p \leq .05$. Any differences between
the HT and VUL reading levels could be attributed to
chance.
Table 8.1

Estimated Marginal and Pairwise Comparison of HT and VUL Means for ITBS 2005

<table>
<thead>
<tr>
<th>Prog.</th>
<th>N</th>
<th>X</th>
<th>Error</th>
<th>Diff.</th>
<th>F</th>
<th>Eta²</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT</td>
<td>434</td>
<td>168.4</td>
<td>.794</td>
<td>-2.415</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VUL</td>
<td>299</td>
<td>170.8</td>
<td>.957</td>
<td>2.415</td>
<td>3.761</td>
<td>.005</td>
<td></td>
</tr>
</tbody>
</table>

p≤.053

p≤.05

Effect size estimates the magnitude of a treatment effect. Cohen (1988) defined $d$ as the difference between the means, $M_1 - M_2$, divided by standard deviation, $\sigma$, of either group. Cohen argued that the standard deviation of either group could be used when the variances of the two groups are homogeneous. To compare the effect size (ES) of the differences in the mean gains between HT and VUL, ES was calculated for each cohort by subtracting the mean of the VUL group from the mean of the HT group and dividing by the standard deviation (SD) of either HT or VUL. The results are as follows:

CRCT: \[
\frac{818.11 - 823.6}{26.584} = -.20918 \text{ or } .21
\]
ITBS: \((168.22-170.57)/19.910 = -.11813\) or \(.12\)

Effect sizes can also be thought of as the average percentile standing of the average treated (or experimental) participant relative to the average untreated (or comparison) participant. An effect size of .21 for CRCT indicated that students who received VUL improved about .2 standard deviation (SD) more than did students who received HT. An effect size of .21 indicates that the mean of the treated group (VUL) is at approximately the 58th percentile of the untreated (HT) group.

**Analysis of Variance (ANOVA)**

One analysis used a 2 x 2 x 2 factorial design to study the differences of treatment for three independent variables, SES, gender and SWD on the dependent variable, CRCT reading performance scores. Table 9 shows the between-subject factors. Main effect_1 - SES included 907 students receiving free/reduced lunches (coded 1) and 159 students not on free/reduced lunches. Of the 1066 students, 542 were males (main effect_2 coded 1) and 524 were females (coded 2). Main effect_3, SWD included 126 students with Individual Education Plans (IEP) (coded 1). Of the participating students 601 were in the HT group and 465 were in the VUL group.
Table 9

Between Subject Factors (1066 students, 2005-06) for 

Analysis of Variance of CRCT

<table>
<thead>
<tr>
<th>Factor</th>
<th>Code</th>
<th>Label</th>
<th>HT N</th>
<th>HT n</th>
<th>VUL N</th>
<th>VUL n</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>0</td>
<td>No F/R</td>
<td>159</td>
<td>74</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Yes F/R</td>
<td>907</td>
<td>527</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>0</td>
<td>Female</td>
<td>524</td>
<td>294</td>
<td>230</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Male</td>
<td>542</td>
<td>307</td>
<td>235</td>
<td></td>
</tr>
<tr>
<td>SWD</td>
<td>N</td>
<td>No IEP</td>
<td>940</td>
<td>534</td>
<td>406</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>Has IEP</td>
<td>126</td>
<td>67</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Rdg</td>
<td>HT</td>
<td>Harcourt</td>
<td>601</td>
<td>n/a</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Prog</td>
<td>VUL</td>
<td>Voyager</td>
<td>465</td>
<td>n/a</td>
<td>n/a</td>
<td></td>
</tr>
</tbody>
</table>

Research Question 2

What difference does VUL or HT reading instruction have on reading achievement of third-grade students in Fulton County, Georgia using (a) SES, (b) gender, or (c) SWD as the main effects measured by state assessments?

A summary of the cell means and standard deviations for the ANOVA analysis appears in Table 10.
Table 10

Summary of 2006 CRCT Means and Standard Deviations

Independent Variables: Reading Programs (HT, VUL);
SES (Free/Reduced Lunch 1,0); Gender (Male, Female);
Special Education (Non-IEP students, IEP students)

<table>
<thead>
<tr>
<th>Main Effect</th>
<th>HT/VUL</th>
<th>SWD</th>
<th>SES</th>
<th>M</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td>823.17</td>
<td>26.144</td>
<td>294</td>
</tr>
<tr>
<td></td>
<td>HT</td>
<td></td>
<td></td>
<td>825.93</td>
<td>28.001</td>
<td>230</td>
</tr>
<tr>
<td></td>
<td>VUL</td>
<td></td>
<td></td>
<td>824.36</td>
<td>26.983</td>
<td>524</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>813.26</td>
<td>26.136</td>
<td>307</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td>821.46</td>
<td>27.640</td>
<td>235</td>
</tr>
<tr>
<td></td>
<td>HT</td>
<td></td>
<td></td>
<td>816.82</td>
<td>27.080</td>
<td>542</td>
</tr>
<tr>
<td></td>
<td>VUL</td>
<td></td>
<td></td>
<td>818.11</td>
<td>26.584</td>
<td>601</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>823.67</td>
<td>27.879</td>
<td>465</td>
</tr>
<tr>
<td>Rdg Tot HT</td>
<td></td>
<td></td>
<td></td>
<td>820.54</td>
<td>27.284</td>
<td>1066</td>
</tr>
<tr>
<td>Prog Tot VUL</td>
<td></td>
<td></td>
<td></td>
<td>827.70</td>
<td>29.056</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>HT</td>
<td>No F/R</td>
<td>833.06</td>
<td>32.774</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VUL</td>
<td>No F/R</td>
<td>930.57</td>
<td>31.117</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>No F/R</td>
<td>816.76</td>
<td>25.966</td>
<td>527</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HT</td>
<td>Yes F/R</td>
<td>820.54</td>
<td>27.284</td>
<td>1066</td>
<td></td>
</tr>
</tbody>
</table>

Table cont.
Table 10 (cont.)

<table>
<thead>
<tr>
<th></th>
<th>Yes F/R</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VUL</td>
<td>821.57</td>
<td>26.255</td>
<td>360</td>
</tr>
<tr>
<td>ALL</td>
<td>818.78</td>
<td>26.181</td>
<td>907</td>
</tr>
<tr>
<td>SWD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT</td>
<td>820.54</td>
<td>25.689</td>
<td>534</td>
</tr>
<tr>
<td>VUL</td>
<td>825.41</td>
<td>27.342</td>
<td>406</td>
</tr>
<tr>
<td>Total</td>
<td>822.64</td>
<td>26.511</td>
<td>940</td>
</tr>
<tr>
<td>HT</td>
<td>798.72</td>
<td>25.806</td>
<td>67</td>
</tr>
<tr>
<td>VUL</td>
<td>811.71</td>
<td>28.831</td>
<td>59</td>
</tr>
<tr>
<td>ALL</td>
<td>804.80</td>
<td>27.923</td>
<td>126</td>
</tr>
</tbody>
</table>

Inspection of mean scores showed that VUL students outperformed HT students as a whole (823.67 : 818.11). Low SES -VUL students outperformed low SES-HT students (821.57 : 816.57). Non-SES VUL students outperformed Non-SES HT students (833.06 : 827.70); as well VUL students with IEPs (VUL - 811.71 : HT-798.72). As a whole, females outperformed males using both reading programs (females, 824.38; males 816.82). VUL males outperformed HT males (821.56 : 813.26).

The criterion for statistical significance in this study was set at $p \leq .05$; however actual significance
levels were reported. Three univariate analyses of variance (ANOVA), in which gender, SES-ED, SWD and the reading programs were the independent variables and reading performance on the CRCT was the dependent variable revealed that the models were significant. The model for gender was significant ($F = 11.611$, df = 3,1062, $\text{sig} = .000$). In addition, main effect of the reading program is reported as significant ($F = 10.882$, df 1/1065, $\text{sig.} = .001$). The main effect of gender was significant ($F = 18.760$, df = 1/1065, $\text{sig} = .000$). The interaction between reading program and gender was not significant. The tests of between-subjects effects are presented in Tables 11-13.

Table 11

Tests of Between-Subjects Effects - Gender

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>25177.718(a)</td>
<td>3</td>
<td>8392.573</td>
<td>11.611</td>
<td>.000</td>
<td>.032</td>
</tr>
<tr>
<td>Intercept</td>
<td>706566.432.7</td>
<td>70</td>
<td>1</td>
<td>977543.48</td>
<td>.000</td>
<td>.999</td>
</tr>
<tr>
<td>Reading_Prg</td>
<td>7865.247</td>
<td>1</td>
<td>7865.247</td>
<td>10.882</td>
<td>.001</td>
<td>.010</td>
</tr>
<tr>
<td>Student_Gender_Desc</td>
<td>13559.638</td>
<td>1</td>
<td>13559.638</td>
<td>18.760</td>
<td>.000</td>
<td>.017</td>
</tr>
<tr>
<td>Reading_Prg * Student_Gender_Desc</td>
<td>1929.633</td>
<td>1</td>
<td>1929.633</td>
<td>2.670</td>
<td>.103</td>
<td>.003</td>
</tr>
<tr>
<td>Error</td>
<td>767611.428</td>
<td>1062</td>
<td>722.798</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>718507935.0</td>
<td>1066</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>792789.145</td>
<td>1065</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* R Squared = .032 (Adjusted R Squared = .029)
A factorial analysis of variance (ANOVA), in which reading program and SES (economically disadvantaged [ED]) were the independent variables reveals that the model was significant (F=11.546, df = 3,1062, sig. = .000). In addition the main effect of reading programs is significant at (F=4.794, df = 1/1062, sig. = .029) and SES-ED is significant at (F= 23.343, df = 1/1062, sig. = .000). The interaction between reading programs and SES is not significant. The tests between-subject effects for SES-ED are presented in Table 12.

