Georgia State University
ScholarWorks @ Georgia State University

Counseling and Psychological Services Dissertations

Department of Counseling and Psychological Services

Spring 5-5-2012

Headsprout Early Reading for Students At Risk for Reading Failure

Donna D. Kreskey
Georgia State University

ACCEPTANCE

This dissertation, HEADSPROUT EARLY READING FOR STUDENTS AT RISK FOR READING FAILURE, by DONNA DEVAUGHN KRESKEY, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree Doctor of Philosophy in the College of Education

The Dissertation Advisory Committee and the student's Department Chair, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty. The Dean of the College of Education concurs.

Stephen D. Truscott, Psy.D.
Committee Chair

Gary Bingham, Ph.D.
Committee Member

Catherine Cadenhead, Ph.D.
Committee Member

Andrea Dixon, Ph.D.
Committee Member

Date

Brian Dew, Ph.D.,
Chair, Department of Counseling and Psychological Services

R. W. Kamphaus, Ph.D.
Dean and Distinguished Research Professor
College of Education

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education's director of graduate studies and research, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

____

Donna DeVaughn Kreskey

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Donna DeVaughn Kreskey
200 Ashley Creek Drive
Newnan, Georgia 30263

The director of this dissertation is:

Dr. Steven D. Truscott, Psy.D.
Department of Counseling and Psychological Services
College of Education
Georgia State University
Atlanta, GA 30303 - 3083

VITA

Donna DeVaughn Kreskey

ADDRESS:			200 Ashley Creek Drive Newnan, GA 30263
EDUCATION:			
	Ph.D.	2012	Georgia State University School Psychology
	Ed.S.	2000	Georgia State University School Psychology
	M.Ed.	1998	Georgia State University School Psychology
	B.S.	1990	The University of Alabama Education (English & Psychology)

PROFESSIONAL EXPERIENCE:

	2000 - Present	School Psychologist Coweta County Schools, Newnan, GA
	1990 - 1999	Teacher Henry County Schools, McDonough, GA

PROFESSIONAL AFFILIATIONS

American Psychological Association
Georgia Association of School Psychologists
National Association of School Psychologists

PUBLICATIONS AND PRESENTATIONS:

Truscott, S.D., Kreskey, D., Bolling, M., Psimas, L., Graybill, E. Albritton, K., & Schwartz, A. (in press). Creating consultee change: Theory-based learning and behavioral change processes in school-based consultation. Invited manuscript for special issue on school-based consultation in *Consulting Psychology Journal: Practice and Research* (S. Rosenfield, guest editor).

Quick, L. M., Driscoll, K., & Kreskey, D. (in preparation). Does a three week self-regulation intervention improve math fluency?

Kreskey, D. & Truscott, S. D. (2010). Kindergarten. In C. Clauss-Ehlers (Ed.) *Encyclopedia of Cross-cultural School Psychology* (pp. 560-561). New York: Springer.

Kreskey, D. (2009). *Best Practices in Behavior Intervention Planning for Children with Attention Deficit Hyperactivity Disorder (ADHD): A Consultation Case Example including Data Collection and Graphing.* Presented to the staff of the Department of Psychological and Social Services, Dallas Independent School District, Dallas, TX.

Kreskey, D. (2009, October). [Commentary on "Crisis intervention for children/ caregivers exposed to intimate partner violence by C. Blodgett, K. Behan, M. Erp, R. Harrington, & K. Souers, (2008)]. *Communique*, pg. 12.

Psimas, L., Bolling, M. A., Graybill, E., Schwartz, A., Albritton, K., Kreskey, D., Truscott, S. D., & Stringer, D. (2009, February). *Designing and conducting exceptional professional learning to create lasting change.* Mini-skills presented at the National Association of School Psychologists 2009 Annual Convention, Boston, MA.

Graybill, E., Kreskey, D., Boney, T., Psimas, L., Bolling, M., Freer, P., & Truscott, S. (2008, February). *Professional development that builds capacity for behavioral interventions*. Paper presented at the National Association of School Psychologists 2008 Annual Convention, New Orleans, LA.

Kreskey, D., Brooks, J. H. & Carroll, M. S. (2008, February). *Administrators' and teachers' perceptions of pre-referral intervention teams*. Poster session presented at the National Association of School Psychologists 2008 Annual Convention, New Orleans, LA.

Truscott, S., Kreskey, D., Graybill, E., Psimas, L., & Bolling, M. (2007). *Professional Development that Builds Capacity for Behavioral Interventions.* Presented to representatives of Marietta City Schools, Marietta, Georgia.

Henderson, J., Kreskey, D., Hargrove, J., & Hannon-White, F. (2006, October). *Disproportionality, SST, and RTI: Coweta's Response to Closing the Achievement Gap*. Paper presented at the 2006 Georgia Association of School Psychologists Fall Conference, Savannah, GA.

Oster, R. & Kreskey, D. (2006). *Vocabulary Infusion.* Presented to the faculty of Yeager Middle School, Douglas County School System, and Douglasville, GA.

Kreskey, D. (2005) *Academic Data Collection for Universal Screening and Progress Monitoring*. Presented to the Department of Psychological Services and the faculty of Canongate Elementary School, Coweta County School System, Newnan, Georgia.

ABSTRACT

HEADSPROUT EARLY READING FOR STUDENTS
AT RISK FOR READING FAILURE
by
Donna DeVaughn Kreskey

This study examined the efficacy of using Headsprout Early Reading (Headsprout, 2007) to supplement a balanced literacy curriculum for kindergarten and first grade students in a suburban public school system. Headsprout, which is an example of computer aided instruction (CAI), provided internet-based, supplemental reading instruction that incorporates the five critical components of reading instruction cited by the National Reading Panel (NRP, 2000). The school system implemented Headsprout as a standard protocol, Tier 2 intervention within their Response to Intervention (RTI) process. The study included kindergarten and first grade students from across the school system who were identified as at risk for reading failure based on fall Dynamic Indicators of Basic Early Literacy (DIBELS) scores. Kindergarten and first grade students identified as at risk for reading failure who participated in Headsprout were compared with matched groups of kindergarten and first grade students who did not participate in Headsprout. Overall, neither kindergarten nor first grade students who participated in Headsprout gained meaningful educational benefit from the CAI instruction provided by Headsprout beyond the benefit they received from participating in the general education, RTI Tier 1, balanced literacy curriculum that was available to all kindergarten and first grade students.

HEADSPROUT EARLY READING FOR STUDENTS
AT RISK FOR READING FAILURE
by
Donna DeVaughn Kreskey

A Dissertation

Presented in Partial Fulfillment of Requirements for the
Degree of
Doctor of Philosophy
in
School Psychology
in
the Department of Counseling and Psychological Services
in
the College of Education
Georgia State University

Atlanta, GA
2012

ACKNOWLEDGEMENTS

There are many people I wish to thank for encouraging me along this journey. Foremost is my dissertation chair, Dr. Stephen Truscott. Steve, thank you for adopting me when my previous advisor left to take another job. You have consistently believed in my ability to make a significant contribution to our profession, and provided wonderful opportunities for me to learn new skills and to develop skills that I did not know that I had. For your guidance, mentoring, and friendship, I will forever be grateful. I also thank my committee members, Dr. Catherine Cadenhead, Dr. Gary Bingham, Dr. Andrew Roach, and Dr. Andrea Dixon for your expertise and encouragement throughout the dissertation process. Thanks as well, to the professors at GSU, I am a better school psychologist and child advocate now than I was when I began the program. I celebrate your vision of all the things that our field can be.

I am also grateful that I had the opportunity to share my doctoral studies with my GSU cohort, Janell, Amanda, and Amy; the GLRS research team, Michelle, Emily, and Lynnae; and special GSU friends, Jessica, Kathryn, Lisa, and Sam. I am lucky to have shared my doctoral program experiences with each of you. I am also fortunate to have had the support of a wonderful supervisor in my work as a school psychologist, Maria Carroll. Without her support of my project, this dissertation would not have been possible.

Finally, I thank my family. To my husband, David, you have been my heart for the past 18 years. I appreciate your patience with my addiction to school more than I can say. I could not have completed this degree without your loving support, your encouragement, and most importantly, your intuitive understanding of when it was past time for me to take a break. Thank you for sharing your life and our children with me. To my children, Ashley and Andrew, thank you for allowing me to be your stepmother. Watching you grow into the adults you are has been amazing. I appreciate your unwavering support through this process—I know you wondered if I was ever going to be done—so did I. And Mom and Dad, there are so many things for which I thank you, but here, as I finish this dissertation, I thank you most for showing me the magic of reading.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEFINITION OF TERMS

Alphabetic Principle - The alphabetic principle is the understanding that there are systematic and predictable relationships between written letters and spoken sounds (National Reading Panel [NRP], 2000).

Analytic Phonics - Phonics instruction that begins with a whole word. Students analyze parts of the word to understand how letters combine to form words (Pressley, 2005).

Balanced Literacy Instruction – Literacy instruction that includes an abundance of authentic reading and writing activities that occur within literacy rich classroom environments that incorporate high-quality literacy centers, the use of authentic children's literature, and opportunities for social collaboration among students (NRP, 2000)

Computer Aided Instruction (CAI) - Instruction that uses a computer to teach the student. The computer holds instruction, and the instruction is designed to teach, guide, and test the student until proficiency is attained (Association for Educational Communications and Technology, 1977).

Comprehension - The process of making meaning from text at both the literal and inferential level (Moats, 2004).

Criterion Referenced Competency Test (CRCT) - The CRCT assessment program is the designated assessment tool for federal accountability in Georgia for grades one through eight. The CRCT assessments and their associated reports provide information about academic achievement at the student, class, school, system, and state levels (Georgia Department of Education, 2009).

Drill and Practice Program – Computer programs that provide the learner with exercises and immediate feedback to reinforce specific skills (Desrochers & Gentry, 2004; Sivin-Kachala & Bialo, 1998; Soe et al., 2000).

Dynamic Indicators of Basic Early Literacy Skills (DIBELS) - DIBELS are a set of brief, individually administered pre-reading and early reading tests designed to assess phonological awareness, alphabetic awareness, and fluency in children. Benchmark assessments are administered to students three times each year, and the results of these screenings are used to evaluate students' progress toward the mastery of early literacy skills. Additional assessments are available to monitor the progress of at risk students at regular intervals (Good & Kaminski, 2007).

Explicit Instruction - Instruction that sequentially reviews previous work, presents new material, provides guided practice, provides feedback and correction, provides independent practice, and provides weekly and monthly reviews (Rupley, Blair, & Nichols, 2009).

Fluency - Reading fluency is the ability to accurately and quickly decode text, so that comprehension can occur (Kamil, 2005).

Graphemes – The written symbols that represent each phoneme (Ritchey & Goeke, 2006).

Initial Sound Fluency (ISF) - ISF assesses the child's ability to recognize and produce the beginning sound in an orally presented word (Kaminski & Good, 1996, 1998).

Integrated Learning System (ILS) – Computerized instructional programs that provide sequential instruction to students across several grades, while keeping records of student progress (Kulik, 2003).

Intensive instruction - Small group or individual instruction that is targeted to the needs of individual students, occurs at least several times a week, lasts for sessions of at least 15 minutes, and continues over several months (Cavanaugh, Kim, Wanzek & Vaughn, 2004).

Nonsense Word Fluency (NWF) - NWF measures knowledge of letter sound correspondences and the ability to blend letters into words (Kaminski & Good, 1996).

Oral Reading Fluency (ORF) - ORF is a standardized, individually administered measure of a student's accuracy and fluency when reading connected text (Kaminski & Good, 2006).

Phoneme –The smallest units of sound in a language (Ritchey & Goeke, 2006).

Phonemic Awareness - The understanding that words are made up of smaller sound-parts that can be manipulated to form other words (Scarborough & Brady, 2002).

Phonics - Instruction is about letter-sound correspondences (NRP, 2000).

Problem Solving Model RTI – An RTI process through which practitioners determine the magnitude of a student's problem, analyze its cause, design a goal-directed intervention, conduct it as planned, monitor the student's progress, modify the intervention as needed based on student responsiveness, evaluate its effectiveness, and determine future actions (Telzrow, McNamara, and Hollinger, 2000).

Response to Intervention (RTI) - A multi-tiered, general education process through which schools can monitor the progress of their students within the curriculum in order to identify students who are at risk for academic failure and provide them with interventions in a timely manner (Torgesen, 2009).

Sight Words – Words that a reader recognizes immediately without having to decode its parts (Ehri, 2005).

Synthetic Phonics – Instruction that centers on the 44 phonemes of the English language and their related graphemes (Ritchey & Goeke, 2006).

Standard Protocol RTI – RTI in which all students in Tier 2 at a given grade level with the same instructional deficit receive the same intervention. The intervention is time-limited, and it is typically implemented with students in a small group setting (Fuchs & Fuchs, 2009).

Universal Screening Measures - Brief screening tools that demonstrate diagnostic utility (e. g., DIBELS, Good & Kaminski, 2007; AIMSWeb, Shinn & Shinn, 2002; STEEP, Witt, 2002) for predicting performance on the reading and math state assessments in the elementary grades or on the local graduation requirements at the secondary level (McCook, 2006).

Vocabulary - Knowledge of the meanings and pronunciations of words that are used in both oral and written language (Kamil, 2005).

CHAPTER 1

INTRODUCTION

Literacy is a skill of universal importance. The National Assessment of Adult Literacy (National Center for Educational Statistics [NCES], 2005) defines literacy as "using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential" (p. 2). Individuals with poor literacy skills in the United States are at risk for a multitude of debilitating problems (Lyon, 1997; NCES, 2005). Adults who do not read well are likely to be unemployed due to their inability to function in a majority of employment situations (NCES, 2005), and research suggests that up to 75% of unemployed adults are unable to read (Lyon, 1997). Individuals with poor literacy skills comprise at least 60% of the prisoners and 85% of the juveniles who appear in court (Hodgkinson, 1991), and only 2 to 3% of prison inmates read at a proficient level (NCES, 2003). Unfortunately, as many as 50% of all children will have difficulty learning to read, and of these children, only half will ever become proficient readers (Lyon, 1997).

Converging research documents that patterns of reading achievement are established early, and once established they are difficult to change (e.g., Good, Simmons, & Kame'enui, 2001). Juel (1998) found that 88% of children who were poor readers at the beginning of first grade continued to be poor readers in fourth grade. In contrast, the likelihood of a student who was at least an average reader in first grade becoming a poor reader by fourth grade was only 12%. This occurs because the gap between poor readers and good readers widens across the elementary school years (Pressley, 2005; Sweet, 2004), particularly after third grade (NCES, 2003). These findings demonstrate that early intervention is essential to effectively address the nation's literacy problems.

In an attempt to change current trends in reading acquisition, President George W. Bush signed the No Child Left Behind Act (No Child Left Behind [NCLB], 2003) into law in 2002. This law mandates that every child read at grade level by the end of third grade. NCLB embodies the four principles of President George W. Bush's education reform plan: 1) stronger accountability for states, school districts and schools regarding student outcomes; 2) expanded flexibility and local control in the use of federal education dollars; 3) expanded options for parents and students, particularly those attending low-performing schools; and 4) emphasis on teaching methods that have been proven to work. The requirement for programs to be funded under NCLB is evidence of effectiveness, which was defined as "programs that have been found through scientifically based research to significantly improve the academic achievement of participating children or have strong evidence that they will achieve this result" (U.S. Department of Education, 2002, p. 45).

The Reading First (Elementary and Secondary Education Act, as amended, Title I, Part B, Subpart 1, 2002) initiative significantly increased Federal investment in scientifically based reading instruction programs for kindergarten through third grade students. NCLB requires that professional development, instructional programs, and materials used by states or school districts must include the five key areas of reading instruction that research has identified as essential. These areas of instruction are phonemic awareness, phonics, vocabulary, fluency and reading comprehension (National Reading Panel [NRP], 2000). Reading First requires evidence-based reading instruction as the core curriculum for all students. It also requires explicit, intensive, and supportive instruction for students who experience difficulty learning to read. Overall, the Reading First initiative emphasizes early identification and the need to screen all children for reading difficulties in order to intervene as early as possible.

While Reading First (2002) significantly increased federal investment in scientifically based reading instruction and intervention programs for children, it does not appear to have resulted in substantial progress toward the goal of universal literacy by third grade (Reading First Impact Study Final Report, 2008). Reading achievement, measured at the fourth grade, has increased for some minority groups and for children as a whole, but a significant percentage of children continue to read below basic proficiency. Specifically, only 48 % of black students, 48% of Hispanic students, and 78% of white students were able to read with at least basic proficiency on the most recent National Assessment of Educational Progress (NAEP; NCES, 2009). Basic proficiency in reading is defined as the ability to make relatively obvious connections between a text and one's own experiences and extend the ideas in the text by making simple inferences when reading text appropriate for fourth-graders (NCES, 2009). Additionally, only 51% of students who were eligible for free and reduced lunch were able to read at the basic proficiency level (NCES, 2009).

Despite the mixed results of recent early literacy initiatives, evidence from research settings continues to demonstrate the potential effectiveness of early intervention for children with early literacy problems (Cooke, Kretlow & Helf, 2010; McIntyre et al., 2005; Wanzek & Vaughn, 2007). For example, Vellutino and his colleagues (1996) found that with remedial help in the form of one on one tutoring for thirty minutes a day over the course of 15 weeks, only 1.5 to 3 percent of poor kindergarten aged readers in their study continued to experience significant reading difficulties. A solid foundation of early reading research and theory exists (NRP, 2000; Pressley, 2005; Tracey & Morrow, 2006; Vellutino, Fletcher, Snowling & Scanlon, 2004) along with a growing body of evidence-based interventions that can serve as a foundation for prevention and remediation (Kamil, 2008). An important avenue for future research is how to translate early intervention that has been demonstrated to be successful in research settings into the public education system without diluting its effectiveness.