Table 12

Tests of Between-Subjects Effects – SES-ED

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>25040.547(a)</td>
<td>3</td>
<td>8348.849</td>
<td>11.546</td>
<td>.000</td>
<td>.032</td>
</tr>
<tr>
<td>Intercept</td>
<td>365146083.4</td>
<td>61</td>
<td>365146083.46</td>
<td>505083.90</td>
<td>.000</td>
<td>.998</td>
</tr>
<tr>
<td>Reading_Prg</td>
<td>3466.071</td>
<td>1</td>
<td>3466.071</td>
<td>4.794</td>
<td>.029</td>
<td>.004</td>
</tr>
<tr>
<td>ED</td>
<td>16875.049</td>
<td>1</td>
<td>16875.049</td>
<td>23.343</td>
<td>.000</td>
<td>.022</td>
</tr>
<tr>
<td>Reading_Prg * ED</td>
<td>10.070</td>
<td>1</td>
<td>10.070</td>
<td>.014</td>
<td>.806</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>767748.588</td>
<td>1062</td>
<td>722.927</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>718507335.0</td>
<td>1066</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>792782.145</td>
<td>1065</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a R Squared = .032 (Adjusted R Squared = .029)

A factorial analysis of variance (ANOVA), in which reading program and SWD were the independent variables reveals that the model was significant (F=21.872, df = 3, 1062, sig. =
.000). In addition the main effect of reading programs is significant at ($F=12.531$, df = $1/1062$, sig. = .000) and SWD is significant at ($F= 49.569$, df = $1/1062$, sig. = .000). The interaction between reading programs and SWD is not significant.

Table 13

Tests of Between-Subjects Effects – Students with Disabilities (SWD)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>46132.794(a)</td>
<td>3</td>
<td>16377.598</td>
<td>21.872</td>
<td>.000</td>
<td>.058</td>
</tr>
<tr>
<td>Intercept</td>
<td>2292845681.8</td>
<td>1</td>
<td>2292845681.8</td>
<td>416526.44</td>
<td>.000</td>
<td>.997</td>
</tr>
<tr>
<td>Reading_Prg</td>
<td>8810.325</td>
<td>1</td>
<td>8810.325</td>
<td>12.531</td>
<td>.000</td>
<td>.012</td>
</tr>
<tr>
<td>Ind_Special_Ed_This_Year</td>
<td>34850.091</td>
<td>1</td>
<td>34850.091</td>
<td>49.569</td>
<td>.000</td>
<td>.045</td>
</tr>
<tr>
<td>Reading_Prg * Ind_Special_Ed_This_Year</td>
<td>1825.208</td>
<td>1</td>
<td>1825.208</td>
<td>2.596</td>
<td>.107</td>
<td>.002</td>
</tr>
<tr>
<td>Error</td>
<td>746696.39</td>
<td>1062</td>
<td>703.066</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>718507955.0</td>
<td>1066</td>
<td>1066</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>792789.145</td>
<td>1065</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a R Squared = .058 (Adjusted R Squared = .056)

Measures of effect size in ANOVA are measures of the degree of association between an effect (e.g. a main effect, an interaction, a linear contrast) and the dependent variable. Partial Eta squared, ($\eta_p^2$) is a commonly used measure of effect size in ANOVA (Becker, 2000). Table 14 summarizes the comparison of progress between HT and VUL, the differences in the means and the partial Eta
squared for the independent variables. Table 14 shows that the gains in the independent variables (VUL, HT, gender, SES and SWD) were larger than those in the comparison groups in all areas. Males and SWD showed the largest differences in means among the two reading programs. VUL males and SWD showed the largest differences in means.
### Table 14

*Comparison of Progress between HT and VUL cohorts on CRCT*

#### 2006 means and differences

<table>
<thead>
<tr>
<th>Factors</th>
<th>HT</th>
<th>VUL</th>
<th>Diff.</th>
<th>Sig.</th>
<th>Partial Eta</th>
<th>HT-VUL</th>
<th>Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Gender</td>
<td>818.22</td>
<td>823.70</td>
<td>-5.48</td>
<td>.001</td>
<td>.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>823.17</td>
<td>825.93</td>
<td>-2.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>813.26</td>
<td>821.46</td>
<td>-8.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SES-ED</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.000</td>
<td>.022</td>
</tr>
<tr>
<td>Total SES</td>
<td>822.23</td>
<td>827.32</td>
<td>-5.09</td>
<td>.029</td>
<td>.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES-ED - N</td>
<td>827.70</td>
<td>833.06</td>
<td>-5.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES-ED - Y</td>
<td>816.76</td>
<td>821.57</td>
<td>-4.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SWD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.000</td>
<td>.045</td>
</tr>
<tr>
<td>Total SWD</td>
<td>809.63</td>
<td>818.56</td>
<td>-8.93</td>
<td>.000</td>
<td>.012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWD - N</td>
<td>820.54</td>
<td>825.41</td>
<td>-4.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWD - Y</td>
<td>798.72</td>
<td>811.71</td>
<td>-12.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this study the 3rd grade mean Pre-2005 CRCT reading scores, the 3rd grade mean Post-2006 CRCT reading scores and mean differences were compared for VUL and HT cohorts by individual schools to the sample means. The cumulative gain in mean differences of pre-2005 CRCT and post-2006 CRCT reading scores of the VUL schools was larger in a positive
direction compared to the cumulative gain in mean differences in the HT schools (pre-2005 CRCT +7.81 : -20.00; post-2006 CRCT +18.63 : -11.73). Table 15 displays the individual school results. Tables 15–18 display similar results for gender, SES and SWD. The VUL program showed higher gains with male, female, free lunch, eligible SWD and students without disabilities than did similar cohorts using the HT program (male= +24.96 : -12.82; female= +10.86 : -10.42; free lunch= +5.95 : -27.12; students with disabilities= -9.86 : -131.08; students without disabilities= +23.70 : +6.81).

The HT schools were more successful with reduced lunch students as noted on the CRCT 2006 CRCT cumulative gain in mean differences (VUL = +32.42: HT = +41.67). One possible cause for the HT gain in this area may be the noticeable mean difference in one particular HT school of +40.26 points. This score is an outlier and does not follow historical trends nor a reasonable trajectory for gains in student achievement and the test group was very small (n=5). If these scores were excluded from the calculations then the pattern noted earlier for VUL schools [larger cumulative mean differences than HT] would most likely hold true for the reduced lunch students.
Table 15

Comparison of Reading Programs by schools, 2005 CRCT

Reading Means, ITBS Reading Means, 2006 CRCT Reading Means

and Mean Differences from sample means

<table>
<thead>
<tr>
<th></th>
<th>CRCT</th>
<th>ITBS</th>
<th>CRCT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2005</td>
<td>2005</td>
<td>2006</td>
</tr>
<tr>
<td>VUL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schools</td>
<td>X</td>
<td>n</td>
<td>Diff</td>
</tr>
<tr>
<td>1**</td>
<td>326.85</td>
<td>20</td>
<td>-11.97</td>
</tr>
<tr>
<td>2</td>
<td>332.87</td>
<td>39</td>
<td>-5.95</td>
</tr>
<tr>
<td>3</td>
<td>345.29</td>
<td>62</td>
<td>+6.47</td>
</tr>
<tr>
<td>4</td>
<td>359.54</td>
<td>76</td>
<td>+20.72</td>
</tr>
<tr>
<td>5</td>
<td>337.36</td>
<td>107</td>
<td>-1.46</td>
</tr>
<tr>
<td>Total</td>
<td>X=343.26</td>
<td>442</td>
<td>+7.61*</td>
</tr>
<tr>
<td>HT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schools</td>
<td>X</td>
<td>n</td>
<td>Diff</td>
</tr>
<tr>
<td>6</td>
<td>331.32</td>
<td>110</td>
<td>-7.50</td>
</tr>
<tr>
<td>7</td>
<td>328.79</td>
<td>68</td>
<td>-10.03</td>
</tr>
<tr>
<td>8</td>
<td>332.79</td>
<td>65</td>
<td>-5.83</td>
</tr>
<tr>
<td>9</td>
<td>340.65</td>
<td>107</td>
<td>+1.83</td>
</tr>
<tr>
<td>10</td>
<td>348.85</td>
<td>60</td>
<td>+10.03</td>
</tr>
<tr>
<td>11**</td>
<td>330.32</td>
<td>28</td>
<td>-8.50</td>
</tr>
<tr>
<td>Total</td>
<td>X=335.77</td>
<td>304</td>
<td>-20.00*</td>
</tr>
<tr>
<td>Sample</td>
<td>X=338.82</td>
<td>746</td>
<td></td>
</tr>
</tbody>
</table>

*Cumulative Mean Gains

**VUL 1 and HT 11 cohorts within one school.
Table 16

Comparison of Reading Programs by schools and Gender, 2006

CRCT Reading Means and Mean Differences from sample means

<table>
<thead>
<tr>
<th>VUL</th>
<th>CRCT 2006 - Males</th>
<th>CRCT 2006 - Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schools</td>
<td>X</td>
<td>n</td>
</tr>
<tr>
<td>1**</td>
<td>820.70</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>824.03</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>821.38</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>827.11</td>
<td>62</td>
</tr>
<tr>
<td>5</td>
<td>815.84</td>
<td>77</td>
</tr>
<tr>
<td>Total</td>
<td>821.46</td>
<td>235</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NT</th>
<th>CRCT 2006 - Males</th>
<th>CRCT 2006 - Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schools</td>
<td>X</td>
<td>n</td>
</tr>
<tr>
<td>6</td>
<td>805.42</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>811.50</td>
<td>38</td>
</tr>
<tr>
<td>8</td>
<td>820.62</td>
<td>39</td>
</tr>
<tr>
<td>9</td>
<td>816.56</td>
<td>79</td>
</tr>
<tr>
<td>10</td>
<td>816.86</td>
<td>43</td>
</tr>
<tr>
<td>11**</td>
<td>817.22</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>813.26</td>
<td>307</td>
</tr>
</tbody>
</table>