Response to Intervention

Prevention and early intervention have long been goals of educators for struggling students. Response to Intervention (RTI) is a process that is intended to systematically addresses both of these goals (Fuchs, Mock, Morgan, & Young, 2003; Glover & DiPerna, 2007). The construct of RTI is not new. It originated in research-based practices such as mastery learning and curriculum-based assessment in the fields of special education and reading intervention, as well as models of consultation within the field of school psychology (Fuchs, 2004). RTI is a multi-tiered, primary prevention approach to education that employs evidence-based instruction and intervention, paired with continuous monitoring of student progress, to either prevent student learning difficulties or intervene to remediate them as early as possible (Reschly, 2005; Torgesen, 2009). Further support for RTI processes was provided by the reauthorization and revision of the Individuals with Disabilities Act (IDEA) in 2004. IDEA now requires the use of evidence-based instruction within the general education program prior to a student's referral for special education eligibility determination. Consequently, increasing numbers of school districts are implementing RTI models to address a variety of student learning and behavior needs (Response to Intervention Adoption Survey, 2010). Effective RTI processes attempt to ensure that all students have equal educational opportunity and prevent long periods of academic failure. Results from schools that have implemented RTI for a number of years suggest that RTI can be an efficient method for identifying and intervening with student difficulties in reading, mathematics, and classroom behavior (Burns, Appleton, & Stehouwer, 2005; Fuchs, Compton, Fuchs, Bryant & Davis, 2008; VanDerHeyden, Witt, & Gilberson, 2007).

The literature on RTI broadly identifies two models for intervention delivery: individualized interventions developed through a consultation based problem solving model and standard protocol interventions (Fuchs et al., 2003). According to Telzrow, McNamara, and Hollinger (2000), the problem solving model was first described in the behavioral consultation literature. As described by Bergan and Kratochwill (1990), behavioral consultation is both inductive and empirical. Solutions for instructional and behavioral difficulties are determined by the frequent and systematic evaluation of a student's response to an implemented intervention. It is a process through which practitioners determine the magnitude of the problem, analyze its cause, design a goal-directed intervention, conduct it as planned, monitor student progress, modify the intervention as needed based on student responsiveness, evaluate its effectiveness, and determine future actions (Telzrow, McNamara, & Hollinger, 2000). Collaborative consultation models, such as the instructional consultation teams approach (Rosenfield & Gravois, 1996), build upon the behavioral consultation model by intentionally emphasizing the development of a school culture that values professional collaboration across disciplines and purposefully enhancing teachers' skills in and application of best practices of instructional assessment and delivery.

Standard protocol interventions, in contrast, provide every child who is identified at risk in a specific area with the same evidence-based intervention. The standard protocol approach is advocated by some early reading researchers (e.g., Fuchs et al., 2003; Fuchs & Fuchs, 2009). Examples of evidence-based reading interventions implemented within the standard protocol approach include alternative curricula such as *Reading Mastery Classic* (McGraw-Hill, 2003) and *Language for Learning* (McGraw-Hill, 2005). These curricula provide instruction that includes the areas of phonemic awareness, phonics, reading fluency, vocabulary, and comprehension. They are characterized by scripted instruction, increased repetition of skills, and a slower pace of instruction. Standard protocol interventions may increase treatment integrity by simplifying training of school personnel and assessment of the accuracy of implementation. Additionally, grouping children with similar problems may increase the number of children who receive evidence-based intervention (Fuchs, et al., 2003).

RTI is conceptualized as a multi-tier model. The first tier is comprised of the school system's standard curriculum and instruction within the general education program. This tier is viewed as primary prevention, and instruction at this level is generally expected to meet the educational needs of 80% of a given school's student population (McCook, 2006). Tier 1 is consistent with typical classroom instruction supplemented by classroom adaptations that require minimal resources to implement. The classroom teacher provides instruction and support at the individual and group level. The defining characteristic of Tier 1 instruction is that resources directed at an identified concern do not exceed what is typically used for classroom instruction. In Tier 1, general education assumes responsibility for universally screening all students at least three times each school year. Within the first month of school, all students are screened to identify those at risk for school failure using brief screening tools that demonstrate diagnostic utility (e. g., AIMSWeb, Shinn & Shinn, 2002; Dynamic Indicators of Basic Early Literacy Skills [DIBELS], Good & Kaminski, 2002; System to Enhance Educational Performance [STEEP], Witt, 2002) for predicting performance on the reading and math state assessments in the elementary grades or on the local graduation requirements at the secondary level (McCook, 2006). School personnel meet regularly to review the screening data, and measurable goals are set for the entire group of students for the next check point. In Tier 1, the focus is on making large-scale changes to the instruction for entire groups of students.

The second tier of RTI is reserved for those students who are not responding to general education instruction. Tier 2 is considered secondary prevention and should include no more than 20% of the general education population (Burns et al., 2005). Through universal screening, students categorized as at risk are identified, and at Tier 2, school personnel identify instructional changes designed to address the educational needs of these students. These instructional changes may include evidence-based standard protocol interventions, the use of problem solving teams to develop more individualized, evidence-based interventions for specific students, or both. The students in Tier 2 receive this instructional treatment in addition to Tier 1 core instruction. Teachers and other school personnel implement the interventions with fidelity, and students who have been identified as at risk are assessed weekly in the area of risk using brief assessments to monitor their progress toward grade level expectations. Adequate progress toward grade level expectations is operationalized using local or national normative data. For students at risk for reading failure, Tier 2 serves two purposes: to prevent reading difficulty by delivering a more intensive, and presumably effective, level of intervention that accelerates reading development, and to assess the child's responsiveness to instructional intensity from which the vast majority of children should profit (Vaughn & Fuchs, 2003).

Students who do not make adequate progress toward grade level expectations at Tier 2 may be moved to Tier 3. In many RTI models, Tier 3 is synonymous with special education services because the students served in Tier 3 have failed to make adequate progress in the general education program (Tier 1 and Tier 2) despite the implementation of increasingly intensive evidence-based interventions, and the presence of a disability has been confirmed through comprehensive evaluation (Fuchs & Fuchs, 2009). Tier 3 is tertiary prevention. Teachers develop ambitious, individualized goals for each student who receives Tier 3 services. These goals may or may not be grade-appropriate. Thus, in Tier 3, students who are at risk for reading failure may receive instruction in foundational skills that are below grade-level but necessary for successful reading achievement (Fuchs & Fuchs, 2009). As in Tier 2, each student progress is monitored weekly; however, the student's progress is compared to his or her individual goals rather than to grade level expectations (Fuchs & Fuchs, 2009).

Although few researchers have taken these interventions to scale, increasing numbers of school systems are implementing RTI processes. Significant barriers to the effective implementation of RTI processes are prevalent in public school settings. These barriers include insufficient teacher training, lack of intervention resources, and lack of resources for instruction and monitoring student progress (Response to Intervention Adoption Survey, 2010). Because school systems consistently report lack of teacher training and lack of intervention resources as barriers to the implementation of RTI, it is important to ensure that interventions that prove effective in research settings are further evaluated within school systems that adopt them.

Computer Aided Instruction (CAI)

According to the Response to Intervention Adoption Survey (2010), the lack of intervention resources, such as personnel to implement interventions, is one of the greatest barriers to the implementation of RTI. Some professionals in the field of education believe that CAI has the potential to reduce educational costs while enhancing educational effects (Kulik & Kulik, 1991; Sivin-Kachala & Bialo, 1998). CAI may provide an avenue to address many of the perceived difficulties of both problem solving and standard protocol models of RTI because it is intended to provide individualized, engaging teaching based on a standardized curriculum without the need for large amounts of extra work from teachers or additional school personnel. Well-designed CAI can also provide immediate feedback, provide frequent opportunities to respond, and create high rates of success by either allowing students control over the instructional sequence or adapting the presentation of content based on student responses (Sivin-Kachala & Bialo, 1998). CAI, therefore, has the potential to provide important instructional opportunities to help children of all ages who struggle to learn to read acquire critical early reading skills such as phonics, word recognition, and word meaning (Hall, Hughes, & Filbert, 2000; MacArthur, Ferretti, Okolo, & Cavalier, 2001; Soe, Koki, & Chang, 2000).

Early meta-analytic reviews of CAI found generally positive effects for CAI on overall student achievement and on reading achievement in particular (Soe, Koki, & Chang, 2000) when it is used as a supplement to traditional teaching across grade levels and subject areas, and on student attitudes toward learning (Kulik & Kulik, 1987; Sivin-Kachala & Bialo, 1998). CAI also reduced the time required for students to learn new information (Hasselbring, 2001; Kulik & Kulik, 1987). Positive outcomes for CAI were also reported specifically in the area of reading instruction based on meta-analytic reviews of a variety of different types of CAI (Blok, Oostdam, Otter, & Overmaat 2002; Soe et al., 2000). In another meta-analytic review, MacArthur and his colleagues (2001) concluded that CAI can be used to teach phonemic awareness and decoding skills. In their review of the literature on CAI for students with reading disabilities, Hall, Hughes, and Filbert (2000) concluded that CAI had positive effects on student achievement in reading decoding and reading comprehension. They also suggested that carefully designed CAI can provide the systematic instructional procedures found to be effective for reading instruction.

In contrast, other researchers have reported neutral or unfavorable outcomes for CAI in the area of reading. First, the authors of meta-analytic reviews of the CAI literature consistently reported that firm conclusions regarding the effectiveness of CAI for reading cannot be drawn because there is a scarcity of acceptable studies in the important sub-skills related to reading achievement (Hall et al., 2000; Kulik 2003; Kulik & Kulik, 1987; MacArthur et al. 2000, Soe et al., 2000; Sivin-Kachala & Bialo, 1998). Evidence that students generalized the skills learned during CAI to other reading tasks was inconclusive (MacArthur et al., 2000). In the area of reading comprehension, in particular, CAI results were inconclusive. Some studies documented student gains in reading comprehension over and above the gains made by students receiving only traditional instruction, but others did not (Cheung & Slavin, 2011; Kulik, 2003). Despite the generally positive findings for reading decoding skills reported above, further research in the specific area of reading CAI is needed to determine whether or not CAI that addresses "the big five" in reading (i.e. phonemic awareness, phonics, vocabulary, fluency, and comprehension) is effective in improving the reading skills of students who are at risk for reading failure.

Headsprout Early Reading. Headsprout is one of several currently available CAI programs designed to provide supplementary reading instruction beyond the standard curriculum adopted by a school system. Headsprout is an internet-based, supplemental reading program designed for students in kindergarten through second grade who are not yet reading or who are at the beginning stages of learning to read. The program incorporates the five critical components of reading instruction cited by the NRP (2000). Headsprout includes numerous instructional strategies in the areas of phonemic awareness, phonics, fluency, vocabulary, and comprehension. Both the publishers of Headsprout (Layng, Twyman, & Strikeleather, 2003) and the Florida Center for Reading Research (FCRR, 2003) identify Headsprout as CAI with content and design that reflect scientific research. In RTI language, Headsprout is an example of a standard protocol, Tier 2 intervention for students at risk for reading failure that can be provided to students via a computer with internet access.

Initial research data suggest that most children who work with the Headsprout program as recommended by the publisher acquire the specific skills it is designed to teach (Clarfield, 2006; FCRR, 2003; Layng et al., 2003). Evidence that Headsprout produces reading gains over and above the gains that result from general classroom instruction, however, is inconclusive. Quasi-experimental research published by the developers of Headsprout (Layng et al., 2003; Layng, Strikeleather, & Twyman, 2004a; Layng, Strikeleather, & Twyman, 2004b) suggests that kindergarten, and first grade students who use Headsprout as a supplement to their regular reading instruction make significantly more progress in reading than those who do not use the program even when the amount of time spent in reading instructional activities is held constant. Using a multiple baseline design, Clarfield and Stoner (2005) found that three kindergarten and first grade students with ADHD made greater gains in reading fluency when they participated in Headsprout in addition to their regular reading instruction. Other researchers (Campuzano et al., 2009; Dynarski et al., 2007), however, found that use of the Headsprout program had no effect on student achievement in reading as measured using group administered standardized tests of reading. These differences in outcomes may be related to the amount of outside support provided to teachers who implemented the Headsprout program and to the fidelity with which the program was implemented.

Need for the Study

The goal of this research was to examine the effectiveness of a specific, internet based reading intervention, Headsprout, for improving the reading skills of young students at risk for reading failure. This study is timely and important for two reasons. First, the need for effective early intervention in the area of reading is well documented in the literature (e.g. Cooke et al., 2010; Lyon, 1997; Vellutino et al., 1996). Secondly, existing research on CAI and reading problems (e.g., Cheung & Slavin, 2011; Hall et al., MacArthur et al., 2001) suggests that CAI holds promise for young students at risk for developing reading problems. Results from many of these studies are limited for several reasons including small sample sizes, lack of adequate control groups, and potential researcher bias (Soe et al., 2000). The existing literature does not yet allow for firm conclusions about the effectiveness of CAI in the area of reading to be drawn. The current study extends the available literature by examining the effectiveness of a specific computer-based reading intervention, Headsprout, which will add to the existing information regarding the effectiveness of CAI in the area of early reading instruction. The current study utilizes an existing data set from a school system that implemented the CAI program using only the resources typically available to the school system, which will supplement the existing literature by providing information about the effectiveness of CAI when it is implemented outside of carefully controlled research settings.

Purpose of the Study

The purpose of this study was to investigate the impact of an internet-based reading program, Headsprout, on the reading skills of kindergarten and first grade students at risk for reading difficulty in one suburban district. In this school district, some kindergarten and first grade students who were having difficulty learning to read as determined by their performance on the DIBELS (Good & Kaminski, 2002) and classroom teacher professional judgment were provided with supplementary instruction through Headsprout, an internet based early reading instructional program. For the purposes of this study, the kindergarten and first grade student population identified as at risk for reading failure was conceptualized as consisting of students who participated in Headsprout Early Reading and students who did not participate in the program. Students with the lowest DIBELS scores at each grade level at each school were assigned to participate in Headsprout.

The impact of Headsprout on the reading skills of kindergarten and first grade students at risk for reading failure was the independent variable. The proposed data analysis statistic was ANCOVA. The first proposed covariate for each of the two outcome measures in both the kindergarten and first grade data sets was student age in months. This covariate was proposed because prior research has shown that student age is correlated with the acquisition of early literacy skills (Paris, 2005). Additionally, a pretest measure of early reading skill was proposed to minimize the effects initial differences in reading skill between the control group and the treatment group at each grade level. That is, for the kindergarten data set fall LNF scores were proposed as a covariate in the analysis of spring LNF scores and fall ISF scores were proposed as a covariate in the analysis of spring NWF scores. For the first grade data set, fall NWF scores were proposed as a covariate in the analysis of spring ORF scores and CRCT scores.

Because early intervention that incorporates the five critical components of reading instruction has been shown to improve the reading skills of students who are at risk for reading failure, it was hypothesized that participating in Headsprout would improve the reading skills of the kindergarten and first grade students at risk for reading failure who participated in the program beyond the improvement in reading skill demonstrated by students at risk for reading failure who did not participate in Headsprout. Specifically, the research questions addressed in this study and answered throughout the remaining chapters include the following:

Research Question 1. Do Kindergarten students at risk for reading failure who participate in Headsprout in addition to their regular classroom reading instruction demonstrate different skill levels in reading as measured by the DIBELS Letter Naming Fluency (LNF) and Nonsense Word Fluency (NWF) tasks than kindergarten students at risk for reading failure who do not participate in Headsprout?

Research Question 2. Do first grade students at risk for reading failure who participate in Headsprout Early Reading in addition to their regular classroom instruction demonstrate different skill levels in reading as measured by the DIBELS Oral Reading Fluency (ORF) task or the Georgia Criterion Referenced Competency Test (CRCT) than first grade students at risk for reading failure who do not participate in Headsprout?

CHAPTER 2

LITERATURE REVIEW

The purposes of this chapter are to review the literature: 1) contributing to the current understanding of effective early reading instruction and intervention for kindergarten and first grade students at risk for reading failure, 2) regarding effective methods to identify and monitor the progress of young students at risk for reading failure, 3) regarding CAI for kindergarten and first grade students in the area of reading, and 4) regarding Headsprout, the internet-based CAI program for early reading instruction that is the focus of this research. Within the area of early reading intervention, the primary focus of the literature review is on instruction and remediation in phonemic awareness, phonics, and sight word recognition because proficiency in these skills by the end of first grade appears to be essential for children to become good readers (NRP, 2000; Pressley, 2005; Shaywitz & Shaywitz, 2004; Sweet, 2004). Fluency, vocabulary development, and comprehension are discussed in less detail. While fluency, vocabulary development, and comprehension are also essential to good reading, poor reading outcomes in kindergarten and first grade appear to be more closely related to deficits in phonemic awareness, phonics, and sight word recognition, which must develop before fluency with and comprehension of written text can occur (Foorman & Torgesen, 2001). The methods used to identify and monitor the progress of students who require intervention to become skilled readers are as important as the content of early reading interventions; therefore, the literature regarding RTI, an evidence-based method for the early identification of students at risk for academic failure (Burns et al., 2005; Cusumano, 2007), is also reviewed in this chapter. Finally, this chapter reviews the literature regarding the effectiveness of computer-based reading interventions designed for kindergarten and first grade students. The widespread availability of computer technology in schools has resulted in the frequent use of computer software by teachers and students as part of the daily instructional curriculum. Research regarding the effectiveness of technology, however, remains in its relative infancy, and researchers continue to explore the efficacy of CAI with typically developing children as well as children at risk for reading failure.

The intuitive appeal of CAI in the schools is based on the idea that computer applications can actively engage students in instruction without requiring direct, immediate involvement by a teacher. Well-designed computer programs have the potential to provide students with practice in specific reading skills and strategies at an individualized pace and with immediate feedback (Hall et al., 2000). Research regarding the efficacy of CAI for early reading instruction and intervention, however, has had mixed results. This chapter reviews the literature regarding early intervention for at risk readers since the publication of the National Reading Panel's report in 2000, reviews the literature regarding RTI, and finally, reviews the empirical research conducted using CAI reading interventions for kindergarten and first grade students who are learning to read in English, including those at risk for reading failure, published since 2000.

Reading Theory and Models of Reading Development

There are few areas of pedagogy that have been debated as extensively as how to teach reading to young children as they enter school. While educators share the goal of providing children with early literacy instruction that creates students who are successful readers, and extensive research documents the importance of children's early literacy achievements as essential for later academic success (Pressley, 2005; Snow, Burns, &

Griffin, 1998; Sweet, 2004), the disturbing reality is that children who do not master early literacy skills remain at risk as learners throughout their school years (Pressley, 2005; Shaywitz & Shaywitz, 2004; Sweet, 2004). Given the importance of early literacy proficiency, a great deal of research has been directed toward understanding the instructional approaches that facilitate early reading success.

Theories of reading development are often broadly categorized as *bottom up*, *top down*, or *interactive* based on the way they model reading development (Frey, Lee, Tollefson, Pass, & Massengill, 2005; Tracey & Morrow, 2006). Bottom up models (i.e. Chall 1976; Ehri, 2005) typically describe reading as a series of stages beginning with letter sound recognition and ending with the construction of meaning. In bottom up models of reading, first letters are identified, then sounds are attached to the letters, next the meaning of the decoded word is added, and finally, after all the words are processed, the meaning of the sentence is understood. Basic reading skills such as phonics are taught in isolation, and the reader learns to decode before he or she is expected to attend to meaning. The context of what is being read and background knowledge of the reader is not considered essential to the process of decoding the words on the page (Tracey & Morrow, 2006).

While bottom up models of reading development provide an understandable model of how reading decoding occurs, they are less able to account for the comprehension of text, which is the end goal of reading instruction (Pressley, 2005). Paris (2005) expands on this criticism of bottom up models of reading development in his presentation of constrained skills theory. Paris argues that phonemic awareness, phonics, and reading fluency are constrained skills because all readers master them fairly early in their reading development. Since these skills are mastered early, they are important predictors of later reading skill only within a narrowly defined period of time when children are first beginning to learn to read. According to constrained skills theory, too much emphasis on the measurement and remediation of phonemic awareness, phonics, and fluency in young children who are beginning to learn to read may have at least two important consequences. First, analysis of the development of constrained skills using traditional parametric statistics may miss-identify students whose literacy skills are developing normally as at risk for reading failure. Secondly, over-emphasis on instruction designed to teach phonemic awareness, phonics, and may reduce the amount of classroom time spent in authentic literacy activities. This may, in turn, result in children who can decode words but are not truly literate beyond the basic ability to decode words. In contrast to the constrained skills of phonemic awareness, phonics, and fluency, vocabulary and comprehension are described as unconstrained skills because they continue to develop across the lifespan (Paris, Carpenter, Paris, & Hamilton, 2005).

Top down reading theories, such as Emergent Literacy Theory (Tracey & Morrow, 2006) posit that reading is one part of the broader concept of literacy development, which begins at birth and includes both oral and written language skills. Emergent literacy theory also emphasizes that literacy development occurs across a wide variety of settings that include homes and communities in addition to schools. Models of reading built from top down theories emphasize the importance of what the reader brings to the reading process. The accurate decoding of each word in a reading passage is not considered essential for comprehension. Background knowledge, including knowledge about the reading topic, knowledge of text structure, knowledge of sentence structure, knowledge of word meanings, and knowledge of letter-sound correspondence, are all important to reading because readers use information from each of these sources to anticipate upcoming text. If what is read is consistent with what the reader anticipated, reading progresses smoothly. If not, reading slows, and the reader attends more closely to the printed text. Top down models of reading provide an explanation for reading comprehension, but they are less able to explain how individuals read unfamiliar texts to learn new information when they have little knowledge of the topic and cannot generate predictions (Tracey & Morrow, 2006).

Interactive theories of reading (i.e. Rumelhart, 2004) recognize the importance of and interaction between bottom-up and top-down processes in skilled reading. They emphasize both what is on the written page and what a reader brings to reading. Readers develop reading skills and reading strategies in context rather than in isolation. Models of reading that develop from interactive theories portray skilled readers as both good decoders and good interpreters of text. Decoding skills become more automatic but no less important as reading skill develops. Skilled reading is the result of interaction between reader and text (Tracey & Morrow, 2006). Interactive models of reading often underlie balanced approaches to the teaching of reading. Balanced literacy programs include the direct instruction and modeling of skills, strategies, and processes as well as student-centered reading and writing activities (Frey et al., 2005; Snow et al., 1998).

Cognitive Load Theory

Research in the area of Cognitive Load Theory (CLT; i.e. Paas, Renkl & Sweller, 2003, 2004) also contributes to understanding how reading develops in both typically progressing students and students at risk for reading failure. CLT suggests that learning happens best when the amount of new information to be learned does not overwhelm the limited capacity of human working memory. It further suggests that working memory can become effectively unlimited when dealing with familiar material stored as schemas. Schemas are organized and automated routines for processing information or performing an action stored in long-term memory. Automated schemas can be used unconsciously, without placing demand on working memory capacity (Chandler & Sweller, 1991). For example, according to the tenets of CLT, skilled reading develops as a learner constructs increasingly complex schemas for reading skills, such as sight word reading, by combining elements of previously mastered schemas for more basic skills, such as phonemic awareness and phonics, as described in bottom up models of reading development.

Three types of cognitive load, intrinsic, extraneous, and germane, are important in CLT and should be considered when providing instruction (Paas et al., 2004). Intrinsic load is the cognitive load imposed by the information to be learned. Extraneous load is imposed by information and activities that do not contribute to the processes of creating and automizing (i.e., learning) schema. Germane load is not imposed by the information to be learned, but is a result of the activities required to learn schemas.

CLT suggests that instruction should be designed to decrease extraneous load, while simultaneously increasing germane load, so that the combined intrinsic and germane loads do not exceed working memory capacity. When this is not possible because reducing extraneous load also reduces germane load, it may be necessary to provide simplified learning tasks even though this may partially compromise full understanding (Pollock, Chandler, & Sweller 2002). This point may be particularly relevant when teaching reading decoding skills to students at risk for reading failure. The application of CLT to beginning reading instruction suggests that for some students it may be necessary to teach decoding skills separately from comprehension skills to ensure that working memory demands are not overloaded to the point that no learning occurs. CLT also suggests that knowledge of the learners level of expertise (Kalyuga, Ayres, Chandler, & Sweller, 2003) is important to ensure that the cognitive demands of presented tasks are neither overwhelming for novice or struggling learners nor too simple for learners with more expertise. Therefore, children who are at risk for reading failure may require qualitatively different instruction than those whose reading skills are developing as expected.

Early Reading Intervention

Expert consensus has identified the instructional strategies that are associated with successful literacy outcomes. These strategies reflect a balanced approach to literacy instruction and include literacy rich classroom environments that incorporate high-quality literacy centers, the use of authentic children's literature as a central component of literacy instruction, opportunities for social collaboration among students, and extensive professional development for teachers (NRP, 2000). Furthermore, *explicit instruction* and meaningful practice in the areas of phonics, phonemic awareness, vocabulary, fluency and comprehension are also important (Foorman & Torgesen, 2001; NRP, 2000; What Works Clearinghouse [WWCH], 2007). Explicit instruction is defined as instruction that sequentially reviews previous work, presents new material, provides guided practice, provides feedback and correction, provides independent practice, and provides weekly and monthly reviews (Rupley, Blair, & Nichols, 2009).

While the debate regarding the amount of explicit instruction required and how best to provide this instruction for typically developing students continues (Frey et al., 2005), there is general agreement that explicit instruction in the areas of phonemic awareness, phonics, fluency, vocabulary, and comprehension, embedded within a balanced reading program that also includes an abundance of authentic reading and writing activities, benefits all students and is crucial for those students who are at risk for reading failure (e.g., Foorman & Torgesen, 2001; Frey et al., 2005; Pressley, 2001). Research also indicates that intensive, supplementary instruction in phonemic awareness and phonics skills can significantly improve long-term outcomes for struggling readers (Al Otaiba & Fuchs, 2002; Foorman & Torgesen, 2001; McIntyre et al., 2005; NRP, 2000; WWCH, 2007). *Intensive instruction*, as the term is typically used in research (e.g., Foorman & Torgesen, 2001), refers to small group or individual instruction that is targeted to the needs of individual students, occurs at least several times a week, lasts for sessions of at least 15 minutes, and continues over several months (Cavanaugh, Kim, Wanzek & Vaughn, 2004).

Phonemic Awareness.

Phonemic awareness is knowledge about the sounds in language (Foorman & Torgesen, 2001; Scarborough & Brady, 2002). Before students are ready to read, they must know that words are made up of smaller sound-parts. They also need to know how the sounds in words work. The focus of phonemic awareness is narrow. Students learn to identify and manipulate the individual sounds in words. Manipulating the sounds in words includes blending or otherwise changing words. Effective phonemic awareness instruction teaches children to notice, think about, and work with the sounds in spoken language (Scarborough & Brady, 2002). For example, a student who is beginning to establish phonemic awareness can rhyme words. At the next level, the student can recognize the odd word within a set of words—if the set of words is fan, can, and bed the student recognizes bed as the odd word. Next, the student learns to blend sounds to make words. For example, /m/, /a/, /t/ becomes mat. The student then learns to segment a word like stop into the sounds /s/, /t/, /o/, /p/. Finally, the student is able to remove a sound from a word to make a new word on request. For example, when asked to stay stop without the /s/, the child says top (Pressley, 2005).

Phonics.

Phonics instruction is about letter-sound correspondences. Before phonics instruction can be effective, however, children must recognize that the words that they hear, and will eventually read, are made up of individual sounds (phonemic awareness). The goal of phonics instruction is to help children to learn and use the alphabetic principle. The alphabetic principle is the understanding that there are systematic and predictable relationships between written letters and spoken sounds. Phonics instruction teaches children the relationships between the letters of written language and the sounds of spoken language. Learning that there are predictable relationships between sounds and letters allows children to apply these relationships to both familiar and unfamiliar words and to begin to read with fluency (Kamil, 2008). While phonemic awareness is important to phonics instruction, research suggests that phonemic awareness and phonics skills can be effectively taught at the same time (Pressley, 2005).

The term phonics is used to describe a wide range of different reading activities and programs. There are two main types of explicit phonics instruction: synthetic phonics and analytic phonics. *Synthetic phonics* involves the development of phonemic awareness as the beginning point for instruction. As part of the decoding process, the reader learns 44 phonemes, the smallest units of sound, and their related graphemes, the written symbols that represent the phonemes. For example, the student learns that the letter *s* represents the phoneme /s/ as heard at the beginning of the word *sit*. In order to read a word the student must recognize each grapheme, sound out the phoneme it represents, and blend the phonemes together to pronounce a word. In other words, to read the word *sit*, a student must recognize each of the letters, s, i, and t; associate each letter with the phoneme it represents, /s/, /i/, /t/; and, finally, blend the phonemes to produce the word *sit*. Synthetic phonics instruction often involves highly systematic whole-class instruction that begins in kindergarten. The sounds and their corresponding written symbols are taught in quick succession, often using a multi-sensory approach in which children see the grapheme, listen to the phoneme, say the phoneme, and accompany this by doing an action. The multi-sensory approach to phonics instruction appears to support most learners in remembering the sound-symbol relationships (Ritchey & Goeke, 2006).

Analytic phonics, in contrast, is an approach that requires the learner to discover letter-sound correspondences from words. Letter sounds are not taught in isolation. Instead, instruction begins with a whole word and students are directed to analyze parts of the word to understand how letters combine to form words. Analogy, or onset-rime phonics, is a specific type of analytic phonics in which children are asked to look at chunks of words or word families. For example students might be asked to look at a group of words that end with *an*. The first sound is added or changed to make different words: *c-an*, *f-an*, *m-an*, or *r-an*.

The NRP (2000) concluded that synthetic phonics programs were especially effective for younger students who were at risk for reading difficulties or identified as disabled as compared to analytic phonics programs. The panel's conclusions were that children with learning disabilities and children who are low achievers make the greatest gains when they are provided with systematic, synthetic phonics instruction. Moreover, systematic synthetic phonics instruction was significantly more effective than other types of instruction in improving the reading skills of children from low socioeconomic levels (NRP, 2000). Other research evidence indicates, however, that some struggling readers make greater gains when they are taught using analytic phonics (Chera & Wood, 2003; Comaskey, Savage & Abrami, 2009; Pressley, 2005).

It appears that good readers use both synthetic and analytic phonics skills when reading. Students who become good readers initially decode words based on letter sound correspondences, but as they encounter groups of sounds that frequently occur together, such as the suffix *–ing* or the suffix *–tion*, they begin to read them as a unit, which is called structural analysis. Finally, they begin to recognize words as a single unit. Word recognition occurs almost instantaneously and is called sight word reading (Ehri, 2005).

Sight Words.

Good readers read familiar words by accessing them as whole words in memory, which is called sight word reading (Ehri, 2005). Sight word reading is one desired result of instruction in phonemic awareness and phonics (Compton, Appleton, & Hosp, 2004; Pressley, 2005; Shaywitz & Shaywitz, 2004). In typically developing beginning readers, sight word reading is not a process of rote memorization (Pressley, 2005; Shaywitz & Shaywitz, 2004). Instead, as the beginning reader successfully sounds out a word, connections are created in memory between the letter patterns of the word on the page and the word in memory. Each successive successful sounding out of the word strengthens these connections until the word is recognized automatically as a sight word. For example, the first time a reader encounters the written word *cat*, it is sounded out as /c/, /a/, /t/. When *cat* is read after several exposures, it may be read as /c/, -at. Finally, the whole word is immediately recognized and almost instantaneously brings to mind the image of a small, furry animal that meows and purrs. Additionally, some high frequency words are phonetically irregular and must be taught as sight words. For typically developing beginning readers, frequently encountered words rapidly become sight words (Ehri, 2005; Pressley, 2005; Shaywitz & Shaywitz, 2004). Some research evidence suggests that explicit, intensive instruction in phonemic awareness and phonics skills helps non-disabled children at risk for reading failure develop sight word recognition skills that are similar to those of typically developing readers (Kamil, 2008; Shaywitz & Shaywitz, 2004). Because sight word reading is automatic, it allows fluent reading to develop and creates a foundation for reading comprehension (Pressley, 2005; Shaywitz & Shaywitz, 2004).

Fluency.

Reading fluency is the ability to accurately and quickly decode text, so that comprehension can occur (Kamil, 2005). Children who have mastered the basic skills of reading are able to identify words quickly and with very little effort. Fluency depends on the automatic recognition of high-frequency words and skilled decoding of less frequently encountered words (Compton et al., 2004; Shaywitz & Shaywitz, 2004). The ability to read words by sight automatically is the key to skilled reading (Ehri, 2004).

Research reviews (i.e. Kuhn & Stahl, 2003) suggest that reading a variety of moderately challenging, as opposed to easy, text with adult support can increase reading fluency and reading comprehension for students at risk for reading failure. Intervention does not, however, appear to completely eliminate the differences in reading fluency between typically developing readers and those at risk for reading failure (Kuhn & Stahl, 2003).

Vocabulary and Comprehension.

Two additional areas that research has identified as important to reading instruction are comprehension and vocabulary (NRP, 2000). Comprehension is the process of making meaning at both the literal and inferential level from what is read (Moats, 2004). Comprehension is generally viewed as the essence of reading (Paris, 2005; Pressley, 2005). The NRP (2000) noted that reading comprehension is an active process that requires an intentional and thoughtful interaction between a reader and the text. Students derive meaning from text only when they engage in intentional, problem solving thinking processes. Research (e.g., NRP, 2000) suggests that text comprehension is enhanced when readers actively relate the ideas represented in print to their own knowledge and experiences and construct mental representations in memory. Formal instruction in the application of a variety of comprehension strategies is highly effective at improving students' understanding of text (Kamil, 2005; NRP, 2000).

Reading comprehension cannot be understood without an appreciation of the role that vocabulary development plays in the understanding of text (NRP, 2000). Vocabulary is the knowledge of the meanings and pronunciations of words that are used in both oral and written language (Kamil, 2005). One of the most consistent findings in the research related to vocabulary and comprehension is that people with larger vocabularies tend to comprehend better (Pressley, 2005; Sedita, 2005). This does not necessarily imply, however, that increasing a reader's vocabulary will result in a corresponding increase in comprehension (Pressley, 2005). Vocabulary knowledge is related to reading comprehension, and it determines how much of a text the student will comprehend. The larger a beginning reader's vocabulary, the easier it is for the reader to understand text because a larger vocabulary increases the likelihood that both sight words and decoded words will have meaning for the reader once they are read (Kamil, 2005).

In its analysis of the research on vocabulary instruction, the NRP (2000) found that vocabulary should be taught both directly and indirectly. Direct instruction in vocabulary means teaching students the meanings of specific words in addition to teaching students how to decode words. For example, pre-teaching vocabulary prior to reading a selection is direct vocabulary instruction. Students are taught an estimated 400 words per year in school through direct instruction (Beck, McKeown & Kucan, 2002). It is impossible, however, to teach students all of the words they need to know through direct instruction, and most vocabulary is learned through indirect methods (Kamil, 2005). Indirect methods of vocabulary instruction occur primarily when children engage in consistent, extensive, and rich verbal interactions with linguistically mature people (Pressley, 2005).

Effective instruction and intervention for young children who are at risk for reading failure requires explicit and intensive instruction in phonemic awareness and phonics over and above the amount of instruction in these skills required by typically developing students. Young readers at risk for reading failure require additional presentations of words to be decoded above the number required by typically developing reader to add words to their store of automatically recognized sight words. For young students, difficulties in the areas of phonemic awareness, phonics, and sight word reading prevent fluent reading from occurring, which in turn, reduces the student's ability to comprehend what has been read. It is important to remember, however, that students who are at risk for reading failure often require remediation in the areas of vocabulary and comprehension in addition to remediation in phonemic awareness, phonics, and fluency to become skilled readers.