Sample | X=816.82 | 542 | 824.38 | 524

*Cumulative Mean Gains

**VUL 1 and HT 11 cohorts within one school.
Table 17

Comparison of Reading Programs by schools and Free/Reduced Lunch - SES, 2006 CRCT Reading Means and Mean Differences from sample means

<table>
<thead>
<tr>
<th></th>
<th>CRCT 2006 - Free</th>
<th>CRCT 2006 - Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VUL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schools</td>
<td>X</td>
<td>n</td>
</tr>
<tr>
<td>1**</td>
<td>822.58</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>824.68</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>823.92</td>
<td>77</td>
</tr>
<tr>
<td>4</td>
<td>823.38</td>
<td>73</td>
</tr>
<tr>
<td>5</td>
<td>914.29</td>
<td>107</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>X=821.57</td>
<td>337</td>
</tr>
<tr>
<td><strong>HT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schools</td>
<td>X</td>
<td>n</td>
</tr>
<tr>
<td>6</td>
<td>808.02</td>
<td>137</td>
</tr>
<tr>
<td>7</td>
<td>810.66</td>
<td>71</td>
</tr>
<tr>
<td>8</td>
<td>825.30</td>
<td>77</td>
</tr>
<tr>
<td>9</td>
<td>823.35</td>
<td>103</td>
</tr>
<tr>
<td>10</td>
<td>815.53</td>
<td>53</td>
</tr>
<tr>
<td>11**</td>
<td>813.26</td>
<td>31</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>X=816.76</td>
<td>472</td>
</tr>
</tbody>
</table>

**Cumulative Mean Gains**

**VUL 1 and HT 11 cohorts within one school.**
Table 18

Comparison of Reading Programs by schools and students with disabilities (SWD), students without disabilities (no-IEPs) 2006 CRCT Reading Means and Mean Differences from sample means

<table>
<thead>
<tr>
<th></th>
<th>CRCT 2006</th>
<th>SWD-IEP</th>
<th>CRCT 2006</th>
<th>No IEP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VUL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schools</td>
<td>X</td>
<td>n</td>
<td>Diff</td>
<td>X</td>
</tr>
<tr>
<td>1**</td>
<td>839.33</td>
<td>3</td>
<td>+18.79</td>
<td>821.05</td>
</tr>
<tr>
<td>2</td>
<td>812.18</td>
<td>11</td>
<td>-8.36</td>
<td>828.74</td>
</tr>
<tr>
<td>3</td>
<td>823.83</td>
<td>12</td>
<td>+3.29</td>
<td>825.51</td>
</tr>
<tr>
<td>4</td>
<td>819.20</td>
<td>10</td>
<td>-1.34</td>
<td>828.23</td>
</tr>
<tr>
<td>5</td>
<td>798.30</td>
<td>23</td>
<td>-22.24</td>
<td>822.87</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>X=811.71</td>
<td>59</td>
<td><strong>-9.86</strong>*</td>
<td>825.41</td>
</tr>
<tr>
<td><strong>HT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schools</td>
<td>X</td>
<td>n</td>
<td>Diff</td>
<td>X</td>
</tr>
<tr>
<td>6</td>
<td>787.00</td>
<td>17</td>
<td>-33.54</td>
<td>812.45</td>
</tr>
<tr>
<td>7</td>
<td>803.00</td>
<td>7</td>
<td>-17.54</td>
<td>814.82</td>
</tr>
<tr>
<td>8</td>
<td>803.75</td>
<td>16</td>
<td>-16.79</td>
<td>831.53</td>
</tr>
<tr>
<td>9</td>
<td>807.41</td>
<td>17</td>
<td>-13.13</td>
<td>826.12</td>
</tr>
<tr>
<td>10</td>
<td>809.00</td>
<td>4</td>
<td>-11.54</td>
<td>820.51</td>
</tr>
<tr>
<td>11**</td>
<td>782.00</td>
<td>6</td>
<td>-38.54</td>
<td>824.57</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>X=798.72</td>
<td>67</td>
<td><strong>-131.08</strong>*</td>
<td>820.54</td>
</tr>
</tbody>
</table>

Sample X=820.54 126 820.54 940

*Cumulative Mean Gains

**VUL 1 and HT 11 cohorts within one school.
For this study the guidelines for effect size (ES) interpretation were: small = .20, Medium = .50, and large = .80 or greater (Becker, 2000). The partial Eta squared was .017 for student gender, .022 for SES and .045 for SWD. The partial Eta squared for SWD of .045 means that SWD by itself accounted for 4.5% of the overall variance in the means.

While there were no large effect sizes, on average VUL students who were low SES or SWD scored higher than HT students in the same subgroups. The profile plots for gender, SES and SWD are presented in Appendix F. All profile plots show ordinal relationships between VUL and HT in gender, SES and SWD.

Interview with Voyager District Coach

In an effort to provide some qualitative descriptions, a single interview was held with the VUL District Coach who worked in all of the Voyager schools. This individual was a full time employee of the district paid exclusively from federal Title I - Part A funds.

The researcher targeted the perspective of the coach regarding program implementation and student achievement. The core questions were structured but the actual interview was informal in nature to allow for depth of responses.
Interview Questions and Responses (see Appendix G for full text)

The interview with the VUL district coach revealed that the school system purchased two HT developed training components in vocabulary and comprehension. System Lead Reading Teachers (LRT) redelivered the HT training components to all elementary teachers shortly after the reading/language arts textbook adoption cycle. The training on the two HT components was delivered once during the textbook cycle. It was incumbent upon the teachers to determine how best to utilize all of the materials for the HT reading program which resulted in independent and isolated reading instruction implementation in the classroom.

The VUL district coach noted the sharp contrast in the professional development components of the two reading programs. VUL producers provided continuous professional development and ongoing support for all VUL classrooms which were job-embedded and major structural components included in the VUL program. HT producers provided neither professional development nor ongoing support.
If HT support programs were available, the school system did not participate nor purchase such programs. VUL producers provided the structure and support to teachers who did not know how to teach reading or additional strategies for teachers wanting to refine reading instruction practices.

Summary

The pre-test CRCT reading performance scores showed significant difference indicating that third-grade VUL students demonstrated higher success on CRCT 2005. The VUL cohort had a significant mean pre-test CRCT difference of 7.49 greater than did the HT cohort. The pre-test and post-test differences between HT and VUL demonstrated that the HT cohort reduced the mean difference by 1.929 points more than did the VUL cohort. Results on the ITBS were not statistically different between the two programs.

The comparisons of CRCT mean differences between the two reading programs and the school system demonstrate large differences. The percentage differences of Level 1 students (students not meeting state standards) in VUL and HT demonstrate similar mean differences between 2005 and 2006 CRCT scores with fewer students not meeting state standards in the VUL cohort.
In Chapter IV data were analyzed primarily using a quasi-experimental, quantitative approach. A single interview with the VUL District Coach was used along with the quantitative analysis to determine integrity of implementation and to add a historical perspective to the curricula aspects of HT and VUL. Data to support the summary tables in Chapter IV are in Appendix D-F. The district will undergo textbook adoption in 2009-2010 and the system leaders will need to decide whether to sustain VUL or switch to the new textbook pending adoption.

Chapter V provides a summary of the analyses, study discussions, limitations and recommendations for policy practice and additional research.
Chapter V

Introduction, Summary, Discussions, Conclusions and Recommendations

The purpose for this study was to determine if any statistically significant (p < .05) differences existed between the reading comprehension performances of third-grade students using VUL and third-grade students using HT in relationship to students' socioeconomic status (SES), students' gender, and students with disabilities (SWD). The study analyses were based on the outcome measures of the cohorts on the Georgia Criterion Referenced Competency Tests (CRCT) and the Iowa Tests of Basic Skills (ITBS).

The cumulative gain in mean difference in the pre-2005 CRCT and post-2006 CRCT reading scores of students in the VUL schools was larger in a positive direction in comparison to the cumulative gain in mean differences in the Harcourt (HT) schools that were used as the comparison group. The VUL program demonstrated higher gains with male, female, low SES, SWD and students without disabilities than did similar cohorts using the HT program. An effect size of .21 for CRCT indicated that VUL students improved about .2 standard deviation (SD) units more than did HT students.
The analyses of the mean reading performances of third-grade students on the Georgia CRCT pre-test and post-test indicates that students instructed in VUL started out ahead of students instructed in HT. The CRCT post-test results may indicate significance at increasing CRCT reading comprehension performances. Third-grade students who participated in VUL and were males and/or SWD were more likely to score higher with VUL than with HT on the Georgia CRCT 2006 post-test. The higher scores for SWD increased the likelihood of selected elementary schools making Adequate Yearly Progress (AYP). This important outcome (AYP) was previously mentioned in Chapter IV, by the VUL District Coach as being positively influenced by the implementation of VUL in the selected schools.

As discussed earlier in Chapter II, the Follow Through (FT) study was perceived a failure because no tested reading instruction models, except one, produced the desired results. The only model that brought children close to the 50th percentile in all subject areas was the Direct Instruction (DI) model (Grossen, 1995).

Other research supports the efficacy of the DI model. O'Neill (1988) previously had reviewed 150 studies and concluded that DI was correlated with improved learning among primary children from working and middle-class
backgrounds - a conclusion supported by almost every relevant study on the topic.

As cited in Kamps, et al. (2003), a descriptive study supported findings of prior studies reporting significant effect sizes for DI and peer-tutoring reading interventions for elementary students with Emotional Behavioral Disorders (EBD) (Coleman & Vaughn, 2000). The same study (as cited in Kamps et al.) showed that the DI reading curriculum demonstrated higher growth in reading fluency than did the literature-based instruction.