Response to Intervention (RTI) in Reading

Response to Intervention (RTI) is a multi-tiered, general education process through which schools can monitor the progress of their students within the curriculum in order to identify students who are at risk for academic failure and provide them with interventions in a timely manner (Fletcher & Vaughn, 2009; McCook, 2006; Torgesen, 2009). RTI processes in schools are based on public health models of disease prevention that differentiate primary (i.e., Tier 1), secondary (i.e., Tier 2) and tertiary (i.e., Tier 3) levels of prevention and intervention that increase in cost and intensity depending on the patient's response to treatment (Fuchs et al., 2003; Vaughn, Wanzek, Woodruff, & Linan-Thompson, 2007). RTI models screen all children for academic and behavioral problems at Tier 1, and monitor the progress of children identified as at risk for these problems at Tiers 2 and 3. Children who do not make adequate progress at Tier 2 are provided with increasingly intensive and individualized intervention at Tier 3 (Fletcher & Vaughn, 2009; Torgesen, 2009). RTI models depend on the implementation of evidence-based interventions designed to prevent or remediate academic difficulties (Fuchs et al., 2003; McCook, 2006).

All students in a school or school district participate in Tier 1 of RTI models (Fuchs et al., 2003). All students are screened using brief screening tools that demonstrate diagnostic utility (e. g., AIMSWeb, Shinn & Shinn, 2002; DIBELS, Good & Kaminski, 2002; STEEP, Witt, 2002) for predicting performance on the reading state assessments in the elementary grades or on the local graduation requirements at the secondary level (Fletcher & Vaughn, 2009). At Tier 1 it is assumed that the general curriculum used in the school is adequate to meet the instructional needs of at least 80% of students (McCook, 2006). In Tier 1 instructional practices are those that can be competently and independently implemented by general education teachers (Fuchs & Fuchs, 2009). Additionally, successful RTI models provide classroom teachers with professional development geared toward helping them meet the needs of their students as identified on universal screening measures (Torgesen, 2009). For example, if 25% of first graders in a given school score below the research-identified cut score on a universal screening measure for sound-symbol relationships, teachers would be provided with professional development opportunities in the area of phonics instruction. One often-cited outcome of successful RTI implementations is that teachers learn to use data to make day-to-day instructional decisions for their students (e.g., Torgesen, 2009).

Children who are not meeting local or national benchmarks when they participate in Tier 1 instruction are moved to Tier 2. These students receive additional instruction via research-validated interventions in their area of difficulty, and they are frequently reassessed using alternate forms of screening measures to track their progress over time (Stecker, Fuchs, & Fuchs, 2005). Tier 2 interventions may be implemented using either a standard protocol or a problem solving model (Fletcher & Vaughn, 2009).

In standard protocol interventions all students in Tier 2 at a given grade level with reading problems receive the same intervention. Standard protocol Tier 2 interventions are time-limited, and they are typically implemented with students in a small group setting. For example, a standard protocol Tier 2 intervention in reading might involve 10 to 15 weeks of 20 to 40 minute lessons delivered three or four times a week using a research-validated curriculum or intervention protocol (Fuchs & Fuchs, 2009). Standard protocol Tier 2 interventions are not intended to be implemented independently by the classroom teacher. Instead, they are designed to be supervised by professional educational support staff such as instructional coaches while paraprofessionals serve as tutors (Fuchs & Fuchs, 2009).

In contrast, problem solving RTI models do not specify a specific intervention that all students identified as at risk in a given instructional area will receive. Tier 2 intervention in the problem solving model of RTI is a process by which teachers and other educational personnel use data not only to determine the magnitude of the student's reading problem, but also to analyze its cause, design a goal-directed intervention, conduct the intervention, monitor student progress, modify the intervention as needed based on student responsiveness, evaluate its effectiveness, and determine future actions (Telzrow et al., 2000). In the problem solving model, a student's progress within an intervention is monitored weekly (Vaughn, Wanzek, & Fletcher, 2007). If a student is not making adequate progress as defined by the problem solving team over a period of three to four weeks, the team changes or modifies the intervention with the goal of improving the student's responsiveness. If, on the other hand, a student consistently exceeds the goal set by the intervention team over a period of three to four weeks, the student's goal may be increased, or if the student has made sufficient progress toward local or national norms, the student may be returned to Tier 1.

Regardless of the type of Tier 2 intervention provided, students who do not make adequate progress in Tier 2, are provided with increasingly intensive and individualized intervention at Tier 3. In many conceptualizations, Tier 3 of the RTI model is synonymous with special education services (Fuchs & Fuchs, 2009). Intervention at this level is generally individualized for the particular student, and it may involve smaller groups, increased instructional time, a more specialized teacher, or instructional content that is below the student’s grade level placement (Fuchs et al., 2003; Fuchs & Fuchs, 2009; McCook, 2006).

For young students at risk for reading failure, Tiers 1 and 2 of the three tiered model are critical because they represent the best opportunity for a student to receive early intervention designed to prevent reading failure. Studies of RTI models suggest that interventions implemented within an RTI model result in successful outcomes for early reading interventions as often as 95% to 98% of the time (Berninger et al., 2003; Mathes et al., 2005; McMaster, Fuchs, Fuchs, & Compton, 2005). Successful district-wide implementations of RTI models across the country report increases in overall academic achievement scores and decreases in special education referrals (e.g., Torgesen, 2009; VanDerHeyden, Witt, & Gilbertson, 2007). Furthermore, in a review of reading interventions that consisted of at least 100 intervention sessions, Wanzek and Vaughn (2007) found that most of the interventions resulted in effect sizes in the moderate to large range. Studies that involved kindergarten and first grade students, used a standard protocol intervention, and delivered the intervention either one-on-one or in small groups generally reported the largest effect sizes. RTI models are a potentially effective method for changing educational outcomes for students identified as at risk for reading failure (Fletcher & Vaughn, 2009; Fuchs & Fuchs, 2009).

The implementation of either problem-solving or standardized protocol RTI models requires significant effort and resources (McCook, 2006). Providing effective Tier 1 instruction to all students requires ongoing professional development for teachers and ongoing screening and progress-monitoring of students. Maintaining these practices demands an extensive professional development regimen from well-trained and committed professionals who may not be readily available (National Association of State Directors of Special Education, 2008). Although effective Tier 1 instruction reduces the number of students who will be identified as at risk, a significant number of students will still require supplemental intervention at Tier 2. These Tier 2 interventions must be provided or at least supervised by trained personnel (e.g., instructional coaches, classroom teachers, paraprofessionals). Tier 2 intervention must also be continuous because each round of universal screening at Tier 1 will likely identify additional students who require intervention.

Unfortunately, many school districts do not perceive that they have the personnel and resources to effectively implement all the elements of RTI models (Response to Intervention Adoption Survey, 2010). Computer Aided Instruction (CAI) may provide an avenue to address many of the perceived difficulties of both problem solving and standard protocol models of RTI because it is intended to provide instruction that is both based on a standardized curriculum and individualized to meet the needs of individual students without the need for large amounts of extra work from school personnel. Well-designed CAI can also provide immediate feedback, provide frequent opportunities to respond, and create high rates of success (Sivin-Kachala & Bialo, 1998). CAI may have the potential to reduce educational costs while enhancing educational effects (Kulik & Kulik, 1991; Sivin-Kachala & Bialo, 1998). Using computer aided instruction to provide intervention for children at risk for reading failure is one potential method for partially addressing the perceived lack of resources in school systems (Blok et al., 2002; Hall et al., 2000; Soe et al., 2000).

Computer Aided Instruction for Reading

In one of the first definitions of CAI, the Association for Educational Communications and Technology (1977) defined CAI as a method of instruction that uses a computer to teach the student. The computer programming provides instruction that is designed to teach, guide, and test the student until proficiency is attained. CAI, which arguably has its origins in B. F. Skinner's fill in the blank teaching machines (Maris, 2005), has changed drastically over the past several decades.

The first CAI programs designed for public schools began at Stanford University in the early 1960s, and this federally funded initiative included instructional components in many subject areas (Atkinson et al., 1970). Several important conclusions drawn from that research remain relevant to CAI today. First, CAI was most effective as a supplement to, rather than a replacement for, the classroom teacher (Atkinson et al., 1970). Second, effective CAI was developed from a theory-based curriculum grounded in empirical research, and the curriculum guiding the CAI needed be consistent with the curriculum used within the classroom (Atkinson et al., 1970). Finally, they concluded that most important benefit of CAI was that it could tailor instruction to the needs of each student based on his or her responses, allowing instruction to be individualized for each learner (Wilson & Atkinson, 1967).

The three main types of CAI are drill-and-practice programs, tutorial programs, and simulation programs (Desrochers & Gentry, 2004). Drill and practice programs provide the learner with exercises that reinforce specific skills learned in the classroom much like traditional practice worksheets (Desrochers & Gentry, 2004; Soe et al., 2000), but they have the advantage of providing learners with immediate feedback and appear to be more motivating to students than traditional worksheets (Sivin-Kachala & Bialo, 1998; Soe et al., 2000).

Tutorials present new instructional material, test the learner's knowledge of the material, and provide feedback to the learner (Desrochers & Gentry, 2004; Soe et al., 2000). Tutorial programs are usually designed to tailor instruction to meet the learner's needs based on his or her responses to questions within the program (Soe et al., 2000). More recent examples of tutorial programs are commonly referred to as integrated learning systems (ILS). The term ILS describes software programs that provide sequential instruction for students, often over several grades, while keeping extensive records of student progress. Most ILS programs provide instruction in the basic skill areas of reading and mathematics (Kulik, 2003). CAI simulations do not teach new material. Instead, simulation programs are designed to model realistic situations in which the learner responds and receives feedback on his or her performance. Simulations may be effective for promoting generalization of learning to natural situations (Desrochers & Gentry, 2004).

Regardless of the type CAI, research in the area of CAI as well as research in the area of CLT (Paas et al., 2003, 2004) has identified several features common to effective CAI. First, effective CAI provides explicit instruction, as previously defined, that, at a minimum, specifies instructional objectives that the program is designed to teach (Desrochers & Gentry, 2004). Effective CAI also minimizes the presentation interesting but non-essential information (Mayer & Moreno, 2003). It prompts desired responses (i.e., highlighting text, providing hints on the screen) during initial instruction and gradually fades the prompts as the learning occurs (Desrochers & Gentry, 2004). Well-designed software presents information in multiple formats. For example, providing some of the information to be learned through pictures, while simultaneously providing audio rather than visual (i.e. text) presentation of supplemental information results in improved learning (Mayer & Moreno, 2003). Another important design feature of effective software is active and frequent student responding, which appears to increase both student engagement and the likelihood that the student will generalize learning to other situations (Kulik, 2003). Finally, providing immediate feedback regarding response accuracy, including providing the correct answer after one incorrect response is important (Desrochers & Gentry, 2004).

Despite the consensus regarding many early literacy instructional practices and the increase in understanding of how to design effective educational software, the appropriate role for the use of technology in early reading instruction remains unclear because the existing research base lacks the plethora of rigorously designed and conducted studies required to draw such conclusions (Blok et al., 2002, Patterson, Henry, O'Quinn, Ceprano, & Blue, 2003). More recent reviews (Cheung & Slavin, 2011) suggest that the types of supplementary CAI programs available currently and in the recent past are not producing educationally meaningful effects in reading for young students. This may be due, at least in part, to the fact that many commercially available software programs approved for teaching reading and writing in early elementary school are non-instructional in that they do not track student progress, provide feedback, or adapt to student needs. Many rely on decades old interface designs, program features, and content features, which limit their usefulness as educational tools (Lovell & Phillips, 2009).

The appeal of CAI for reading, in particular, is the belief that well-designed computer programs have the potential to supplement instruction provided by classroom teachers in meaningful ways (Blachowicz et al., 2009; Hasselbring, 1986). Well-designed computer programs that incorporate both the elements of effective general instruction (i.e., explicit, strategic, and scaffolded instruction; high percentages of engaged time; high success rates; immediate, corrective feedback; and frequent reinforcement; [Hall, Hughes & Filbert, 2000]) and components of effective reading instruction as identified in research (e.g., NRP, 2000) are becoming available. These programs have the potential to teach students at their own pace, provide immediate feedback, increase motivation, and create tailored practice schedules that provide only the necessary amount of repetition (Cheung & Slavin, 2011; Soe et al., 2000; Torgesen, 1986; Wood, Pillinger & Jackson, 2010). Despite the challenges that exist for researchers studying technology in literacy education, a number of recent studies have begun to investigate the components of reading instruction as provided through CAI, which has added depth to the field. The increased availability of computers and software in schools has allowed researchers to investigate the effectiveness of CAI by grade level, subject area, and program type.

Studies assessing drill and practice CAI.

Drill and practice programs are CAI in which the computer provides the student supplemental instruction in the form of exercises that reinforce the learning of specific skills taught in the classroom. The programs provide immediate feedback on the correctness of the student's response and often tailor instruction to meet the needs of the individual student based on his or her responses within the program (Hall et al., 2000; Soe et al., 2000). Although drill and practice programs can provide instruction in phonemic awareness, phonics, vocabulary, fluency and reading comprehension (Hall et al., 2000), the skills most commonly taught using this type of software are phonemic awareness and phonics.

DaisyQuest and Daisy's Castle. A number of drill and practice CAI programs have explicitly targeted phonological awareness skills. Two of the first were the DaisyQuest (Erickson, Foster, Foster, Torgesen & Packer, 1992) and Daisy's Castle (Erickson, Foster, Foster, Torgesen & Packer, 1993) programs, which are no longer available. They were developed at the University of Florida, and studies of these programs provide some of the strongest evidence for the potential effectiveness of CAI in the area of early reading instruction. DaisyQuest taught phonological awareness skills such as recognizing words that rhyme, recognizing words that have the same beginning, middle, and ending sounds, and combining phonemes to create words. Daisy's Castle was a continuation of the DaisyQuest program. It taught and reinforced the skills of segmenting words into individual phonemes and blending phonemes into words (Barker

& Torgesen, 1995). The programs used storylines about a dragon to motivate students to learn skills. As a student completed each instructional activity he or she was given clues to help locate Daisy the Dragon's lost eggs in the DaisyQuest program or to find Daisy's hiding place in the Daisy's Castle program (Foster et al., 1994). The programs also included a computer-based phonological awareness skills assessment called Undersea Challenge (Foster et al., 1994).

Foster and colleagues (1994) investigated the effectiveness of the DaisyQuest (Erickson et al., 1993) program with kindergarten aged students. The researchers, who were involved in the development of the DaisyQuest program, randomly assigned second semester kindergarten students into two groups. All students received standard screening measures for verbal ability; a Phonological Awareness Test, a Production Test of Blending, and a Production Test of Segmenting, which were developed for the study; and Undersea Challenge, which was part of the software. Thirty-four students were randomly assigned to the treatment group, and 35 matched students served as the control group. The students in the treatment group used DaisyQuest in groups of four students under the supervision of project staff for a total of 16, 20-minute sessions. The control group continued to receive their typical kindergarten curriculum. Results indicated that the treatment group significantly outperformed the matched, no-treatment control group on Undersea Challenge (effect size of .88) and the Production Test of Blending (effect size of 1.32), but not on the Test of Phonological Awareness or the Production Test of Segmenting (Foster et al., 1994).

In another study of the DaisyQuest and Daisy's Castle programs, Barker and Torgesen (1995), who were also associated with the development of the programs, compared computer driven phonological awareness training using the DaisyQuest and Daisy's Castle programs, with two other computer programs. Fifty-four at risk first graders were randomly assigned to either the DaisyQuest program or one of two control groups. Prior to training, there were no statistical differences between the three groups on pre-test measures of phonological awareness skills, verbal intelligence, or word reading skills. The treatment group spent 25 minutes a day, four days a week for eight weeks systematically working through the DaisyQuest and Daisy's Castle programs. The first control group spent the same amount of training time using an alphabetic decoding program, and the second control group worked with a math skills program. Results indicated that the treatment group outperformed both control groups on post-test measures of phonological awareness. These results reached significance for Undersea Challenge (effect size of 1.1), segmenting (effect size of 1.1), and elision (effect size not reported) but not for sound categorization or blending (Foster et al., 1994).

In contrast to studies investigating the efficacy of CAI alone, Mitchell and Fox (2001) investigated effectiveness of CAI for phonological awareness when directly compared to teacher-delivered instruction. The authors were also associated with Florida State University. The authors examined the potential of DaisyQuest and Daisy's Castle to increase the phonological awareness skills of kindergarten and first grade students who demonstrated below grade level performance in reading. Thirty-six kindergarten and first grade students at risk for reading difficulties were randomly assigned to one of three conditions, a group that received computer-based phonological awareness instruction by using the DaisyQuest and Daisy's Castle programs, a group that received teacher-delivered phonological awareness instruction, and a technology control group that used computerized drawing and mathematics software during the intervention period. Each group worked on their designated intervention for 30 minutes a day, five days a week, for a total of four weeks. Pre- and post-tests for this study included measures of rhyming, segmentation, phoneme isolation, and blending. The results of this study indicated that the children who received computer-assisted and teacher-delivered phonological awareness instruction demonstrated a significant increase in phonological processing skills as compared with those in the control group who used drawing and mathematics software. There were no statistically significant differences between the group who used DaisyQuest and Daisy's Castle and the group that received teacher-directed phonological awareness instruction. This study provided support for the notion that computer assisted interventions have the capacity to deliver phonological awareness training with the same success as teachers.

Mathes, Torgesen, and Allor (2001) investigated the effects of DaisyQuest and Daisy's Castle on low achieving first grade students who were already receiving the peer-assisted literacy strategies (PALS) intervention. Thirty-six first-grade teachers and their classes from eight schools participated in the study. Twelve teachers implemented PALS, 12 teachers implemented PALS and the CAI intervention, and 12 teachers served as a control group. Teachers and their classes were randomly assigned to one of the three treatment conditions, with the exception of 6 teachers who had participated in the PALS intervention the year before and requested to participate in PALS again. These six teachers were randomly assigned to either the PALS or PALS and CAI condition and matched classrooms from the same schools were included in the control group. In each of the PALS classrooms, students participated in the PALS intervention for 35 minute sessions, three times per week for 16 weeks. Students within these classrooms who were identified as low performing also participated in three 20 to 30-minute CAI sessions each week for eight weeks using the DaisyQuest (Erickson et al., 1992) and Daisy's Castle (Erickson et al., 1993) computer programs under the supervision of project staff in addition to participating in the PALS intervention. The control group of teachers and students participated in typical classroom instruction. Study results indicated that the PALS intervention enhanced reading performance, both in terms of statistical significance and in terms of educational relevance, achieving effect sizes that ranged from .37 to .74. The addition of CAI in phonological awareness did not impact student performance beyond the implementation of PALS alone. The authors note, however, that only the lowest achieving students, based on pre-testing results, participated in the PALS with CAI condition, which provides an important alternate explanation of their non-significant findings.

In summary, research into the effectiveness of the DaisyQuest and Daisy's Castle CAI programs suggests that CAI can be as effective as teacher instruction for improving the phonological awareness and early reading skills of kindergarten and first grade students at risk for reading failure in research settings. It is not clear from the existing literature that DaisyQuest and Daisy's Castle would have had the same effects on student performance without the considerable support provided by research staff. In each of the reviewed studies, at risk students participated in the CAI conditions under the supervision of research personnel in groups as small as four students to one researcher. Early intervention research consistently demonstrates that any well-designed phonemic awareness intervention with appropriate content delivered in this type of small group setting is likely to produce statistically significant and educationally meaningful results whether or not it is delivered by computer.

Lexia Learning Systems. Macaruso and colleagues, in three separate studies (Macaruso, Hook, & McCabe, 2006; Macaruso & Rodman, 2011; Macaruso & Walker, 2008), investigated the efficacy of three CAI reading programs produced by Lexia Learning Systems. Early Reading (Lexia Learning Systems, 2003) is for kindergarten students, Phonics Based Reading (Lexia Learning Systems, 2001) is for students in the early primary grades, and Strategies for Older Students (Lexia Learning Systems, 2001) is for students through adulthood. The programs are designed to provide intensive, structured, and systematic practice in learning and applying word–attack strategies with the goal of improving word recognition skills. Phonological awareness skills are taught in conjunction with decoding strategies. The activities in each of the programs make use of visual graphics and offer frequent opportunities for students to respond. The programs often require students to respond within a specified time limit, and the student's response is then followed by immediate feedback. Activities branch automatically based on the student's individual performance, reviewing when necessary and moving to more advanced items when easier ones have been mastered.

The Macaruso et al. (2006) study was sponsored by Lexia Learning Systems, and it appeared as a technical report on their website prior to its publication in *The Journal of Research in Reading*. The study compared the reading performance of first grade students using the Phonics Based Reading and Strategies for Older Students with control group students receiving similar classroom instruction without the use of CAI. Students from 10 first grade classrooms across five schools in an urban school district participated in this study. One class in each school was assigned to the treatment group and a second class to the control group. There were 83 students in the treatment group and 84 students in the control group. No students receiving special education services were included. Fifteen students receiving Title I supplementary reading services were included in each group. All students received daily reading instruction, which included explicit phonics instruction, from the same reading curriculum. The students in the treatment group received between two and four weekly CAI sessions lasting from 20 to 30 minutes for a total of six months. Students in the control group spent this time receiving regular classroom language arts instruction. Standardized reading measures were administered to all participants pre- and post-intervention. Results indicated that the post-test scores of the treatment group were greater, but not significantly greater, than the post test scores of the control group that did not receive CAI. When analyses were restricted to low-performing students eligible for Title I services, the treatment group obtained significantly higher post-test scores than the students who were eligible for Title I services in the control group. These results support the hypothesis that intensive phonics-based CAI can be beneficial for young students at risk for reading difficulties.

Macaruso and Walker (2008), in a continuation of Macaruso and his colleague's 2006 study, examined the benefits of drill and practice CAI designed to supplement regular, phonics-based reading instruction for kindergarten students in an urban public school system. Three kindergarten teachers and their six half-day kindergarten classes participated in the study. The classes were located in two schools. One class from each teacher was randomly assigned to the treatment group while the teacher's other class was assigned to the control group. Classes in the treatment group used the Early Reading

(Lexia Learning Systems, 2003) software for two or three, 15 to 20 minute sessions each week. Intervention sessions were conducted in the schools' computer labs under the supervision of the classroom teacher and computer lab staff. Teachers and lab staff were trained to implement the software. Students who completed at least 45 intervention sessions were included in the final data analysis. No student completed more than 62 intervention sessions during the six month intervention period. Students in the control group participated in typical classroom language arts instruction during the time that the treatment group received the intervention. While no significant differences between the treatment and control groups at post-test were found using DIBELS Phonemic Segmentation Fluency task, students in the treatment group significantly outperformed the control group on the Oral Language Concepts subtest of the *Gates-MacGinitie Reading Test*, which is also a measure of phonological awareness. While effect sizes for all students based on the *Gates-MacGinitie Reading Test* fell within the moderate range (.53), the effect size for the lowest performing students was 1.24, which is considered large.

The Macaruso and Rodman (2011) investigation consisted of two studies that extended the work of Macaruso and Walker (2008) by implementing Early Reading with preschool classes and a larger sample of low-performing kindergarten students within the context of daily classroom instruction. Teachers for these classrooms were trained in best practices for integrating supplemental reading software into their classrooms, and they assumed responsibility for implementing the CAI. Students worked on Early Reading independently after being introduced to the program by their teacher. Study 1 included students from 14 preschool classes in 3 schools. Because each of the seven teachers who participated in the study taught a morning class and an afternoon class, one of each teacher's classes was randomly assigned to the control group and the other was assigned to the treatment group. Students in the treatment classes used Early Reading within their classrooms for two to three sessions a week with each session lasting for 10 to 15 minutes during free choice or centers time. Results indicated significant gains for the treatment group as compared to the control group on a standardized test of reading skill with an associated effect size of .69. However, it should be noted that data from three of the seven pairs of classes originally included in the study were excluded from final data analysis because the students in the classes did not complete a minimum of 200 minutes of instruction in Early Reading. The final treatment and control groups included only 19 students each.

Study 2 investigated the effects of Early Reading, when it was provided to low performing Kindergarten students in addition to their traditional classroom instruction. Low performing students were described as students who scored one standard deviation below the mean on a standardized test of reading. Six kindergarten classes from two schools served as the treatment group and two kindergarten classes from a third school served as the control group. Students in the treatment classrooms who participated in less than 600 minutes of Early Reading were excluded from the treatment group, and this resulted in a treatment group that consisted of 47 students. The control group consisted of 19 students. Results indicated that while both groups made significant gains in reading as measured by a standardized test of reading, the students in the treatment group made significantly greater gains than students in the control group, with an associated effect size of .64.

The results of these studies, like the results from studies of DaisyQuest and Daisy's Dragon, suggest that CAI can be as effective as teacher instruction for improving the phonological awareness and early reading skills of kindergarten and first grade students. It also appears that CAI for reading intervention may be of particular benefit for students at risk for reading failure, even when the CAI is implemented without additional support from researchers.

Fast ForWord Language. Fast ForWord Language (Scientific Learning, 2002) attempts to develop oral language skills to create a foundation for reading. Fast ForWord Language differs from the other programs discussed in this section because it uses acoustically modified speech as an intervention tool. The speech component of the program adapts with the child's progress, so that the amount of speech modification decreases as the student becomes more successful. The program also incorporates other language training elements in an attempt to train multiple skills at the same time (Strong, Torgerson, Torgerson, & Hulme, 2010; Troia & Whitney, 2003). The program addresses four major areas of language acquisition: phonological awareness, listening comprehension, language structures, and sustained focus and attention (Rouse & Krueger, 2004). Although the programs were initially developed for children with spoken language disorders, the publishers suggest that Fast ForWord programs can also benefit children at risk for reading failure, and research conducted by Scientific Learning suggested positive outcomes for students who used the programs (Hook, Macaruso, & Jones, 2001). It is important to note, however, that these results were primarily from studies that did not utilize control groups and often had small sample sizes (Rouse & Krueger, 2004).

In one of the first independent investigations of Fast ForWord's impact on early reading skills, Hook et al. (2001) explored the effects of the program that would become Fast ForWord Language on the reading and oral language skills of children identified as having difficulties in the areas of phonemic awareness and word identification. The study design included three groups of children between the ages of seven and twelve. Two groups of children were treatment groups, and one group served as the control group. The first group, called the Fast ForWord group, consisted of 11 children. The children in the group were identified by means of a flyer posted in a major newspaper. The children in this group attended a computer lab for two hours a day, five days a week for up to two months during the summer to participate in the Fast ForWord program. Treatment was discontinued when the child successfully completed five of the seven activities in the program with 90% accuracy, or until an obvious plateau had been reached. Children varied in completion time from 22 to 44 days. The second group, which included 9 children, was identified as the Orton-Gillingham group. These students were chosen from a summer school for children with reading difficulties. They were matched to the Fast ForWord group based on age, full scale IQ, phonemic awareness ability, and reading level. The Orton-Gillingham group received an hour a day of one-on-one remediation five days a week for five weeks using Orton-Gillingham tutoring methods. The Orton-Gillingham method is a multisensory, structured language approach that explicitly and systematically teaches phonics skills. The third group was the control group, and it consisted of 11 students. This group of students was also matched to the Fast ForWord group. Each of the three groups participated in similar educational programs and curricula during the school year. Results indicated that both treatment groups made gains in phonemic awareness immediately after treatment but only the Orton-Gillingham group made gains in word attack skills (phonics). Neither of the treatment groups made significant gains in word identification (sight word reading) skills as compared to the control group. All students made gains in reading over the two years that the researchers followed them, but Fast ForWord *Language* did not result in additional or faster improvement as compared to the control group.

Troia and Whitney (2003) evaluated the efficacy of Fast ForWord Language on the academic performance of 37 children in the first through sixth grades using a pre-test, post-test, matched no-contact control group design. The 25 students in the in the Fast ForWord *Language* treatment condition and the 12 students in the control group were nominated by their teachers due to poor academic performance, and they had received academic support services through the school system. Students in the treatment group participated in the Fast ForWord Language program during the school day in pull out sessions that generally replaced the students' language arts instruction. Students participated in 100 minutes a day of program training five days a week for a minimum of four consecutive weeks. Results indicated that the Fast ForWord Language group made significant gains in oral language (effect size .53), but not in phonological awareness skills or basic reading skills as compared to the control group.

Loeb and colleagues (2009) also investigated the effects of Fast ForWord Language and two other interventions on the phonemic awareness and reading skills of children between the ages of six and eight using a quasi-experimental design. The 103 children in the study were a subset of a larger, randomized sample obtained for a related study. The subset of children included in this study displayed both language impairment and poor reading skills on standardized tests, had not participated in more than eight hours of computer-based language intervention, and were not enrolled in any other language intervention during the treatment phase of the study. The children were divided into four groups, and each group participated in an intervention activity for 1.5 hours a day, five days a week for six weeks. Group one was the Fast ForWord Language group. The second group received computer aided language intervention using a variety of commercially available software that did not use acoustical modification. The third group received individualized language interventions delivered by a speech language therapist, and the fourth group was an attention control group that played computer games focused on mathematics, social studies, and science. Each group except the attention control group received an intervention intended to improve their phonemic awareness skills. Study results indicated that immediately after the intervention, all groups except the attention control group had made significant gains in blending sounds in words as measured by standardized tests with an associated moderate effect size of .71. Long-term gains in sound blending six months after treatment, however, were not significantly different from the gains made by the attention control group, and none of the interventions led to significant changes in reading skills as measured by standardized tests. Overall, research results for Fast ForWord do not suggest that it is an effective early reading intervention.

Unpublished drill and practice CAI. Wild (2009) reported the results of a randomized control trial that directly compared the results of skills practice using CAI to the results of traditional instructional practice activities such as worksheets for teaching phonological awareness skills to beginning readers. One hundred twenty seven students in six schools were assigned to one of three groups. The six schools were randomly chosen from a list of schools in a school district in England. One class from each of five schools and two classes from another school participated in the research. Students within existing classes were randomly assigned to one of three groups. Each group was taught by the researcher two days a week for six weeks. Two intervention groups were taught phonological awareness skills using the same phonological awareness program. The treatment group completed practice exercises on a computer using software purchased as part of the phonological awareness program. The control group completed practice exercises using worksheets provided as part of the phonological awareness program. The third group completed a math program which included no explicit literacy instruction or CAI components. The children in each of the three groups were pre- and post-tested on phonological skills and their ability to apply those skills using standardized tests. Statistical analysis indicated a significant learning advantage for children in the computer-based practice group compared with each of the other groups, particularly in relation to phonological awareness with a modest but significant effect sizes ranging from .14 to .25 reported. The author suggested that the fact that the CAI condition provided students with corrective feedback on their performance after each response while the students who completed worksheets received performance feedback only after completing a worksheet that required numerous responses may have benefitted the students in the CAI condition. She also noted, however, that students of higher ability appeared to be annoyed by the fixed nature of the feedback received from the computer.

Overall, the research regarding drill and practice CAI reviewed in this section suggests that drill and practice CAI in reading may be effective for improving targeted reading skills of at risk students when they are used to supplement traditional teacher instruction. It appears important that the CAI provide explicit and systematic instruction in the specific skills that have been identified as important to early reading rather than attempting to remediate underlying processing deficits as a means to improve the phonemic awareness, phonics, and sight word reading skills of students.

Studies assessing the effectiveness of ILS.

ILS are instructional programs that move beyond drill and practice activities to provide sequential instruction for students, often across several grades, while keeping records of student progress. Most ILS programs provide instruction in the areas of reading and mathematics (Kulik, 2003). Early reading skills frequently taught by ILS include letter recognition, phonemic awareness, word-recognition and word-attack skills, vocabulary building, and text comprehension across a variety of different kinds of text such as fiction, nonfiction, and poetry (Desrochers & Gentry, 2004).

Early ILS programs. One early ILS, developed by Zhao and his colleagues (2000), was called Technology Enhanced Learning Environment of the Web (TELE-Web). It was created to complement the Early Literacy Project (as cited in Zhao et al., 2000), a curriculum designed for use with students with learning disabilities in primary grade classrooms. TELE-Web software allowed the teacher to input words and select reading activities that were best suited for a particular reader or a group of students. Thus, TELE-Web could be integrated into the teacher's reading curriculum, offering potentially greater benefits than software separate from the curriculum. The TELE-Web software used digitized speech dictation and feedback; word models or prompts, and context clues to develop word identification skills of students. When investigating the effectiveness of

TELE-Web, Englert, Zaho, Collings, & Romig (2005) found that 160 minutes of independent reading practice using the program effectively improved the word recognition performance of first grade students at risk of school failure as compared to a previous class that did not use the program but was taught by the same teacher. Students in the TELE-Web class who began the first grade with minimal reading skills showed more than 1.5 months of reading gain per month of instruction as measured on a standardized test of reading achievement, with a resulting effect size of 1.25. It is impossible, however, to attribute these gains solely to the TELE-Web program due to the fact that individual and group differences within the students and the class as a whole that might explain the differences in reading achievement cannot be ruled out.

Another early ILS prototype, Intellitools, was investigated by Howell, Erickson, Stanger and Wheaton (2000). They studied the use of Intellitools with first graders from a range of geographical areas within the United States who were at risk for or diagnosed with reading difficulties. Intellitools reading was based on a balanced approach to the teaching of reading, and it included reading connected text for comprehension, word study to build word identification and decoding skills using analytic phonics, and structured writing activities. These activities were built around anchor stories written for the program using decodable text. Use of the program for 30 minutes, four times a week resulted in significant and meaningful gains (effect sizes ranging from .66 to .85) in phonemic awareness and word reading for students from pretest to posttest after just 16 weeks of intervention. An important practical feature of the Intellitools program was that it required only 15 minutes of instruction initiated by an adult who did not have specialized training in reading instruction.

PLATO Beginning Reading for the Real World. The PLATO Beginning Reading for the Real World (PLATO; PLATO Learning Corporation, 2001) program was an early ILS that provided opportunities for students to learn and use letter-sound connections, context cues, comprehension, and reasoning skills. It was designed to supplement the standard curriculum within the school. In a technical paper for PLATO Learning, Inc., Foshay (2002) summarized the research base supporting PLATO through 2001, which included thirteen studies by independent evaluators, and concluded that PLATO had generally positive effects on student achievement. Specifically, the use of PLATO software by students for at least 30 instructional hours over the course of at least one school semester resulted in achievement gain effect sizes of up to two standard deviations on standardized tests. Quinn and Quinn (2002), who evaluated the implementation of the PLATO ILS during a summer school program for the PLATO Learning Corporation, reported “a generally positive correlation between the level of PLATO Elementary program use and posttest student achievement scores” (p. 11) for first grade remedial students after students had used the program for a total of less than six hours.

Bauseman, Cassady, Smith, and Stroud (2005), conducted an independent evaluation of the effectiveness of PLATO Beginning Reading for the Real World using a quasi-experimental pre-test/ post-test design. Urban kindergarten students from two schools within the same school system participated in the study. Students at one school served as the treatment group, and they used the PLATO for between five and six hours over the course of eight weeks during their center time. The students in the treatment group completed an average of 12 PLATO lessons on the computer. The students in the control group used alternative kindergarten literacy and math computer programs during their center time. Results indicated that the children in the treatment group, who used PLATO, outperformed control group children on standardized tests of phonological awareness, knowledge of print concepts, and listening comprehension. Effect sizes were large for phonological awareness (.47) and knowledge of print concepts (.43) and moderate for listening comprehension (.35). It is important to note, however, that two classes of students in the treatment group participated in a full day kindergarten program and two classes participated in alternate day kindergarten programs while only one class in the control school participated in full day kindergarten program and the other three classes participated in half-day programs. More students in the treatment group, therefore, participated in full day kindergarten, which provides an alternate explanation for their improved performance as compared to the control group.