DI emphasized the systematic teaching of phonemic awareness and language skills among other techniques. Follow up studies (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977; Bock, Stebbins, & Proper, 1977; Meyer, 1984) showed that the effects of DI are sustainable. Further studies (as cited in e.g., Gersten & Keating, 1987; Gersten, Keating, & Becker, 1988; Adams, 1995) showed the long-term benefits of DI on student achievement.

Rosenshine (2002) examined the effectiveness of DI and Reading Mastery (RM) as instructional methods to help students read at grade level. Rosenshine noted that in future years, reading researchers should report standardized test scores, means and standard deviations for third- and fourth-grade students grouped by years in a
program. Standardized tests in reading begin with a focus on decoding in first grade and shift to a focus on reading comprehension in third grade and higher. Therefore, success for a program with students from poor families must involve nothing less than helping students read [with comprehension] at grade level, on standardized tests, at the end of third grade and higher (Rosenshine). Hodgkinson (1995), Rothstein (2004) and others have shown that the best predictor for school failure is poverty. Given that reality, educators must proactively seek solutions to address the social issues and risk factors in tandem with the academic achievement issues.

Class-size reduction is one solution that has been shown to, among other things, improve academic achievement for all students and particularly for low-income and minority students (Achilles, Finn, Gerber, & Pannozzo, 2000). Summers and Wolfe (1977) found that low-achieving students performed worse when in larger classes, while high-achieving students performed better. Thus the literature suggests that disadvantaged students - minorities, low-SES, and low-achievers - gain the most from small classes. This conclusion is also reinforced by the findings from the STAR class-size experiment (Finn et al., 2005; Finn et al., 2006; Word et al., 1990; Mosteller,
(1995). Levin, Belfield, Muennig & Rouse (2007) identified class size reduction as one of only five interventions that demonstrated improvements in high school graduation rates. The study by Levin et al. (2007, p. 22) has shown that by focusing resources on students who are receiving inadequate education, it is possible to obtain benefits far in excess of the costs of those investments.

For example, research has shown that attending small classes in the elementary grades (K-3) is accompanied by important long-term advantages, namely taking advanced course work in high school, taking college entrance examinations, and graduating from high school (Krueger & Whitmore, 2001; Achilles & Finn, 2005; Finn et al., 2005; Finn et al., 2006).

In the present study independent 2-tailed t-tests were performed using mean gain on the Georgia CRCT and ITBS reading assessments to determine the success of HT and VUL. These two assessments were administered to both the HT third-grade cohort and the VUL third-grade cohort. The Georgia CRCT 2005 assessment was used as the pre-test to establish a baseline for the study.

To determine the efficacy of expanding the VUL reading program in other schools with similar populations, the sample used in this study consisted of 1066 students who
attended HT schools \( n=601 \) or VUL schools \( n=465 \). Using data from a district database, the researcher analyzed the reading scores, based on the reading programs and instruction used exclusively at the specific schools.

All third-grade students who took the CRCT and the ITBS in the school system in 2005-2006, and those having CRCT and ITBS scores served as the population for the study. Those third grade students who attended HT schools with similar mobility (mobil) and SES percentages made up the comparison group. Those third grade students who attended VUL schools represented the treatment group for CRCT scale scores and ITBS standard scores. All students received reading instruction (HT or VUL) from Georgia certified teachers.

A summary of reflections and comments regarding implementation with integrity appears in this chapter because perceptions oftentimes influence the quality and delivery of instruction. Implementation with integrity has been a critical component that plays an important part in any program's success or failure. VUL has a larger initial cost than has traditional basal reading programs. If the VUL program could save systems monies in the long run if targeted groups such as male students, low SES students become more successful in reading then those students would
be less likely to become eligible for special education services. The majority of students identified as learning disabled have deficits in reading (Lyon, 1996). In Georgia, as in other states, most students identified for special education services are disproportionately African American and male.

The following sections of this chapter present a summary of findings, discussion, limitations of the present study, conclusions and suggested implications for future policy, practice and research.

Summary of Findings

The demographic data for this study revealed that most students were African American [HT - 84%; VUL - 93%] and eligible for free or reduced price (FR) meals as a proxy for socioeconomic status (SES) [88 : 82]. The district had an overall population of African American (40%) and FR meals (40%) much lower than the sample. The percentages for gender in the HT and VUL schools [females 49:49; males 51 : 51] and students with disabilities (SWD) [11 : 13] were consistent with the district percentages of 49% female, 51% male and 15% SWD.

Pre-test Scores

The 2005 CRCT pre-test scores of HT and VUL students showed a significant difference indicating that third grade
VUL students demonstrated higher success on CRCT 2005 reading scores than did third-grade HT students. The 2005 CRCT pre-test scores may have existed because the VUL program included structured direct instruction in phonics, phonemic awareness and fluency in grades K-1 as well as additional time and support for struggling readers. The VUL cohort had a significant mean pre-test difference of 7.49 greater than the HT cohort [See Table 7 data and t-test results, p.126].

Students in both cohorts were tested in 2005 and 2006 using the Georgia Criterion Referenced Competency Tests (CRCT) and Iowa Tests of Basic Skills (ITBS). Scale scores and standard scores for reading were extracted from the district database. ITBS 2006 scores were not included.

Post-test Scores

The 2006 CRCT reading scores of HT and VUL cohorts indicated that there were significant differences [See Table 7. p.126 and Table 7.1, p.129, data and t-test results]. The CRCT mean difference of HT and VUL scores indicated that there was a significant difference between the cohorts in favor of the VUL cohort at pre-test. The mean difference in gains [See Table 19] between HT and VUL suggests that the HT cohort had a large mean gain that might suggest more success in moving students ahead quickly
but visual inspections of school-to-school scores reveals one outlier for the reduced-lunch, low SES students that may have influenced the HT scores in a positive direction. The pre-test and post-test differences in HT and VUL demonstrate that the HT cohort reduced the mean difference between the groups by 1.929 points [7.490-5.561 = 1.929].

Table 19

**Comparison of Reading programs by Mean Differences and 2005 CRCT pre-test and 2006 CRCT post-test**

<table>
<thead>
<tr>
<th>Test</th>
<th>Rdg.</th>
<th>X</th>
<th>Mean Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-test</strong></td>
<td>HT</td>
<td>335.77</td>
<td>-7.490</td>
</tr>
<tr>
<td><strong>CRCT 2005</strong></td>
<td>VUL</td>
<td>343.26</td>
<td>-7.490</td>
</tr>
<tr>
<td><strong>Post-test</strong></td>
<td>HT</td>
<td>818.11</td>
<td>-5.561</td>
</tr>
<tr>
<td><strong>CRCT 2006</strong></td>
<td>VUL</td>
<td>623.67</td>
<td>-5.561</td>
</tr>
<tr>
<td><strong>Mean Diff</strong></td>
<td>HT</td>
<td>482.34</td>
<td></td>
</tr>
<tr>
<td><strong>of Gains</strong></td>
<td>VUL</td>
<td>480.41</td>
<td></td>
</tr>
</tbody>
</table>

The researcher did mention the difference in scores to the area assistant superintendent of the particular school. All other scores for that particular school were within
normal range of the other reported scores. The small \( n \) \( [n=5] \) in the category discussed was most likely not large enough to significantly influence the overall mean scores of the sample.

The researcher has shown comparability on HT and VUL, and Table 20 below shows the comparison of the reading programs to system CRCT reading scores for 2005 and 2006.

Table 20

Comparison of Reading Programs by Mean Differences and Percentage of Level 1 students

<table>
<thead>
<tr>
<th>Test</th>
<th>Prog</th>
<th>X</th>
<th>Diff.</th>
<th>Avg. %</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>HT</td>
<td>335.77</td>
<td>-7.490</td>
<td>11.8</td>
<td>-4.0</td>
</tr>
<tr>
<td>CRCT 2005</td>
<td>VUL</td>
<td>343.26</td>
<td>-7.490</td>
<td>7.8</td>
<td>-4.0</td>
</tr>
<tr>
<td>FCSS 2005</td>
<td>ALL</td>
<td>357.20</td>
<td>N/A</td>
<td>5.0</td>
<td>N/A</td>
</tr>
<tr>
<td>Post-test</td>
<td>WT</td>
<td>818.11</td>
<td>-5.561</td>
<td>22.3</td>
<td>-4.8</td>
</tr>
<tr>
<td>CRCT 2006</td>
<td>VUL</td>
<td>823.67</td>
<td>-5.561</td>
<td>17.5</td>
<td>-4.8</td>
</tr>
<tr>
<td>FCSS 2006</td>
<td>ALL</td>
<td>838.00</td>
<td>N/A</td>
<td>12.0</td>
<td>N/A</td>
</tr>
</tbody>
</table>

In 2004-05 comparisons of mean differences between the reading programs and the district scores demonstrate large
differences \([\text{VUL} = -13.94; \text{HT} = -21.43]\). The 2005-06 comparisons of mean differences between the reading programs and the district scores demonstrate large mean differences \([\text{VUL} = -14.33; \text{HT} = -19.89]\). The percentage differences of Level 1 students in VUL and HT demonstrate similar mean differences between 2005 and 2006 CRCT scores with fewer students not meeting standards in the VUL cohort. It should be noted that the GaCRCT changed from objective-based test items in 2005 to standards-based test items in 2006. Visual inspection reveals that the VUL cohort had fewer third grade students in Level I than did the HT cohort.

A program evaluation (Emstar Research Inc. et al., 2006) conducted for the Fulton County School System (FCSS) found that in general, VUL students and comparison students performed comparably (p.10). There were some cases in which VUL students had higher mean CRCT scores than did comparison students, but differences between the groups were not statistically significant, with one exception — Reading scores for 3rd graders receiving four years of VUL instruction (Emstar Research Inc. et al., p. 10).

The following null hypotheses resulted from the primary research questions:
1. There is no difference in the reading comprehension performances of third-grade VUL students when compared to the performances of third-grade HT students as measured by the Iowa Tests of Basic Skills (ITBS).

2. There is no difference in the reading comprehension performances of third-grade VUL students when compared to the performances of third-grade HT students as measured on the Criterion Referenced Competency Tests (CRCT).

3. There is no difference in the reading comprehension performances of third-grade VUL students when compared to the performances of third-grade HT students using three variables as main effects as measured by the CRCT:
   a. socio-economic status (SES),
   b. gender, or
   c. students with disabilities (SWD).

There was not a significant difference between HT and VUL in 2005 in reading comprehension as measured by the ITBS. The ITBS was used as a post-test for the spring 2005 Georgia CRCT. With a mean ITBS standard score of 168.22 for the HT cohort and a mean ITBS standard score of 170.57 for the VUL cohort, a 2-tailed t-test of difference of means indicated that the difference [2.35] was not significant at
p \leq .05 \text{ [See Table 8, p.130 for data and t-test results].} 

The null hypothesis 1 is retained. Although the null hypothesis was retained there was a marginal significant difference (p \leq 0.053) \text{ [that could have been influenced by the size of the samples (n)], with the VUL cohort being more successful on the ITBS as demonstrated by mean difference of 2.415 for the VUL cohort.}

The results indicate that VUL students in 2005 and 2006 had significantly higher mean scores in CRCT reading performance than did HT students; thus null hypothesis 2 was not accepted.

ANOVA Results for post-test 2005-06 CRCT

The univariate analysis of variance (ANOVA) in which gender, SES, and SWD were the independent variables and reading performance on the CRCT was the dependent variable revealed significance differences for gender, ED and SWD in favor of the VUL cohorts. Comparison of progress between HT and VUL cohorts, means and differences are presented in Chapter 4, (Table 14 p.140 as well as the Table 21 below. Null hypothesis 3 was not accepted.
Table 21

Comparison of progress between HT and VUL cohorts on CRCT

2006 means and differences

<table>
<thead>
<tr>
<th>Factors</th>
<th>HT</th>
<th>VUL</th>
<th>HT-VUL</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td>.000</td>
</tr>
<tr>
<td>Reading pg.</td>
<td>818.22</td>
<td>823.70</td>
<td>-5.48</td>
<td>.001</td>
</tr>
<tr>
<td>SES</td>
<td></td>
<td></td>
<td></td>
<td>.022</td>
</tr>
<tr>
<td>Reading pg.</td>
<td>822.23</td>
<td>827.32</td>
<td>-5.09</td>
<td>.029</td>
</tr>
<tr>
<td>SWD</td>
<td></td>
<td></td>
<td></td>
<td>.000</td>
</tr>
<tr>
<td>Reading pg.</td>
<td>809.63</td>
<td>818.56</td>
<td>-8.93</td>
<td>.000</td>
</tr>
</tbody>
</table>

The results of the CRCT are used to categorize students into levels indicating (a) did not meet, (b) meets or (c) exceeds state standards. The VUL means indicated that most VUL students met or exceeded state standards on the 2006 CRCT. Table 22 shows that influence of VUL on third grade reading performance on the Georgia CRCT in 2006.
Table 22

The Influence of VUL on reading performance on Georgia CRCT 2006 - post-test

<table>
<thead>
<tr>
<th>When students...</th>
<th>students were more likely to...</th>
</tr>
</thead>
<tbody>
<tr>
<td>who were males took the CRCT,</td>
<td>score higher with VUL than with HT</td>
</tr>
<tr>
<td>who were females took the CRCT</td>
<td>score higher with VUL than with HT</td>
</tr>
<tr>
<td>who were SWD took the CRCT</td>
<td>score higher with VUL than with HT</td>
</tr>
<tr>
<td>who were low SES took the CRCT</td>
<td>score higher with VUL than with HT</td>
</tr>
<tr>
<td>who were not SWD (NO-IEP)</td>
<td>score higher with VUL than with HT</td>
</tr>
</tbody>
</table>

The interaction effects between gender and the reading program, SES and the reading program and SWD and the reading program were not significant in the test of between-subject effects. In the profile plots the lines are parallel in a positive direction and the interaction is not significant [See appendix G]. The VUL students were more successful than were the HT students in all categories.
However VUL students started higher than did HT students and the differences between groups reduced slightly (1.93 points) in favor of the HT group.

When CRCT 2005 reading scores are the co-variate, the tests between-subject effects reveal that the reading program is the best variable to explain the difference in the CRCT 2006 reading scale scores. The VUL cohort had higher mean scores than the HT cohort [See Table 7.1. p.128] When a similar test for covariance was applied to the ITBS scores, results were not significant [See Table 8.1. p.131].

Comparisons of CRCT reading scores by individual schools and reading programs demonstrate students in VUL schools scored higher than did students in HT schools in all categories. Females in either school (HT or VUL) scored higher on the 2005 CRCT regardless of the reading program.

Effect size as noted by Cohen (1988) measures the magnitude of a treatment effect. An effect size of .21 for CRCT indicated that students who received VUL improved about .2 standard deviation (SD) units more than did students who received HT. The partial Eta squared for SWD is 0.045 which means that by itself SWD accounted for 4.5% of the overall variance in the means. SES had the next
largest partial Eta squared at 0.022 or 2.2% of the overall variance in the means.

Limitations of the Present Study

Limitations

1. The study was limited by ethnicity because the majority of students were African-American; therefore, ethnicity was not used as an independent variable.

2. The small sample size for schools (n=10), student mobility rates, pre-assignment of student groups, and pre-established teacher assignments limit the ability to generalize results.

3. HT had been used since 2003-2004 in FCSS compared to VUL which was used since 2004-2005 with third-grade students. The difference in length of usage between the two reading programs may influence the results.

4. Implementation with fidelity and integrity was beyond the researcher's control and may influence findings. Using group data could not provide as valid an assessment as a design that would embed students into class units, recognizing the influence of teacher, peers, and setting (etc.) on student achievement.

5. Another limitation is the cross-sectional design because it becomes difficult to establish time order (condition 2 of the necessary conditions for causality).
If data are collected from research participants at one point only, there can be no direct measurable changes that occur over time (University Of South Alabama). This study had only 2 points for comparison.

6. No qualitative data other than the single interview with the VUL District Coach are available for this study as survey information was not allowed for use in independent studies not designed or conducted by the county assessment and evaluation office.

7. The researcher was an administrator in the district at the time of the study which suggests potential for researcher bias, in interpretation and discussion.

**Implications for Policies, Practice and Future Research**

1. Replicate district program evaluations using a quasi-experimental design and achievement pretests to establish equivalency of comparison group to intervention group at baseline. A district evaluation conducted in 2004 did not meet the What Works Clearinghouse (WWC) evidence screens because it lacked establishment of a baseline within the quasi-experimental design.

2. Most SWD showed substantially higher pre-test and post-test means in the VUL cohort when compared to the HT pre-test and post-test means. A similar study
should be designed to establish control and treatment special education cohorts over two or more years to determine effectiveness of the treatment.

3. Further studies should investigate the implementation effect on reading comprehension.

4. Voyager Universal Literacy Program (VUL) stops at third grade. Are gains in reading at grade 3 attributed to VUL sustained in grades 4 and 5?

5. Replicate this study or a similar study in schools with large English Language Learners (ELL) populations.

6. Compare VUL reading combined with Voyager Passport intervention for extended day to HT reading combined with Voyager Passport intervention.

7. Frechtling, Zhang and Silverstein (2006) studied kindergarten students by using the DIBELS Letter Naming Fluency subtest as one of the assessment instruments. A similar study should be designed to include DIBELS as one of the assessment instruments noting the extensive research conducted by Good, Kaminski, and Simmons (2001) related to the validity of DIBELS as a predictor of reading success.
8. A study should be conducted regarding the value of the tests used [e.g. DIBELS, Vital Indicators of Progress (VIP)].

Conclusions and/or Discussions

Cooley (1993) calculated that over 60% of the "variation in average school performance (test scores) among these school districts can be explained by these three simple census factors" of a child:

1. living in poverty,
2. with a single parent,
3. who is not a high-school graduate. (p. 5)

The conclusions drawn from this study validate the research and theories of several noted educators. Bloom (1984a, 1984b) discussed the "2-sigma problem" to find methods of group instruction that were as effective as one-to-one tutoring.

Achilles (1999) developed a formula for education improvement (EI): improved education = QE² (quality, equality and equity) using the class-size research as the theoretical framework. Small classes offer quality, equality, and equity. The idea of "individualized" instruction becomes a reality when class size is low and more manageable. If we objectively analyze the costs and benefits study of Levin, Belfield, Muennig, & Rouse (2007)
and revisions of Belfield & Levin (2007), we realize that policy changes must be forthcoming that address the social problems that contribute greatly to the school academic achievement gaps associated with minority students, SWD and low-SES. New results from a recent study published by Muennig and Woolf (2007) conclude that reducing class size may be more cost-effective than public health and medical interventions. (p.2020) The Muennig and Woolf (2007) study shows that estimated savings increase to $196,000 per additional low-income student who graduates.

The What Works Clearinghouse (WWC) listed two studies (Frechtling, Zhang, & Silverstein, 2006; Hecht, 2003) which investigated the effects of VUL and that met WWC evidence standards with reservations. The WWC considers the extent of evidence for VUL to be moderate to large for alphabettics and small for comprehension. No studies that met WWC evidence standards with or without reservations addressed fluency or general reading achievement (WWC, 2007). The extent of evidence for comprehension is small because the domain only included the study by Hecht who reported a negative effect on the Stanford Binet Intelligence Expressive Vocabulary subtest in that study. According to Slavin (2007):
the clearinghouse gives its highest ratings for evidence of positive effects to programs supported by studies that are often very small, very brief, very biased, and/or very seriously flawed in other ways, failing to give educators valid or meaningful information on the programs they might use to improve their students' achievement. (p.36)

General reading achievement has yet to be fully examined for VUL.