Waterford Early Reading Program. The Waterford Early Reading Program (WERP; Pearson Digital, 2003) is a current example of an integrated learning system that provides computer-based instruction in reading. It adapts to each student's learning pace. It consists of a planned curriculum that integrates classroom-based assessments, instructional activities, and aligned materials to provide systematic instruction in the five reading essentials as defined by the NRP Report (2000). Activities are presented through a mixture of songs, interactive games, videos, and digital books. WERP allows teachers to select activities within the program to meet each student's current instructional needs. Alternatively, the program can be allowed to direct a student's instruction based on his or her responses to assessments within the program. Research on the effectiveness of WERP has been mixed. Several studies with kindergarten and first grade students suggest that use of WERP results in significant and meaningful gains in student reading achievement (Cassady & Smith, 2003, 2005; Powers & Price-Johnson, 2006; Tracey & Young, 2007), especially for students identified as at risk for reading failure (Cassady & Smith, 2005; Powers & Price-Johnson, 2006; Tracey & Young, 2006). Other studies (Campuzano et al., 2009; Dynarski et al., 2007; Patterson et al., 2003) failed to show that WERP resulted in reading achievement gains for kindergarten and first grade students.

Evidence for the effectiveness of WERP in early reading instruction when implemented across a school system comes from both Powers and Price-Johnson (2006) and Tracey and Young (2007). In their research report, Powers and Price-Johnson, who were employed by Creative Research Associates, a private company that provides educational support services, used quasi-experimental methods to investigate the effectiveness of WERP with at risk kindergarten students in the Tucson Unified School District. Their results indicated that while all kindergarten students made significant gains in reading, kindergarteners who used WERP for the amount of time recommended by the publisher made greater gains on DIBELS measures and the Arizona state test for reading achievement than the students who did not use WERP. Effect sizes ranged from low (.28) on the Arizona state test for reading to moderate (.56) on DIBELS Initial Sound Fluency in favor of the students who used WERP, after initial differences in pretest scores were controlled using ANCOVA. These gains were consistent across various subgroups including ethnicity, socioeconomic status, and gender. Tracey and Young (2007), in a study partially funded by Pearson Digital Learning, Inc. and published in a peer-reviewed journal, investigated a year-long implementation of WERP with 265 kindergarten children from an urban, high-risk, community using a pre-test, posttest design with control and treatment groups. Results obtained by comparing student gain scores on the Test of Early Reading Ability, Second Edition (TERA2) indicated that students who participated in the intervention significantly outperformed students in the non-intervention classrooms on the TERA2 and the *Waterford Reading Inventory*. Effect sizes associated with these results were not reported.

In two quasi-experimental studies, published in peer-reviewed journals, Cassady and Smith (2003, 2005) investigated the use of the WERP as a supplement to an existing balanced literacy curriculum with kindergarten (2003) and first grade students (2005). The kindergarten students participated in WERP for 20 minutes each day completing phonemic awareness and phonics activities selected for them by their teacher to meet their current instructional needs and correspond with classroom lessons. Study results indicated that students in the WERP school acquired phonological awareness skills significantly faster than students in the control group as measured by standardized tests, but the associated effect size of .16 was small. The control group was comprised of kindergarten students from a similar school that did not implement WERP or any other ILS program. The students in the treatment group maintained their advantage through the end of the school year. First grade students (Cassady, 2005) participated in WERP for 20 minutes each day completing phonemic awareness and phonics activities selected for them by their teacher to meet their current instructional needs and correspond with classroom lessons. When compared to students from the previous year at their school, first grade students who used WERP made meaningfully greater gains (effect size = .43) in reading achievement as measured by standardized tests. The gains were greatest for students who demonstrated the weakest initial reading skills.

In contrast, results of three independent studies found that WERP was not any more effective than typical classroom instruction for improving student reading skills. Patterson et al. (2003) investigated the implementation of WERP using mixed methodology in an urban school system. The quantitative part of their analysis was a quasi-experimental investigation that compared eight existing classrooms- seven kindergarten and one first grade- that implemented WERP with eight existing classrooms- seven kindergarten and one first grade that did not implement WERP or any other type of ILS. The participating schools were matched in terms of overall socioeconomic level and teaching style, but only the classes with the greatest number of at risk students received the WERP intervention. The researchers did not provide any implementation support beyond that typically provided by Pearson Digital, Inc. Results indicated that WERP had no significant impact on kindergarten and first grade students' emergent reading skill development. Two additional studies (Campuzano et al., 2009; Dynarski et al., 2007) included first graders from numerous school districts located in various geographical parts of the United States and reported no meaningful effects as measured by standard and local tests of reading as a result of first grade students using WERP for a full school year (Dynarski et al., 2007). In the study, however, approximately one-third of teachers did not feel qualified or ready to use the technology after the single day of professional development training that was provided several weeks before the intervention. Details on the ways the technology was implemented in the classroom were not reported. A follow-up study, conducted with the same teachers after they had a year of prior experience implementing the ILS (Campuzano et al., 2009) also failed to show meaningful effects on standardized and local tests of reading. Overall, the effectiveness of the WERP program has not been consistently demonstrated in the literature. It is possible that WERP is an effective program that has not been consistently implemented as intended by the publisher by teachers when they are not supported by research personnel. It is equally possible that WERP has no meaningful effects on the reading skills of young students beyond that provided by typical classroom instruction.

ABRACADABRA. Unlike most other ILS programs, which must be purchased, ABRACADABRA is a free access, web-based literacy tool developed by the Centre for the Study of Learning and Performance in Canada (Hipps et al., n.d.). The program was designed to supplement instruction provided by classroom teachers. It uses a balanced curriculum that includes texts and strategies designed to support phonics, word reading, reading and listening comprehension, and reading fluency. It has a modular design that allows instruction to be customized to individual student needs. Program content was developed utilizing evidence from systematic research reviews of effective reading interventions for phonics and letter skills, reading fluency, and reading comprehension (Hipps et al., n. d.). Evidence for the effectiveness of the ABRACADABRA program, collected by the developers of the program, generally supports its effectiveness as an adjunct to classroom instruction. For example, a pilot study conducted in two Montreal area schools demonstrated moderate effect sizes in the following areas: decoding skills, processing speed, word reading, and text comprehension for students who participated in ABRACADABRA as compared to control group students who received only classroom instruction (Hipps et al., n.d.). Other researchers have found positive effects of medium to large magnitude for kindergarten students in letter sound knowledge, blending skills and segmenting skills (Comaskey et al., 2009) and first grade students in letter-sound knowledge, phonological awareness, listening comprehension, and reading comprehension skills (Savage, Abrami, Hipps & Deault, 2009).

Comaskey and colleagues (2009) conducted a pre-test, post-test experimental study that compared the effects of synthetic versus analytic phonics instruction using the ABRACADABRA program in kindergarten classrooms. Fifty-three children from two different kindergarten classes in the same school were randomly assigned to either a synthetic or analytic phonics intervention group. The students in each group used the ABRACADABRA program in small groups three times a week for 10 to 15 minutes for up to 16 weeks. All children participated in a total of 40 sessions resulting in 10 hours of instructional time per child. The analytic phonics group's core activities revolved around word families, identifying words that rhymed, and manipulating and articulating words at the onset–rime level. The synthetic phonics group's core activities focused on blending and segmenting simple two-phoneme words, identifying words with shared initial and final consonants and forming new words by blending single phonemes. In general, all synthetic phonics groups mastered the consonant-vowel and vowel-consonant levels of their core blending activities. All analytic phonics groups showed rime word generation skills. Both groups demonstrated gains in reading as measured by standardized tests. Because all students in the study used the ABRACADABRA program, no conclusions regarding the program's effectiveness in contrast to typical classroom instruction can be drawn.

Savage, Abrami, Hipps and Deault (2009) investigated the effectiveness of ABRACADABRA for first grade children in a study with a randomized control trial pre-test, post-test design. One hundred forty-four first grade students from 13 classes participated in the study. Students in each classroom were randomly assigned to one of three groups: a synthetic phonics CAI group, an analytic phonics CAI group, or a teacher-led, balanced instruction group. The teacher-led balanced instruction group served as the control group. Each of the ABRACADABRA groups participated in 20 minutes of computer based instruction in word analysis, text comprehension, and fluency using ABRACADABRA four times a week in groups of four students. The synthetic phonics intervention was focused on developing students' skills at blending and segmenting words at the level of the individual phoneme unit. The analytic phonics intervention taught students to identify shared rimes in word families and to attend to familiar spelling patterns when reading words. Results, as measured by standardized tests, at post-test indicated moderate effect sizes for blending words (.59 to .70) across intervention conditions that remained apparent, but smaller in magnitude (.20 to .33) at delayed post-test seven months later. Word attack skills and reading comprehension skills test scores indicated small effect sizes (.22 to .37) at immediate posttest across interventions. These small effect sizes (.16 to .20) were still apparent at delayed post-test for the analytic phonics intervention group but not for the synthetic phonics intervention group. When the synthetic and analytic phonics interventions were considered together at post-test and delayed post-test as an overall measure of the effectiveness of CAI for reading, the mean effect size for all standard scores at immediate posttest was small at .23. The mean effect size for all synthetic phonics standard scores was .17, and the mean effect size for all analytic phonics standard scores was .18. Both of these effect sizes are small. The authors also calculated the effect sizes for the control group across post- and delayed posttest. The mean effect size for all control standard scores was -.02, which indicates that the children in the control group made expected progress in reading with regular classroom teaching during the intervention.

In a quasi-experimental pre-test, post-test design, Savage, Erten, Abrami, Hipps, Comaskey, and van Lierop (2010) investigated the effectiveness of ABRACADABRA with first grade students when it was implemented by classroom teachers after they received one day of ABRACADABRA training. Four first grade teachers from three schools were randomly assigned to either the treatment or control condition. Each teacher in the treatment condition agreed to implement the ABRACADABRA program for a minimum of 2 hours a week for 8 consecutive weeks. The teachers who were implementing the ABRACADABRA program also had access to a program facilitator for the first four weeks of the study. The program facilitator did not provide any direct instruction to students. The teacher in the control condition taught the regular literacy program based on a balanced literacy approach. No additional literacy intervention of any type was used in the control classroom. Sixty students participated in the research. Results indicate that the ABRACADABRA web-based literacy program can produce significant growth in standard measures of reading ability (mean effect size = .48), when used effectively by teachers. Additionally, teacher variation in the implementation of ABRACADABRA had a substantial impact on student learning outcomes, as indicated by the wide range, from .07 to .48, of effect sizes in student reading gains across teacher skill in the implementation of the ABRACADABRA program. The mean effect size for the control group was .12. Overall, ABRACADABRA appears to be a promising, no-cost example of an ILS for early reading instruction.

Headsprout Early Reading. Headsprout (Headsprout Corporation, 2005) is an internet based, ILS that uses behavioral principles (Layng et al., 2004a) to teach early reading skills to students in pre-kindergarten through second grade who are not yet reading or who are in the beginning stages of the reading process (FCRR, 2003; Layng et al., 2004a; Twyman, Layng, Strikeleather & Hobbins, 2004). Headsprout uses a carefully planned and researched sequence of instruction to provide direct instruction in the five areas identified by the NRP (2000) as critical for reading acquisition, which are phonemic awareness, phonics, fluency, vocabulary, and comprehension (FCRR, 2003). Headsprout teaches phonemic awareness through learning routines (Layng et al., 2003). These learning routines require students to hear letter sounds and match them with letters, say sounds out loud and choose the character that "said the sound just like you did" (p. 3), combine sounds and hear them slowly blended, say sounds slowly blended, and hear the sounds said fast as whole words. Headsprout teaches phonics through the introduction of 84 phonetic elements, most of which maintain consistent pronunciation across 85% of the words in which they appear. For example, the first sounds taught in Headsprout are /ee/, /v/, /cl/ and /an/, which generally sound the same across most words in which they appear. Students are taught to combine these sound elements to make words. They are also taught that some sounds have other sounds inside them and that sound units can be combined to make new sounds. Headsprout instruction is designed to produce learners who can reliably use these insights to read unfamiliar words. Students are also taught to sound out words in isolation, as parts of sentences, and when reading stories with words they have not been taught directly.

In order to teach reading vocabulary, the first words that students learn to decode in Headsprout are words that are likely to be in their spoken vocabularies (Layng et al., 2003). By the time that the sounding out skills and all 84 Headsprout sound elements have been taught, typical students have been exposed to a reading vocabulary of 5000 words in 30 hours of instruction. Instruction in reading fluency begins as students learn their first sounds. After a sound is taught, the student completes a task that requires quick identification of the sound. By episode four, students should be building fluency on words made up of the sounds they have learned, and by episode five students should be reading their first story. From as early as episode five, students should be learning that the sentences they read have meaning.

Comprehension is verified through the use of increasingly complex question and answer routines (Layng et al., 2003). For example, early episodes require the student to choose one of three pictures that go with a sentence after each reading exercise. In later episodes, students are required to select pictures that represent the meaning of whole stories, construct meaning by building sentences that result in an animated picture that represents the sentence, express meaning by building sentences that describe a picture, complete sentences that describe a picture by selecting a missing word from among four choices, and read a text passage and select the best answer to a written question about the passage from among three written choices (Layng et al., 2003).

Initial research data generated by the Headsprout Corporation during program development and early implementation suggest that most children who work with the Headsprout program as recommended by the publisher acquire the specific skills it is designed to teach. These skills are phonemic awareness, phonics, vocabulary, fluency, and comprehension (Layng et al., 2003; FCRR, 2003; Headsprout, 2007). Research authored by Headsprout (Layng et al., 2003; Layng et al., 2004b; Headsprout, 2007) suggests that students who use Headsprout as a supplement to their regular reading instruction make significantly more progress in reading than those who do not use Headsprout even when the amount of time spent in reading instructional activities is held constant.

For example, results from investigations conducted by the Headsprout Corporation indicate that kindergarten and first grade students who were at risk for reading failure in New York City who completed at least 70 of the 80 Headsprout episodes as part of their classroom literacy instruction made significant and substantial gains in reading as measured by the Letter Word Identification Subtest of the *Woodcock Johnson, Third Edition* and the reading subtests of the *Iowa Test of Basic Skills* over those made by students in classrooms that did not use Headsprout (Headsprout, 2007). Kindergarten students in Los Angeles classrooms that used Headsprout as part of their literacy instruction and completed at least 6 episodes of the Headsprout program also significantly and meaningfully outperformed kindergarteners in classrooms that did not use Headsprout as part of their literacy curriculum as measured by the Reading Total score of the *Gates-MacGinitie Reading Test* (Headsprout, 2007). Research conducted by Headsprout with students identified as having special needs indicates that these students are able to maintain a correct response rate of 90% across all Headsprout episodes, which is the minimum accuracy rate suggested by Headsprout as necessary for successful completion of the program (Headsprout, 2005). Research also suggests that students with disabilities who participate in Headsprout demonstrate increased on task behavior and improved oral reading fluency (Headsprout, 2005; Clarfield & Stoner, 2005).

Evidence in research not authored by the Headsprout Corporation regarding outcomes for students who participate in Headsprout, however, is inconclusive. Using a multiple baseline design, Clarfield and Stoner (2005) found that three students with ADHD made greater gains in reading fluency when participating in Headsprout in addition to their regular reading instruction than when they received their regular reading instruction alone. In her dissertation research, Clarfield (2006) found that typical students (N = 18) and at risk kindergarten students (at risk in reading N = 9, at risk in behavior N = 8, at risk in both reading and behavior N = 9) who participated in Headsprout significantly outperformed those who participated in Lexia (typical students N = 16, at in reading students N =11, at risk in behavior students N = 8, and at risk in both reading and behavior N = 9), another ILS for reading, for the same amount of time on measures of early reading skill. Effect sizes were small on both the DIBELS Phonemic Segmentation Fluency (.06) and the DIBELS Nonsense Word Fluency, ranging from .06 to .18.

In contrast, other researchers (Dynarski et al., 2007; Campuzano et al., 2009) found that use of the Headsprout program, as well as several other ILS for reading, had no effect on student achievement in reading. Dynarski and his colleagues (2007) reported on the effectiveness of five ILSs designed to improve reading achievement in first grade students. Using hierarchical linear modeling, which allowed the researchers to compare effects at the classroom, school and district levels, the researchers found that none of the studied ILSs, including Headsprout, resulted in improved the reading performance of first grade students (treatment group N = 1,516) beyond that achieved by control group (N = 1,103) students in any of the 43 schools studied.

In the second year of the study conducted by Dynarski and colleagues (2007), Campuzano and colleagues (2009) again investigated the effectiveness of four of the original five computer programs, including Headsprout, after teachers had used the products for a year. The researchers determined that first grade student scores on standardized reading tests did not change as a result of teacher familiarity with the programs by calculating the difference between the second-year product effect on test scores and the first-year product effect on test scores. The product effect score is the difference in spring student test scores between treatment and control classrooms caused by the assignment of treatment classrooms to use a software product. In fact, the amount of reading instructional time devoted to software use by the students actually decreased by more than 50% (statistically significant decrease in use). Campuzano et al. (2009) did not observe classrooms or interview teachers, so no additional information about how or why the teachers' use of the programs changed was available. Use of the programs did not improve first grade student reading scores as compared to the control group.

In summary, this chapter reviewed literature related to early reading success, RTI, and CAI. The research demonstrates that early reading failure is a significant problem, and suggests that RTI processes may be an effective means for addressing the problem. There are significant potential barriers to the effective and successful, widespread implementation of RTI, including scarcity of the resources in terms of personnel, time, training and materials required to effectively implement RTI processes. CAI, particularly in the area of early reading instruction and intervention, provides a potential method for implementing the explicit and intensive instruction that students at risk for reading failure require in a cost effective manner. Evidence from CAI implemented in school settings that received substantial support in terms of personnel, software and hardware, and professional development often, but not always, report significant and meaningful benefits to students, especially at risk students, from the use of CAI in conjunction with traditional instruction. Few studies, however, have investigated the effectiveness of CAI when it is implemented as part of the school curriculum without the benefit of additional resources from outside entities, and the results from the available studies are inconclusive (Campuzano, et al. 2009; Dynarski et al., 2007; Savage et al., 2010). Further research investigating the effectiveness of CAI as it is implemented within the day to day operations of schools is warranted.

CHAPTER 3

METHOD

The purpose of this chapter is to describe the research design and method that were used for this quasi-experimental study. First, this chapter discusses the research questions and research design for the study. Next, the selection of participants and the existing data set are described. The chapter continues with a description of the instrumentation used for the study. Finally, an explanation of the statistical procedures that were proposed for the study is provided.