The 2005-06 school year was the first year for assessments on the new Georgia Performance Standards (GPS) aligned to the Georgia CRCT. In the present study, the researcher did use the 2005 CRCT scores as pre-test data and 2006 CRCT scores as post-test data. Given that VUL students scored well on the new CRCT aligned to GPS which suggests a strong alignment to the state test in the area of reading.

The Annual Measurable Objectives (AMO) for AYP calculations have increased from 66.7% to 73.3% students meeting or exceeding state standards in Reading/English/Language Arts. Thus educators, in spite of the federal appeals court ruling (City of Pontiac v. Department of Education, 2008) against the No Child Left Behind (NCLB) law; are working to ensure quality curricular
materials that can assist with the teaching of reading and language arts. VUL students, especially in the male and SWD subgroups were more successful in meeting state standards in reading using the VUL reading program than students using the HT program. However, the costs for the VUL program are more expensive than the costs for the HT program. The Sixth Circuit Court ruled that if Congress wants to control the direction of public education, it will have to follow the plain language of Section 9527 of NCLB and provide the funds needed to implement its mandates (City of Pontiac v. Department of Education, 2008).

Considering the diversity of schools in the district and the large disproportionality of students identified as SWD and as being African American males – the VUL reading program did significantly and positively influence the reading performances of those two subgroups (SWD, gender- [African-American males]) at least in the short term of the present study. The majority of students identified as SWD have deficits in reading (Lyon, 1996). Struggling readers benefit from small classes (Achilles, 2007; Achilles & Finn, 2006); Direct Instruction (Snow, Burns, & Griffin, 1998; Grossen, 1995) systematic phonics instruction, fluency and vocabulary instruction (NRP, 2000). The VUL reading program offers a more systematic program that
utilizes more of the specific strategies in the aforementioned areas than does the HT program.

Using Crane's (1998) criteria to analyze this study, (see Appendix H) the results reveal mostly positive answers that suggest the VUL program could be considered a favorable reading program to address the needs of selected populations.

A significant number of FCSS schools have large populations of low-SES, reading scores less than the system average, less experienced teachers, higher mobility rates and lower teacher retention rates. The VUL program is a more structured reading program which provides extension professional development for less seasoned teachers who may be less prepared to effectively address the reading instruction needs of struggling readers. The use of the DIBELS-like assessments and Direct Instruction (DI) combined with a balanced approach to reading and writing help provide mechanisms to assist schools meet varying student needs.

Future program evaluations are necessary to determine the sustainability of VUL and HT beyond third grade. Preliminary information gathered in this study suggests that using VUL in the primary grades would be beneficial to schools having high percentages of poverty or SWD. Making
AYP with the SWD subgroup is increasingly challenging, especially knowing that the Annual Measurable Objectives (AMO) will increase until all students and subgroups are 100% proficient in reading, language arts and math by school year 2013-2014. At a minimum FCSS should consider the influence that small class size, direct instruction, duration, intensity and $OE^2$ may have on student achievement, especially for SWD, African-American males and schools heavily impacted by low SES.
References


Achilles, C. M., & Finn, J. D. (2006). *Class-Size Policy: the STAR Experiment and Other Class-size studies (Policy Brief (Draft)).*


http://www.questia.com/PM.qst?a=o&amp;d=5008813963


http://epsllasu.edu/epru/ttreviews/EPSL-0702-225-EPRU.pdf


Bloom, B. S. (1984a). The search for methods of group instruction as effective as one-to-one tutoring. Educational Leadership, 41(8), 4-17.


California English Language Arts Committee (1987). *English Language arts framework for California’s public schools (kindergarten through grade twelve)*. Sacramento California Department of Education.


Frechtling, J. A., Zhang, X., & Silverstein, G. (2006). The Voyager Universal Literacy System: Results from a study of kindergarten students in inner-


Furr, D. (2001). Leave No Child Behind? We may be institutionalizing the very cycle of failure we seek to break. Education Week, 21(9), 34.


Georgia Department of Education. *Georgia's Criterion-Referenced Competency Tests (CRCT) - Questions and Answers for Parents of Georgia Students.* Retrieved May 13, 2006, from Georgia Department of Education Web Site:


Georgia Department of Education. (2002). *2002 State Recommended Textbook/Instructional Materials List.* Retrieved November 12, 2006, from Georgia Department of Education Web Site:


36A3C27&type=D


Harris, A., & Serwer, B. (1966a). Comparison of Reading Approaches in First-Grade Teaching with Disadvantaged Children (The Craft Project). ED 010 037.


http://www.hope.edu/academic/education/wessman/2block/unit3/history.htm


International Reading Association (or IRA). (2002). In What is evidence-based reading instruction? (Position statement of the IRA). Retrieved August 28, 2005, from IRA Web Site:
http://www.reading.org/resources/issues/position_evidence_based.html


http://nces.ed.gov/nationsreportcard


National Reading Panel [cited in text as "Subgroups"], (2000). Report of the National Reading Panel: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction, reports of the subgroups. Washington, DC: National Institute of Child Health and Human Development. 33pp [We have referenced pages from the version posted online as of Nov. 2002. These page numbers agree with those in one of the bound, undated versions].


of reading comprehension and reading."


The reauthorization of the Office of Educational research and Improvement: Hearing before the House Committee on Education and the Workforce, Subcommittee on Education Reform, 107th Congress, (2002, February 28) (testimony of Grover Whitehurst). Also available on-line:
http://edworkforce.house.gov/hearings/107th/edr/oeri22802/whitehursts.htm


Sweet, R. W. (2004). The Big Picture - Where We are Nationally on the Reading Front and How We Got Here. In P. McCrandle & V. Chhabra (Eds.), The Voice of Evidence in Reading Research (pp. 13-46). Baltimore, MD: Paul H. Brookes.
Texas Education Agency (1990). *Proclamation of the state Board of Education advertising for bids on textbooks.* Austin, TX: Author.


University Of South Alabama. Cross-sectional research. Retrieved February 15, 2007, from University of South Alabama Web Site:


Appendixes
Appendix A

Superintendent's Approval Letters

Test Coordinator's Approval of Research
June 5, 2006

As Superintendent of the Fulton County School System,

I, James Wilson, give Tawana D. Miller, a Doctoral student in Educational Administration at Seton H. University, permission to conduct the research as described in her letter dated June 5, 2006 according to the policies established by Seton Hall University Institutional Review Board for Human Subjects Research and with the parameters of the local Board of Education policy.

James Wilson  
Signature  
6-5-06  
Date
June 5, 2006

Mr. James Wilson
Superintendent
Fulton County Schools
Atlanta, Georgia 30315

Dear Mr. Wilson,

I am a doctoral candidate in Educational Administration at Seton Hall University and am presently working on my dissertation. The purpose of my dissertation is to study the effect of the Voyager Universal Literacy Program as compared to the Harcourt Trophies Language Arts Reading program on reading achievement of third grade students who attend Parklane ES, Liberty Point ES, Seaborn Lee ES, Brookview ES, S. L. Lewis ES, Heritage ES, Oak Knoll ES, Conley Hills ES, Campbell ES and Mount Olive ES.

The letter requests permission to use the Criterion Referenced Competency Tests, The Iowa Tests of Basic Skills, Checkpoints and other summative or formative assessments for the 2005-2006 third grade students for this study. Historical test data for these same students from kindergarten to third grade will be examined in this study. No individual or school district will be identified at any time before, during, or after the study, and all data will be confidential. I will be assigning numbers to each student in order to examine their historical progress and the key (my coding system) will be destroyed as soon as the data is calculated. The proposal for the study has been reviewed with by Ms. Connie Maggett, Director – Assessment and Evaluation and approval to conduct the study has been granted.

Please indicate your approval to grant me permission by endorsing the enclosed Letter of Permission and

inserting it in the enclosed, self-addressed, envelope.

Thank you for your consideration and assistance of this request. I would be happy to provide the district a copy of the results of this study, if you are interested.

Sincerely,

Tawana D. Miller

Tawana D. Miller
Appendix A (continued)

Test Coordinator's Approval of Research
April 12, 2006

Fawore D. Miller
Executive Director
Curriculum and Instruction
786 Cleveland Ave.
Atlanta, GA 30315

Dear Ms. Miller:

Your request to conduct the research study [A Comparison of the Influence of Two Reading Programs (Voyager Universal Literacy & Harcourt Brace) on Third Grade Reading Achievement Including Cost Benefit Analysis] in the Fulton County Schools has been reviewed. We are pleased to inform you that you have been granted permission for this study. Approval means that a school may choose to participate in this study; however, it is not mandatory that they do so, as the choice remains a local option.

No identification of Fulton County Schools (students' names, teachers' names, etc.) is to be included in your findings. Also, all confidentiality of records must be maintained. Once this study is completed, please send to me at the address below a copy/summary of the completed study. If I can provide additional information, please contact my assistant, Jennifer P. Arnold, or me at (404) 763-5600 ext 123.

Sincerely,

[Signature]
Connie Maggart
Director of Assessment & Evaluation
Appendix B

IRB Non-Review Certification Letter
IRB non Review Certification

STUDENT: Tawana D. Miller

Title of Dissertation: A comparison of the influence of two reading programs on third grade reading achievement

I certify, by my signature below, that the above indicated study does not require IRB review as a result of a lack of involvement with human subjects (see OHRP flow chart) and as indicated by any or all of the following (check all that apply).

1. Historical research
2. Public data base
3. Proprietary data base
4. Freedom of Information
5. Right to know – sunshine law

Student signature: ____________________________
Advisor approval: ____________________________

Reviewed by: ____________________________________________
Marty Finklestein – Higher Ed  Daniel Gulmore – K-12

• Proprietary data that do not identify individuals
Appendix C

Glossary

Achievement gap is usually defined as the difference between the test scores of black and Hispanic students on the other (Bracey, 2005).

Alphabetic principle is “knowledge of letter names.” (Bond & Dykstra, 1997). Snow, Burns, and Griffin (1998) define the alphabetic principle as the idea that written spellings systematically represent spoken words. Alphabetic principle includes alphabetic understanding or the mapping of print to speech (Elliott, Lee, & Tollefson, 2001).