Research Questions

Research Question 1

Do Kindergarten students at risk for reading failure who participate in Headsprout in addition to their regular classroom reading instruction demonstrate different skill levels in reading as measured by the DIBELS LNF and NWF tasks than kindergarten students at risk for reading failure who do not participate in Headsprout?

Research Question 2

Do first grade students at risk for reading failure who participate in Headsprout in addition to their regular classroom instruction demonstrate different skill levels in reading as measured by the DIBELS ORF task or the CRCT than first grade students at risk for reading failure who do not participate in Headsprout?

Research Design

The research design for this study was a modified pretest, posttest quasi-experimental design. All kindergarten and first grade students within the school system identified as at risk for reading failure using the DIBELS were considered for inclusion in both the treatment and control groups. All kindergarten and first grade students were screened to assess their literacy skills at the beginning, middle, and end of the school year using DIBELS. Kindergarten students were considered for inclusion in the study if they demonstrated deficits in early reading skill acquisition as measured by the LNF (any score below 8) or ISF (any score below 8) subtests of the DIBELS in the fall. First grade students were considered for inclusion in the study if they demonstrated deficits in early reading skill acquisition as measured by the LNF (any score below 37) or the NWF (any score below 24) subtests of the DIBELS in the fall. The students at greatest risk for reading failure, as indicated by DIBELS performance and teacher judgment, were assigned to participate in Headsprout. The treatment group in this study consisted of students assigned to participate in Headsprout. Participation in Headsprout was the independent variable.

Kindergarteners' fall DIBELS ISF and LNF scores and first graders' fall DIBELS NWF scores served as the pretest measures in the study. The dependent variables for kindergarten students were the scores from the spring administration of the DIBELS LNF and NWF subtests. The dependent variables for first grade students were the score from the spring administration of the DIBELS ORF subtest and the score from the reading portion of the CRCT administered in the spring of 2009.

Participants

District. The data pool for this study was an existing data set collected during the 2008-2009 school year by a suburban school district located within a major metropolitan area in the southeastern United States. Census bureau population estimate for the county for 2008 was 122,924, and per capita income for 2007 was $30,377. The district served a total enrollment of 21,525 students in grades pre-kindergarten through twelve during the year the data were collected. The overall student population within the district was 68% white, 38% black, 6% Hispanic, 3% multiracial, and 1% Asian. Nine students within the district qualified for services through the Migrant Education Program. Additionally, 24.1% of the student population received early intervention (grades K-5) or remedial education services (grades 6-12), 10.9% of the student population was enrolled in special education, and 1.7% of the student population received services through the English to Speakers of Other Languages (ESOL) program. Ten of the 18 elementary schools in the district qualified for Federal Title I services. Two elementary schools, both of which received Title I funding, did not make adequate yearly progress (AYP) for the previous school year. These two schools did not make AYP because they failed to increase the number of their students within the subgroup of students identified as black who met the state standard (as measured by the CRCT) in mathematics. Both schools met all other AYP indicators. Georgia CRCT results from the spring of 2008 indicate that among first grade students within the school district, 95% of white students, 80% of black students, 87% of Hispanic students, 90% of multiracial students and 96% of Asian students met or exceeded expectations in reading on the CRCT. Additionally, 68% percent of first grade students with disabilities and 73% of students with limited proficiency in English met or exceeded expectations in reading on the CRCT (Governor’s Office of Student Achievement, 2009).

Students. The potential sample for this study included 1688 kindergarten students from 91 classrooms and 1604 first grade students from 90 classrooms. These classrooms and students represent all of the kindergarten and first grade classes within the school district during the 2008-2009 school year. Specific demographic information by school for the schools included in the final data set is presented in Table 1.

Kindergarten Students. At the school level, 7 schools were eliminated from the kindergarten data set because no kindergarten students in those schools participated in Headsprout. In the remaining schools, 463 kindergarten students were identified as at risk for reading failure based on their performance on the DIBELS Letter Naming Fluency (LNF) and Initial Sound Fluency (ISF) benchmarks administered in the fall of 2008. Of the kindergarten students identified as at risk for reading failure, 61 students were not included in the data set because they were not enrolled in the school system for the entire study year. Twenty-two students were not included in the data set because they were repeating kindergarten. Upon checking the data set for outliers, seven additional control group students were removed from the data because visual inspection of the data suggested that they were not best described as students at risk for reading failure because one of their pretest scores was much higher than those of the other control group students. Additionally, all winter DIBELS scores for these students fell within the DIBELS low risk category. The remaining 373 students were considered for inclusion in the study, and specific demographic information for these students is presented in Table 2. The kindergarten teachers at each school met as a grade level and assigned each student identified as at risk for reading failure to either participate in Headsprout or not participate in Headsprout based on the student's fall DIBELS scores and the teachers' judgment of the student's reading skill. The resulting treatment group of students who participated in Headsprout included the students at greatest risk for reading failure based on the available data.

Table 1

Demographic Information by School

School	1	2	3	4	5	6	7
Total Enrollment	625	512	423	801	543	350	349
% White	79	75	52	78	83	78	32
% Black	10	15	35	9	11	18	36
% Hispanic	5	4	8	7	1	1	19
% Asian	1	0	1	2	1	0	3
% Native American/ Alaskan Native	1	1	0	0	1	0	0
% Multiracial	4	4	3	4	3	4	9
% Male	49	48	51	52	48	50	52
% Female	51	52	49	48	52	40	48
K Enrollment-Fall	109	89	64	123	84	61	56
K Enrollment- Spring	113	87	68	118	81	63	59
1st Enrollment-Fall	101	71	68	111	100	46	55
1st Enrollment-Spring	96	69	69	126	95	43	61
Title I Status	No	Yes	Yes	No	No	Yes	Yes
% Economically Disadvantaged	24	53	66	22	40	54	70
% Scored $\geq$ 800 on CRCT Reading	98	87	86	92	96	86	81
Special Education	8.3	8.2	7.8	5.6	9.8	12.9	12.6
EIP	13.6	18.2	13.5	13.1	21.2	24.6	30.7
ESOL	1.4	2.1	5.4	3.9	0	0	8.6

School	8	9	10	11	12	13	14
Total Enrollment	436	779	684	443	862	431	754
% White	79	51	72	87	49	64	79
% Black	13	30	22	5	35	27	8
% Hispanic	6	8	3	2	8	3	8
% Asian	0	5	0	1	3	0	1
% Native American/Alaskan Native	0	1	1	0	1	0	0
% Multi	3	6	3	3	5	6	4
% Male	50	50	53	53	51	53	53
% Female	50	50	47	47	49	47	47
K Enrollment-Fall	76	140	117	64	173	69	118
K Enrollment- Spring	79	139	117	63	178	67	123
1st Enrollment-Fall	66	146	109	68	140	67	125
1st Enrollment-Spring	69	146	111	73	136	65	130
Title I Status	Yes	Yes	No	No	Yes	Yes	No
% Economically Disadvantaged	45	39	41	15	42	54	27
% Scored $\geq$ 800 on CRCT Reading	91	98	91	100	85	94	92
Special Education	7.8	8.3	11.3	6.3	9.3	10	6.6
EIP	14.9	13.4	13.2	14.4	19.3	19.7	13
ESOL	5	4.7	0	0	4.2	0	4.5

Table 2

Demographic Information for Students At Risk for Reading Failure

	Kindergarten		First Grade	
	Headsprout	Non-Headsprout	Headsprout	Non-Headsprout
Number of Students	130	243	181	178
Mean Age in Months	65.60	65.34	78.06	77.67
Age Std. Deviation	3.37	3.49	3.68	3.68
% Girls	41.50	46.40	38.70	44.40
% Boys	58.50	53.60	61.30	55.60
Mean Fall LNF Score	6.02	12.82	34.69	35.14
Fall LNF Std. Deviation	8.1X	11.84	11.90	11.76
Mean Fall ISF Score	5.07	5.86		
Fall ISF Std. Dev.	3.93	3.43		
Mean Fall NWF Score			15.74	22.10
Fall NWF Std. Dev.			7.30	11.57

First Grade Students. In the first grade data set, one school was eliminated from the first grade data set because DIBELS measures were not administered to first grade students. Three schools were eliminated from the first grade data set because no first grade students from those schools participated in Headsprout. Across the remaining schools, 470 first grade students were identified as at risk for reading failure based on their performance on the DIBELS Letter Naming Fluency (LNF) and Nonsense Word Fluency (NWF) benchmarks administered in the fall of 2008. Of the first grade students identified as at risk for reading failure, 47 first graders were not included in the data set because they were not enrolled in the school system for the entire study year. Thirty-nine first graders were not included in the data set because they had either repeated kindergarten or were currently repeating the first grade. Twenty-five first graders were excluded from the data set because they had participated in the Headsprout program as kindergarten students. The remaining 359 first graders were considered for inclusion in the study, and specific demographic information for these students is presented in Table 2.

Reading instruction. All students engaged in daily reading instruction. The typical language arts block of instruction for kindergarten and first grade students consisted of 135 minutes of instruction. This block of time was subdivided into 30 minutes of word work, which included phonics instruction and vocabulary practice; 60 minutes of whole group instruction using the basal reader or other children's literature; and 45 minutes of flexible group instruction. Flexible group instruction was designed to provide students with teacher-led instruction targeting specific skills. Some schools assigned students to flexible groups within their classrooms while other schools assigned students to flexible groups across classrooms. Implementation of flexible group instruction within a classroom generally involved the use of centers or stations so that while the teacher was providing small group, targeted instruction to one group of students, other groups of students were involved in independent reading tasks such as listening to books on tape, writing or copying words or sentences from a model or prompt, or playing phonics games. When students were assigned to flexible groups across classrooms, each instructional group was somewhat larger, but a longer period of time was available for teacher led instruction.

Reading Curriculum. The system-wide basal reading series in use at the time of data collection was *Harcourt Trophies* (Beck et al., 2003). Harcourt Publishers describe their *Trophies Basal Reading Series* as a research-based, developmental reading/language arts program. According to the publisher, *Harcourt Trophies* includes, "explicit phonics instruction, direct reading instruction, guided reading strategies, phonemic awareness instruction, systematic intervention strategies, integrated language arts components, and state-of-the-art assessment tools to ensure every student successfully learns to read." (Houghton Mifflin Harcourt School Publishers, n.d.). An independent review of *Harcourt Trophies* completed by the Oregon Reading First Center (2006) found that *Harcourt Trophies* provided instruction in each of the five critical elements of effective reading instruction. Rankings of *Harcourt Trophies* inclusion of instruction in the five critical elements of early reading instruction ranged from 75% (kindergarten phonics instruction) to 100% (first grade phonemic awareness instruction). *Harcourt Trophies* has been approved by several states (e.g. Vermont, Louisiana, Virginia) for use in their Reading First schools. Each elementary school within the school district also had the option, at the discretion of the faculty, to supplement the core reading curriculum supplied by the school system with materials of their choice. The range of materials that various schools used to supplement the basal reading series during the language arts instructional block included *Saxon Phonics*; *Orton-Gillingham Phonics*; *Sing, Spell, Read, and Write*; and *Animated Literacy* (see Table 3).

Table 3

Supplemental Instructional Programs & Headsprout Implementation Method

School	Kindergarten	First Grade	Head Sprout
1	Orton Gillingham	Saxon Phonics Orton Gillingham	Classroom teachers implemented Headsprout during their regularly scheduled computer lab time.
2	Sing, Spell, Read, & Write	Sing, Spell, Read, & Write	The Early Intervention Program (EIP) teacher implemented Headsprout in the computer lab at a scheduled time within the school day.
3	Did not participate in Headsprout.	None	A paraprofessional implemented Headsprout in the computer lab before the start of the school day.
4	Orton Gillingham	Orton Gillingham	The school counselor and a kindergarten paraprofessional implemented Headsprout in the computer lab at a scheduled time during the school day.
5	Animated Literacy	Animated Literacy	The EIP teacher implemented Headsprout in the computer lab during EIP time.
6	Saxon Phonics	Saxon Phonics	A paraprofessional implemented Headsprout in the computer lab at a scheduled time within the school day. The Literacy Coach completed benchmarks with students.

School	Kindergarten	First Grade	Head Sprout
7	Orton Gillingham	Orton Gillingham	A paraprofessional implemented Headsprout in the computer lab at a scheduled time within the school day. The paraprofessional supervising the program rotated from week to week so that the same person did not monitor the program across the year.
8	Saxon Phonics	Saxon Phonics	The paraprofessional assigned to each kindergarten classroom implemented Headsprout with the students assigned to the class. The EIP teacher implemented Headsprout with first grade students at a scheduled time within the school day
9	Did not participate in Headsprout.	Orton Gillingham	A paraprofessional implemented Headsprout in the Computer Lab at a scheduled time during the school day.
10	Orton Gillingham Sing, Spell, Read & Write	None	A paraprofessional implemented Headsprout with kindergarten students in the Computer Lab at a scheduled time during the school day; The first grade classroom teacher implemented Headsprout within the classroom.
11	Orton Gillingham	Orton Gillingham	A paraprofessional implemented Headsprout in the Computer Lab at a scheduled time during the school day.
12	Orton Gillingham	Orton Gillingham	EIP teachers implemented Headsprout in the computer lab at a scheduled time within the school day.

School	Kindergarten	First Grade	Head Sprout
13	Did not participate in Headsprout.	Saxon Phonics	The speech language pathologist and a special education paraprofessional implemented Headsprout in the computer lab before school.
14	Saxon Phonics Orton Gillingham	Saxon Phonics Orton Gillingham	A paraprofessional implemented Headsprout in the Computer Lab at a scheduled time within the school day.

Instrumentation

Headsprout Early Reading

Headsprout is an internet-based, supplemental reading program for students in pre-kindergarten through second grade who are not yet reading or who are in the beginning stages of the reading process. Headsprout uses one-on-one, computer based instruction to teach the alphabetic principle, the use of sound elements to decode words, print awareness, vocabulary, and comprehension (Florida Center for Reading Research [FCRR], 2003; Layng et al., 2004b). Headsprout was available as a standard protocol, Tier 2 intervention for elementary aged students at risk for reading failure at each school in the system during the 2008 -2009 school year. School personnel were encouraged by central office administrators to implement the program with kindergarten and first grade students at risk for reading failure. Headsprout was chosen for implementation by the school system because it employs explicit instruction in phonemic awareness, phonics, vocabulary, fluency, and reading comprehension. The Headsprout Corporation also guaranteed that every kindergarten or first grade student who completed the 80 episodes included in Headsprout would be reading at grade level. The Headsprout Corporation promised to refund the price paid for any kindergarten or first grade student who was not reading at grade level after completing the program (Headsprout, 2007). Students were chosen to participate in Headsprout based primarily on their performance on the DIBELS in the fall. Table 2 provides demographic information about the kindergarten and first grade students who participated in Headsprout.

Headsprout introduces letters and sounds following a prescribed scope and sequence, provides fluency building exercises, and teaches segmenting and blending strategies. The program also provides explicit instruction in the area of sight word vocabulary development and recognizing and using punctuation cues to aid comprehension. Each episode adapts to the individual student's learning rate as the student progresses through the program. That is, students who demonstrate mastery of a presented skill move on to the next activity, but students who demonstrate the need for addition practice in a skill by answering questions incorrectly, receive additional practice. Headsprout is designed so that each student experiences a success rate of at least 90% in each episode. A majority of the Headsprout activities involve the student completing tasks, which then results in an animated character moving toward a desired destination. After completing a set of six episodes, the student receives a Headsprout reader, which is a colorful story booklet that contains the sounds and words that the student has learned through the program. The Headsprout reader also serves as a benchmark of student progress through the program. The reader is designed to be read aloud to an adult who scores the student's reading based on Headsprout guidelines and manually enters the score into the program.

Based on the benchmark score, teachers can also have a student who has not mastered the content within a series of episodes repeat those episodes before beginning the next set of six episodes. Feedback is interspersed within the Headsprout program, as every student response is acknowledged with feedback, encouragement, and correction if necessary. For example, after each correct response, the computer tells the student, "yeah" or "you did it". The program also provides brief (10 to 30 second) humorous movies to entertain students between activities. Each animated episode lasts approximately 20 minutes, and Headsprout suggests that each student complete three 20-minute sessions each week (FCRR, 2004; Layng, et al. 2004b).

The 80 episodes provided through Headsprout Early Reading are subdivided into sets of episodes that represent kindergarten content, first grade content, and beginning second grade content. All students start with the first episode in the program. Episodes 1 through 23 are collectively referred to as Cracking the Code. These episodes teach early reading skills and are designed to prepare students to independently sound out words. Episodes 24 through 40 are called Make Sense Out of Reading. Episode 24 introduces independent sounding out of words to students. After the successful completion of episode 40, a student should have achieved an early first grade reading level. Episodes 41 through 56 are referred to as Accelerate and Diversify. Beginning with Episode 41, the program places increased emphasis on the development of vocabulary, fluency and comprehension. The final group of episodes is called Reading for Meaning and Enjoyment.

Headsprout states that after successfully completing Episode 80, a student should be reading at the mid-second grade level and have a reading vocabulary of 5000 words. In general, the Headsprout Corporation recommends that kindergarten students complete through Episode 40 by the end of a school year and first grade students complete through Episode 80 by the end of a school year. For second grade students, Headsprout recommends that the student complete all 80 episodes by midyear (P. Clayton, personal communication, 2010).

All elementary schools within the district were provided with access to Headsprout, and at least one administrator or teacher from each elementary school attended the implementation training provided by the Headsprout Corporation. This training consisted of a 2 1/2 hour workshop provided at the district office. The primary focus of this training was administrative issues related to entering students into the program data base and managing their records. Additional technical support was provided by Headsprout via telephone and in person by district employed instructional technology staff. Personnel at each school were allowed to choose the method by which the school would deliver Headsprout to students based on their evaluation of the resources at hand. No school received additional support for the implementation of Headsprout in any form (e.g., additional computer hardware, additional personnel). The method of implementation chosen by each school included in the study is described in Table 3. Students who participated in Headsprout spent more time in reading instructional activities than did the students who did not participate in Headsprout.