Cohort is a group with statistical similarities sharing a common factor such as the same age or the same income bracket, especially in a statistical survey (The Encarta Dictionary).

Criterion Referenced Competency Tests (CRCT) are state-mandated achievement tests for students in grades 1 through 8 that cover the subject areas of reading, English/language arts, and mathematics. Students in grades 3 through 8 are also required to take the CRCT in science and social studies in the state of Georgia. The CRCT is based on the RASCH model of measurement.
Curriculum-Based Measurement is one of the few alternative forms of assessment where an impressive body of data supports the technical adequacy and the practical application of these techniques in the assessment of young children (Elliott et al., 2001).

Direct Instruction refers to a rigorously developed, highly scripted method of teaching that is fast-paced and provides constant interaction between students and the teacher (Lindsey, 2004). Direct instruction is skills-oriented and its teaching practices are teacher-directed. It emphasizes the use of small-group, face-to-face instruction by teachers and aides using carefully articulated lessons in which cognitive skills are broken down into small units, sequenced deliberately, and taught explicitly (Carnine, 2000; Traub, 1999).

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

DIBELS are ten reliable and valid measures of
essential early literacy skills used to document growth
toward outcomes. DIBELS measures of language
development include: Story Retell, Picture Description,
and Picture Naming Fluency. DIBELS tasks designed to
measure knowledge of alphabetic print are: Letter
Naming Fluency (LNF) and Sound Naming Fluency (SNF).
DIBELS measures of phonological awareness include:
Rhyming Fluency, Blending Fluency, Onset Recognition
Fluency (OnRF), Initial Sound Fluency (ISF), and
Phonemic Segmentation Fluency (PSF).

Economically Disadvantaged (ED) is a term used by
government institutions to categorize "a student who is
a member of a household that meets the income
eligibility guidelines for free or reduced-price meals
(less than or equal to 185% of Federal Poverty
Guidelines. (Wikipedia, 2006). In this study, ED is used
interchangeably with low socioeconomic status (SES).

Fluency is smooth and natural oral production of written
text(Rasinski, 1989).
Guided reading is a teaching approach designed to help individual students learn how to process a variety of increasingly challenging texts with understanding and fluency. Guided reading occurs in a small-group context because the small group allows for interactions among readers that benefit them all. The teacher selects and introduces texts to readers, sometimes supports them while reading the text, engages the readers in a discussion, and makes teaching points after the reading. Sometimes, after reading a text, the teacher extends the meaning of the text through writing, text analysis, or another learning experience. The lesson may also include work with words based on the specific needs of the group (Fountas & Pinnell, 2001).
Iowa Test of Basic Skills (ITBS) are sets of norm-referenced standardized tests given annually to school students in the United States. These tests are given to students beginning in kindergarten and progressing until Grade 8 to assess educational development. The ITBS was designed by the University of Iowa's College of Education as part of a program to develop a series of nationally accepted standardized achievement tests (Wikipedia, 2006). The ITBS covers numerous areas of educational knowledge, including English language use, math concepts and problem solving, social studies, science, geography and map use, and source use.

Level 1 refers to a range of scores that define a specific level of performance on the CRCT. Level 1 performance scores range from 150 to 299 or (650 to 799) and indicate a level of performance that Does Not Meet the Standard set for the CRCT test.

Level 2 refers to a range of scores that define a specific level of performance on the CRCT. Level 2 performance scores range from 300 to 349 or (800 to 849) and indicate a level of performance that Meets the Standard set for the CRCT test.
Level 3 refers to a range of scores that define a specific level of performance on the CRCT. Level 3 performance scores range from 350 to 450 or (850 to 950) and indicate a level of performance that Exceeds the Standard set for the CRCT test.

Metaphonological skills refer to thoughts about the way sounds of the language operate (Snow, Burns, & Griffin, 1998).

Nonexperimental research is systematic empirical inquiry in which the scientist does not have direct control of independent variables because their manifestations have already occurred or because they are inherently not manipulable. Inferences about relations among variables are made, without direct intervention from concomitant variation of independent and dependent variables. (Kerlinger, 1986, p.348: italics in original)
Norm-referenced tests (NRTs) are designed to “rank-order” test takers – that is, to compare students’ scores. A commercial norm-referenced test does not compare all the students who take the test in a given year. Instead, test-makers select a sample from the target student population (say, ninth graders). The test is “normed” on this sample, which is supposed to fairly represent the entire target population (all ninth graders in the nation). Students’ scores are then reported in relation to the scores of this “norming” group (Fair Test, 2007).

Phonemic awareness is the insight that every spoken word can be conceived as a sequence of phonemes. Because phonemes are the units of sound that are represented by the letters of an alphabet, an awareness of phonemes is key to understanding the logic of the alphabetic principle and thus to the learnability of phonics and spelling. (Snow, Burns, & Griffin, 1998, p.52)
Phonemes are speech phonological units that make a
difference to meaning. Thus, the spoken word rope is
comprised of three phonemes: /r/, /o/, and /p/. It
differs by only one phoneme from each of the spoken
words soap, rode, and rip. (Snow, Burns, & Griffin,
1998, p.52)

Phonics refers to instructional practices that emphasize
how spellings are related to speech sounds in
systematic ways. (Snow, Burns, & Griffin, 1998, p.52)

Phonological awareness is the ability to manipulate and use
the sounds of language in the absence of print (Smith
et al., 2001).

Phonological awareness is a more inclusive term than
phonemic awareness and refers to the general ability
to attend to the sounds of language as distinct from
its meaning. Phonemic awareness generally develops
through other, less subtle levels of phonological
awareness. Noticing similarities between words in
their sounds, enjoying rhymes, counting syllables, and
so forth are indications of such "metaphonological"
skill. (Snow, Burns, & Griffin, 1998, p.52)
Phonological sensitivity is the identification and manipulation of sound structure of language (Lyon & Chhabra, 2004).

Reader-response theory posits that literature is not an object to be studied nor does it have one correct interpretation (Iser, 1978); rather, meaning in the text is constructed by the readers’ own interpretation of their experiences while they are reading (Rosenblatt, 1978).

Scientifically based reading instruction (SBRI) - This term was first defined in law in the Reading Excellence Act and was carefully written to reflect the manner by which valid research is conducted by the National Science Foundation, the National Institutes of Health, and the National Academy of Sciences (Sweet, 2004).

Socioeconomic Status (SES) is an individual's or group's position within a hierarchical social structure. Socioeconomic status depends on a combination of variables, including occupation, education, income, wealth, and place of residence. Sociologists often use socioeconomic status as a means of predicting behavior (Hirsch, Kett, & Trefil, 2002). SES is used to refer to the economically disadvantaged (ED).
Standard score describes a student’s performance along standard scale. These are scores that should be expected of students at that grade level.

Student progress monitoring is defined as an ongoing procedure that uses the Vital Indicators of progress to ensure informed and targeted instruction. Ongoing progress monitoring identifies struggling readers within the first few weeks of school and measures their performance weekly.

Vital Indicators of Progress (VIP) is a timed oral measure individually administered to Voyager students twice in the fall, once in winter, and once more in spring. VIP is brief, predictive, and highly reliable oral fluency measures of student progress every nine weeks - to ensure informed and targeted instruction.

Vocabulary is all the words used by or known to a particular person or group, or contained in a language as a whole.
Appendix D

Standards Setting, Cutscores and Scale Scores for the
Georgia CRCT

Georgia educators employed a modified Angoff methodology in
making their recommendations for standards setting and the
cutscore that was used for 2005-2006 CRCT. Appendix E
includes the Angoff methodology and cut scores for the
CRCT. In brief, this method involves the steps listed
below.

1. Panelists operationalize performance-level descriptors
or PLDs.

2. Panelists take the full test.

3. Panelists use PLDs to rate each item in terms of the
percent of students meeting or exceeding the standard
who should respond correctly.

4. Panelists study field-test data on items and
share/discuss the ratings

5. Panelists rate items a second time.

6. After viewing 'impact data', panelists make a final
recommendation (Georgia Department of Education,
2006).

After the first round of standard setting an
Articulation Committee is convened to review and evaluate
the standards recommendation in the initial process. Whereas the initial process involved setting standards independently within each grade and content area, the Articulation Committee considers standards collectively across all grades and content areas.

In evaluating the standards, the Articulation Committee considered such factors as 1) consistency across grades (including a review of statistical smoothing, 2) the expectations described in the Georgia Performance Standards and 3) relevant policy and/or accountability implications. The Articulation Committee submitted a final recommendation to the State School Superintendent, the Governor's Office of Student Achievement, and the State Board of Education (SBOE) for review and approval. The SBOE approved the recommended standards at the April board meeting 2006.

Since the CRCT measure performance and are criterion-referenced tests, questions would naturally be asked regarding the reliability and whether the instrument could yield consistent results over time. The Georgia BOE adopted the new Cutscores for the CRCT.

New cutscores were established in the grades and content areas based on the Georgia Performance Standards rollout. New scale scores have been developed for those areas which test the new GPS curriculum. The DOE set the
scale scores after reviewing a sampling of test results in these areas. The new scale has a range of 650-950. 800 will be the scale score required to meet expectations. The new cutscores for selected grades are indicated in Table 23.

Table 23
The cutscores for the 2006 CRCT for GPS-Based Tests in grades 1 through 4 are:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Reading</th>
<th></th>
<th>ELA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Meets</td>
<td>Exceeds</td>
<td>Meets</td>
<td>Exceeds</td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>37</td>
<td>28</td>
<td>46</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>37</td>
<td>30</td>
<td>46</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>35</td>
<td>25</td>
<td>43</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>35</td>
<td>27</td>
<td>33</td>
</tr>
</tbody>
</table>

Number correct out of 40
Number correct out of 50

Student performance standards for the CRCT are established through a standard setting process in which educators from around the state participate. Educators make recommendations on what scores define categories of student performance. After this process, student scores on the CRCT are reported in the scale scores and performance levels listed in Table 24.
Table 24

*Performance levels and Scale Scores for the CRCT*

<table>
<thead>
<tr>
<th>Level</th>
<th>Performance</th>
<th>Scale Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Does Not Meet Standard</td>
<td>650-799</td>
</tr>
<tr>
<td>2</td>
<td>Meets the Standard</td>
<td>800-849</td>
</tr>
<tr>
<td>3</td>
<td>Exceeds the Standard</td>
<td>850-950</td>
</tr>
</tbody>
</table>
Appendix E

Standard Scores for Iowa Tests of Basic Skills (ITBS)

A standard score describes a student's performance along a standard scale. These are scores that should be expected of students at that grade level.

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>K</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>130</td>
<td>150</td>
<td>168</td>
<td>185</td>
<td>200</td>
<td>214</td>
<td>227</td>
<td>239</td>
<td>250</td>
<td>280</td>
</tr>
</tbody>
</table>

Source: Georgia Department of Education Web Site:

http://public.doe.k12.ga.us/DMGetDocument/aspz/342,19,standardscore(SS)
Appendix F

Profile Plots for Reading Programs, Gender, SES [also referred to as ED] and Students with Disabilities (SWD)

Estimated Marginal Means of RDG_Scale Sc

[Diagram showing estimated marginal means for RDG_Scale Sc across different reading programs and gender]

Reading Prg
--- HT
--- VUL

Student Gender Desc
FEMALE
MALE
Estimated Marginal Means of RDG_Scale Sc

Reading Prg
--- HT
--- VUL

Estimated Marginal Means

ED

.00
1.00
Estimated Marginal Means of RDG_Scale Sc

Reading Prg
- - HT
- - VUL

Estimated Marginal Means

<table>
<thead>
<tr>
<th>Ind Special Ed This Year</th>
<th>SWD</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>

Values:
- 830
- 820
- 810
- 800
Appendix G

Interview with Voyager District Coach
Voyager District Coach Interview - Friday August 10, 2007. The following questions were asked during the interview.

Question 1. State your credentials as a Voyager District Coach.

Response 1: I have a Reading Specialist (Ed.S.) Degree from Georgia State University. I have been trained by Voyager in Universal Literacy and Passport and I have a leadership degree in Educational Leadership from Lincoln Memorial University.

Question 2: How long have you worked as a Voyager District Coach?

Response 2: This is my 6th year.

Question 3: Describe your duties and responsibilities as the Voyager District Coach.

Response 3: I'm responsible for obtaining all materials for classroom instruction, supporting implementation, contact person between FCSS and Support people at Voyager which includes materials and data analysis, all administrators in the schools as well as FCSS admin. Analyzing data and conducting meetings, I developed the Ongoing Professional Development (OPD) offerings and
facilitated the delivery of the OPDs to 3 of the 4 Voyager schools. I actually developed curriculum for 30 hours of professional development to be redelivered across 8 months. We’re not going to do it the month of April.

Go back to your Harcourt days and...

**Question 4:** Describe the strengths and weaknesses of the Harcourt Trophies (HT) Reading program. Specifically please refer to third grade but you can discuss HT wholistically.

**Response 4:** Harcourt provided a good variety of reading materials and different genres. They had a lot of support materials.

**Please comment on the professional development piece.**

Fulton County purchased training components that were developed by Harcourt. The ones that we did use were comprehension and vocabulary development. So Harcourt produced those and we were lucky that Fulton County had money to buy those and we as Lead Reading Teachers (LRT) went out and provided the staff development in those two genres. All of the LRTs redelivered that piece after the textbook adoption. It wasn’t the first year. It was just a one shot deal. The thing about that was - there was very good information but that didn’t come with the purchase of Harcourt materials. LRTs didn’t have the time to provide in
depth on-going support. Every delivery consisted of two schools coming together.

**Question 5:** Any other strengths with the Harcourt reading? What would you say about the structure of the lesson?

**Response 5:** I wouldn’t put that under strengths - that would go under weaknesses.

Separate if you can the HT piece and exclude the Harcourt Language program.

**Question 6:** What are the weaknesses of HT?

**Response 6:** Weaknesses - It was up to the teachers to figure out how to pull all of the materials together and decide what to teach, they had the anthology as the reading material but there’s more to reading than just going through the story and following what it says in the textbook. Based upon your memory you were an LRT at the time.

**Question 7:** Were CSTs trained or others trained, or any training or modules purchased to help the teachers?

**Response 7:** There was some training but it was more familiarity with the materials that came - it could have been but I don’t remember. They talked about the other components and what was embedded in the curriculum guides.
They had things in there like ELL support. There wasn’t as much support in implementing it [HT] as... [in VUL].

**Question 8:** What’s your overall assessment - drawing on all your years and expertise as a reading specialist?

**Response 8:** Overall assessment of the Harcourt Reading Program compared to everything else you’ve seen. I think it was decent. I guess it was adequate - good variety of story and genres -If I remember correctly there was narrative and expository, poems in there too. - Nice job of that. They really did provide a significant amount of resource materials to use as well.

**Question 9:** What is your overall assessment of Harcourt as a program to teach reading?

**Response 9:** With Harcourt just like with any other basal...if someone doesn’t know how to teach reading - Harcourt is not the way to do it - they gave you the materials but if someone doesn’t know how to use and how to develop the foundation the kids need- Harcourt cannot do it. A teacher couldn’t pick up this reading program and adequately teach reading? NO

We will follow the same process and interview questions for Voyager.

**Question 10:** Describe the strengths and weaknesses of VUL.
Response 10: Strengths - VUL is based on research and best practice. As far as developing the foundation in K and 1st grade - I haven’t seen anything better. Support from the company that is on-going - not just the first year. An online data system and an assessment program are in place. This is going to be our 6th year and they [Voyager Expanded Learning] are still trying to help us.

It could have been a marketing tool - but they continued to revise. They revised more the first and second year. They are getting close to their adoption so they haven’t done as much - gone to hardbound anthology. Oh I’m really reading a book. Always wanting feedback - wanting to know what works and what doesn’t work. With Peg Marin - she developed a mock test for 3rd grade that looked like the CRCT. It looks like what the teachers wanted.

Question 11: What are the weaknesses of VUL?

Response 11: Biggest weakness - not enough language and writing but we must remember they are reading. They need the writing component in there. They do have the writing component at third built into the curriculum. Grammar is embedded and it does use the writing process. Writing process is taught. We use Writing to Win at grades 1 and 2.

Question 12: What’s your overall assessment of VUL?
Response 12: I personally think it beats just using reading textbooks because it provides that structure to teach kids how to really read. Best practice is embedded in the curriculum. Comprehension strategies (Main idea, retelling, cause and effect) are embedded as well. The curriculum is structured around the five essential components: phonemic awareness, phonics, fluency, vocabulary and comprehension. HT has it as well but VUL has it built in comprehensively and also has structured language in there that is designed for the benefit of the students. The teacher should be modeling and using think-alouds. Not to say that it didn’t exist in Harcourt but it was more apparent in VUL. VUL is good at providing the structure to teachers who did not know how to teach reading.

The other weakness that VUL has is that - 60% of text is expository - but not enough different genres for students who will be tested. They don’t provide enough for our learners to be successful on the CRCT

Question 13: What is your overall assessment of VUL?

Response 13: Overall assessment - I think - VUL is an excellent tool for teachers to use to give students a foundation and beyond. You would say that it is more rigorous than HT. Absolutely!
Question 14: What is your assessment of the cost of VUL as compared to the cost of HT for classroom materials?

Response 14: Given the cost - VUL is much more expensive. HT has a wealth of materials - VUL has sufficient materials.

Question 15: Does the cost justify the results?

Response 15: I'm not saying that VUL is the only reason that the schools made AYP but I would venture to say they might not have made AYP had it not been for VUL. If you had to add extra costs for those schools - I believe that the costs would be more closely aligned. When you look at just student materials cost - you have to look at what else we get for the money. We have extended support. We don't get that with Harcourt - you hand them (HT) the check - they hand you the materials and it's "hasta la vista".
Appendix H

Critique of VUL/HT results using Crane's (1998) Criteria
Critique of VUL/HT results using Crane's (1998) Criteria*

Crane Criteria and Questions

<table>
<thead>
<tr>
<th>Crane Criteria and Questions</th>
<th>VUL/HT facts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Do the benefits outweigh the costs? <strong>YES</strong></td>
<td>The VUL might not have made AYP had it not been for the VUL program. The professional development component may increase teacher effectiveness across all curricular areas.</td>
</tr>
<tr>
<td>2. Does the program have a statistically significant effect on the treatment group? <strong>YES</strong></td>
<td>Yes. This significant difference was found in the baseline data as well as the following year and by others.</td>
</tr>
<tr>
<td>3. What is the magnitude of the program's effect?</td>
<td>Effect-size (ES) results were .21 in the analyses for the entire treatment group. Smaller effect sizes were noted for students with disabilities and low SES.</td>
</tr>
<tr>
<td>4. How long do the effects of the program last? <strong>TBD</strong></td>
<td>Future studies are needed to determine the long-term effects into MS and HS.</td>
</tr>
<tr>
<td>5. What is the relationship of the evaluator to the program? <strong>Independent</strong></td>
<td>The researcher was an independent reviewer and neither a teacher or principal at a VUL school.</td>
</tr>
<tr>
<td>6. Can the program and its results be replicated? <strong>YES</strong></td>
<td>YES. Reported gains in other program evaluations are similar.</td>
</tr>
<tr>
<td>7. Can the program maintain its effectiveness on a larger scale? <strong>TBD</strong></td>
<td>It has yet to be determined whether to expand the program to other ES schools having similar demographics and test scores.</td>
</tr>
</tbody>
</table>