Headsprout recommends that a student spend a minimum 20 minutes, three times a week using the Headsprout software until all 80 episodes have been completed

(Headsprout, n.d.). Following this recommendation from Headsprout would mean that a student who completed Headsprout would have received a minimum of approximately 26 hours of reading instruction beyond that provided by the typical curriculum, depending on how quickly the student completed each Headsprout episode. The average amount of time that kindergarten students spent participating in Headsprout was 7.25 hours, but participation ranged from a low of 30 minutes to a high of 21.38 hours, with a standard deviation of 4.75 hours. The average amount of time spent that first grade students spent participating in Headsprout was 14.5 hours, but participation ranged from a low of 5 hours to a high of 29.75 hours, with a standard deviation of 14.47 hours.

Measures

Dynamic Indicators of Basic Early Literacy Skills, Sixth Edition (DIBELS, Good & Kaminski, 2002). DIBELS benchmarks are a set of brief, individually administered pre-reading and early reading tests designed to assess phonological awareness, alphabetic awareness, and fluency in children. DIBELS was developed from the principles of curriculum-based measurement (CBM). Like CBM, DIBELS was developed to be economical and efficient indicator of a student's progress toward achieving an important learning outcome. Unlike initial forms of CBM which were linked to a specific curriculum, DIBELS measures are generic and draw their content from sources other than a specific school's curriculum. The use of CBM methods without the link to a specific curriculum is referred to as General Outcome Measurement (GOM), and it is designed to measure the progress students are making toward a long term goal, such as reading at grade level (Fuchs & Deno, 1994).

DIBELS benchmarks are administered to students individually by teachers or other educational professionals. All DIBELS benchmarks are timed, and the oral instructions and directions for administering individual items are standardized. DIBELS benchmark assessments are criterion measures, and benchmark goals and risk levels have been established using longitudinal student data. DIBELS measures are considered indicative of future student success or failure in reading (Carlson, Romhild, McCormick, Chin, and Geisinger, 2010). Specifically, students who fall into the DIBELS low risk category have an 80% chance of becoming proficient readers, students who fall into the DIBELS some risk category have a 50% chance of becoming proficient readers, and students who fall within the DIBELS at risk category have an 80% chance of not becoming proficient readers (Good & Kaminski, 2002).

DIBELS benchmark assessments are administered to students three times each year, and the results of these screenings are used to evaluate students' progress toward the mastery of early literacy skills. DIBELS benchmark assessments include the following measures: Initial Sound Fluency (ISF), which is administered in the fall and winter of kindergarten; Letter Naming Fluency (LNF), which is administered beginning in the fall of kindergarten and continuing through the fall of first grade; Phoneme Segmentation Fluency (PSF), which is administered beginning in the winter of kindergarten and continuing through the end first grade; Nonsense Word Fluency (NSF), which is administered beginning in the winter of kindergarten and continuing through second grade; Word Use Fluency (WUF), which is administered beginning in kindergarten and continuing through third grade; and Oral Reading Fluency (ORF), which is administered beginning in the winter of first grade and continuing through third grade.

The main purpose of DIBELS benchmark assessment is to identify and monitor the progress of students who have weak basic literacy skills in the attempt to insure that they will eventually become proficient readers. Because DIBELS benchmarks are intended to serve as a brief screening measures rather than a comprehensive reading assessment, each reading sub-skill is included as a separate benchmark. The set of benchmarks administered at each benchmarking period changes across time to ensure that the administered benchmarks assess the reading skills that typically developing students are learning and beginning to master at the time of the benchmark (Carlson et al., 2010; Kaminski, Cummings, Powell-Smith, & Good, 2008). Total assessment time is minimized by administering only the benchmarks designed to measure the reading skills that are likely to distinguish between students who are developing reading skills as expected and those who are at risk for reading failure at each benchmark period (Hintz, Ryan, & Stoner, 2003). For example, LNF is not routinely administered after the beginning of first grade due to ceiling effects that limit its usefulness as a predictor of children who are at risk for reading failure (Kaminski & Good, 1996). Similarly, ORF is not administered until the middle of first grade because many typically developing readers are not successfully reading long passages of connected text prior to mid-first grade (Kaminski et al., 2008). DIBELS results may be used to assess the effectiveness of reading instruction or the effectiveness of a particular intervention for individual students or for groups of students (e.g., by class, by school, or by district).

All Kindergarten and first grade students within the school system were administered selected subtests of the DIBELS at approximately the same times (fall, winter, and spring of the 2008-2009 school year). The decision to administer only selected subtests was made by school level administrators to reduce the amount of time spent on assessment activities. In the fall and winter, DIBELS subtests were administered by a system wide team of educational professionals (school psychologists, speech language therapists, instructional coaches, and Early Intervention Program Teachers, Special Education Teachers) that included representatives from each school in the system who were trained to administer the DIBELS measures. In the spring, DIBELS subtests were administered by the educational professionals from this team at their assigned schools. At all three benchmarking periods (fall, winter, and spring), students were tested in a centralized location at their school but outside of their classroom.

Initial Sound Fluency. The ISF measure is a standardized, individually administered measure of phonological awareness. It assesses the child's ability to recognize and produce the beginning sound in an orally presented word (Kaminski & Good, 1996, 1998). The ISF measure is a revision of an earlier measure, Onset Recognition Fluency (OnRF). To administer ISF the examiner presents four pictures to the child, names each picture, and asks the child to point to or say which picture begins with the sound said by the examiner. For example, the examiner says, "This is sink, cat, gloves, and hat. Which picture begins with /s/?" and the student points to the picture of the sink. The child is also asked to say the beginning sound of a word said by the examiner that matches one of the given pictures. The examiner calculates the amount of time taken to complete the tasks and converts that score into the number of initial sounds correctly produced in one minute. The ISF measure takes about 3 minutes to administer and score. Technical data regarding the ISF measure is somewhat limited (McBride, Ysseldyke, Milone, & Stickney, 2010), but a summary of the available data is presented in Table 4.

Letter Naming Fluency. The LNF benchmark is a standardized, individually administered measure of letter naming ability. It assesses a student's ability to recognize and name the upper and lower case letters of the alphabet (Kaminski & Good, 1996, 1998). To administer the LNF benchmark, the examiner presents the student with a page of upper- and lowercase letters, and asks the student to name as many letters as he or she can. The student's score is the total number of letters correctly named in 1 minute. Technical data for LNF is summarized in Table 4.

Nonsense Word Fluency. The NWF measure tests letter sound correspondence and the ability to blend letters into words (Kaminski & Good, 1996). The student is given a sheet of paper with randomly ordered vowel consonant and consonant-vowel-consonant non-words (e.g. sug, fud, af). The student is told that the "words" are not real words and asked to either read the word or say each sound. The final score is the number of letter-sounds produced correctly in a minute. Because the measure is fluency based, a student who phonologically recodes the words rather than providing sounds in isolation will automatically receive a higher score. Technical data for NWF is summarized in Table 4.

Oral Reading Fluency. The ORF measure is a standardized, individually administered test that measures a student's accuracy and fluency when reading connected text (Kaminski & Good, 2006). The passages are selected to reflect the grade level at which students should be reading at the end of the school year. The assessment requires each student to read aloud for one minute to the examiner. Student errors are marked, and the number of words read correctly in one minute becomes the student's score. Several studies, which are summarized in Table 4, have confirmed the technical adequacy of ORF.

Table 4

DIBELS Reliability and Validity Data

			Criterion Related Validity		
Grade	Test	Reliability	Concurrent	Predictive	Source
K	LNF		.61 NWF	.61 (NWF); .73 (TOWRE); .64 (ORF);	Burke et al. (2009)
K	LNF		.50 (ISF); .67 (NWF)	.71, .62 (ORF) .59, .72 (TOWRE); .51 (WRMTR)	Hagan-Burke, et al. (2006)
K	LNF	.89	.70 (WJ)	.72 (NWF); .66 (WJ); .72 (CBMORF)	Good et al. (2004)
K	LNF	.94	.58, .53, .52 (CTOPP)		Hintz et al. (2003)
K	LNF			.46 (TOPA2); .61, .55, ,62 (WJ3)	Nelson (2008)
K	LNF		.62 (DRA); .52, .59, .41, .32 (TERA3)	.67 (DRA); .48, .63, .57 (TerraNova)	Rouse & Fantuzzo (2006)
1	LNF			.40, .40, .36 (DIBELS ORF); .30, .22 (SAT10)	Chard et al. (2008)
1	LNF	.86	.53 (WJ)	.68 (NWF); .74 (CBMORF); .66 (NWF); .62 (WJ)	Good et al. (2004)
1	LNF		.58 (NWF); .62, .47 (TOWRE)		Hagan-Burke et al. (2006)
1	LNF			.44 (GRADE); .40 (TerraNova)	Riedel (2007)
1	LNF			.57, .52, .54, .52, .56, .30 (ITBS)	Schilling et al. (2007)
K	ISF		.51 (NWF)	.43, .38 (ORF); .32, .44 (TOWRE); .46 (WRMT-R)	Burke et al. (2009)
K	ISF	.61	.47 (PSF); .38 (WJ)	.35 (PSF); .29 (NWF); .37 (WJ); .36 (CBMORF)	Good et al. (2004)
K	ISF	.86	.60, .46 (CTOPP)		Hintz, Ryan, & Stoner (2003)
K	ISF		.66 (STAR) Early Literacy)		McBride et al. (2010)
K	ISF		.56 (TOPA2); .38, .31, .37 (WJ3)		Nelson (2008)

Grade	Test	Reliability	Criterion Related Validity: Concurrent	Criterion Related Validity: Predictive	Source
K	NWF			.73, .58 (ORF); .67, .67 (TOWRE); .56 (WRMTR)	Burke et al. (2009)
	NWF		.56 (TOPA2); .74, .71, .76 (WJ3)		Nelson (2008)
K	NWF	.86	.79 (LSF)	.76, .75, .65 (WRMTR); .72 (CBM ORF); .72 (LSF); .72 (NWF);	Ritchey (2008)
K	NWF		.62 (DRA); .53, .56, .44, .35 (TERA3)	.63 (DRA); .50, .57, .55 (TerraNova)	Rouse & Fantuzzo (2006)
K	NWF	.94	.36 (PPVTR); .65 (CTOPP); .27, .52 (TPRI); .91 (WJR)	.59, .59 (WJR); .77 (NWF); .71(CBM ORF)	Speece, Mills, Ritchey & Hillman (2003)
1	NWF		.69 (TOWRE)	.57 (ORF)	Burke et al. (2009)
	NWF		.68 (ORF); .75, .68 (TOWRE)		Burke & Hagan-Burke (2007)
1	NWF	.83	.51 (WJ)	.71, .75, .77 (CBMORF); .67 (WJ)	Good et al. (2004)
1	NWF		.73, .75 (TOWRE)		Hagan-Burke et al. (2006)
1	NWF		.77, .78, 77, .74 (ORF)	.62, .76 .63, .43, .72 (NWF); .82, .73, .74, .73, .56, .72 (ORF)	Harn et al. (2008)
1	NWF		.46 (GRADE)	.45, .45 (GRADE); .39, .38, .37 (TerraNova)	Riedel (2007)
1	NWF		.60, .54, .58, .59, .56 (ITBS)	.57, .51, .56, .54, .54, .57, .57, .56 (ITBS)	Schilling et al. (2007)
1	NWF		.71, .75 (WJR) .74 (CBMORF)		Speece et al. (2003)
1	ORF	.94, .98	.82 (SAT10)	.71, .63, .72 (SAT10)	Baker et al. (2008)
1	ORF		.77, .92 (TOWRE)		Burke & Hagan-Burke (2007)
1	ORF			.81 (ORF); .61 (WRMTR)	Burke et al. (2009)

Grade	Test	Reliability	Criterion Related Validity: Concurrent	Criterion Related Validity: Predictive	Source
1	ORF			.69, .91, .62, .85 (NWF)	Harn et al. (2008)
1	ORF		.67 (GRADE)	.59 (GRADE); .49, .54 (TerraNova)	Reidel (2007)
1	ORF		.75, .61, .74, .69, .71 (ITBS)	.69, .61, .69, .61, .63 (ITBS)	Schilling et al. (2007)

Abbreviations

CBMORF	Curriculum Based Measurement Oral Reading Fluency
CTOPP	Comprehensive Test of Phonological Processing
GRADE	Group Reading Assessment and Diagnostic Evaluation
ISF	DIBELS Initial Sound Fluency
ITBS	Iowa Test of Basic Skills
LSF	Letter Sound Fluency
NWF	DIBELS Nonsense Word Fluency
ORF	DIBELS Oral Reading Fluency
PPVTR	Peabody Picture Vocabulary Test, Revised
SAT10	Stanford Achievement Test, Tenth Edition
TOWRE	Test of Word Reading Efficiency
TOPA2	Test of Phonological Awareness, Second Edition
TPRI	Texas Primary Reading Inventory
WJ	Woodcock Johnson Psycho Educational Battery
WJR	Woodcock Johnson Psycho Educational Battery, Revised
WJRMT	Woodcock Johnson Reading Mastery Test

Georgia Criterion Referenced Competency Test. The No Child Left Behind Act of 2001 requires that schools, districts, and the state be accountable for the academic performance of all students. The *Georgia Criterion Referenced Competency Test* (CRCT) assessment program is the designated assessment tool for federal accountability in Georgia for grades one through eight (Georgia Department of Education [GaDOE], 2009). Georgia chose to administer the CRCT to first grade students even though the assessment of first graders was not required in the federal law. According to the Georgia Student Assessment Program Student Assessment Handbook 2008-2009, the CRCT is “designed to measure student acquisition of the knowledge, concepts, and skills set forth in the Georgia Performance Standards” (p. 45) in grades one through eight in reading, English language arts, and mathematics, and in grades three through eight in science and social studies (GaDOE, 2009). Student CRCT scores are reported as scale scores, which can range from 650 to 900. GaDOE cut scores indicate that a score of at least 800 meets state standards and a score of at least 850 exceeds state standards.

For the 2008-2009 school year, all Georgia public school students enrolled in grades 1-8 were required to participate in the CRCT by the GaDOE. The GaDOE determined a 30-day state testing window in the spring, and each school district selected a nine-day testing window within that 30 day period to administer the CRCT. The CRCT was administered to first grade students in this study over three consecutive days (Monday through Wednesday) during the third week in April. Students who were absent during testing, took the portion or portions of the test that they missed on Thursday and Friday. Students were administered one subject each day, and the reading portion of the CRCT was administered on the first day of testing. It consisted of two test sections that lasted about

60 minutes each. On the reading test, students read comprehension passages independently and teachers read each question and its 4 answer choices aloud to the students (GaDOE, n.d.).

The content of the first grade CRCT included the domains of vocabulary and comprehension. Vocabulary was defined as “the skills required to read, interpret, and acquire new vocabulary in a variety of texts” (GaDOE, 2007, p. 3). It included skills such as differentiating between words with multiple meanings; identifying antonyms and synonyms; understanding root words, prefixes and suffixes; and understanding compound words. Comprehension was defined as “the skills required to read, recall, comprehend and explore various texts including literary texts (narratives, stories, poems, fables), and information from a wide variety of informational articles” (GaDOE, 2007, p. 5). It included skills such as identifying narrative elements, summarizing and paraphrasing, identifying the main idea and supporting details, and interpreting information from graphic features such as charts and diagrams.

Validity. The GaDOE followed the guidelines published in the *Standards for Educational and Psychological Testing (Standards*, 1999*)* developed by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) when developing the CRCT (GaDOE, 2009). The *Standards* define validity as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (9). One source of evidence for establishing a test’s overall validity is to examine its content validity, which begins with the clear identification of the purpose of the test by the test developer.

According to the Georgia legislature, the main purpose of the CRCT is to measure how well students have mastered the state's curriculum, the Georgia Performance Standards (GPS; O.C.G.A. § 20-2-281). The CRCT is also intended to identify areas where individual students and groups of students need improvement, to inform various stakeholders of progress toward meeting academic achievement standards set by the state and the federal NCLB (2003), and to gauge the overall quality of education in the state of Georgia. The GaDOE asserts, therefore, that, the validity of the CRCT depends primarily on its match to the curriculum as expressed in the GPS and taught in Georgia schools and on how well its score reports inform the various stakeholders – students, parents, and educators – about student performance at the student, class, school, system, and state levels (An Assessment and Accountability Brief: Validity and Reliability for the 2009 CRCT). While this information clearly addresses content validity and the GaDOE provides extensive information regarding the method used to develop CRCT test items and cut scores as described below, it does not address criterion validity.

Criterion validity is the degree to which a test's results match other known measures of the same or similar content. The only information provided by the GaDOE regarding criterion validity was the following statement found on page 4 of *An Assessment and Accountability Brief: Validity and Reliability for the 2009 CRCT*, "The department has also conducted analyses as evidence of external validity by comparing how the constructs the CRCT measures compare with other well-recognized assessments (e.g., ITBS)." (GaDOE, 2009). No data from these comparisons was available.

With regards to content validity, the GaDOE asserts that the CRCT is a valid instrument for its stated purpose because its development began with committees of educators from around the state who reviewed the GPS to create a test blueprint and test specifications that identified the concepts, knowledge, and skills that would be assessed by the CRCT and how they would be assessed on the CRCT. Guidelines for the item writing phase were developed from the test blueprint and test specifications. For example, the item guidelines for the reading test address the genres, complexity, and lengths of the reading passages produced for the test. This process resulted in the development of the CRCT Content Descriptions and the CRCT Content Weight documents which were posted on the state department of education's website so that all stakeholders were informed of the test's content and method of assessment. These documents and the inclusion of Georgia educators in the test development process serve as evidence of the CRCT's validity as a measure of the state's curriculum (GaDOE, 2009).

Once the CRCT Content Descriptions and the CRCT Content Weight documents were developed, multiple choice test items were written by qualified, professional assessment specialists specifically for the CRCT. Emphasis was placed on the assessment of higher order thinking skills. Each item was to have one clearly correct answer and appropriate, relevant, and reasonable distracters. Items were to be free from bias toward or against any particular group. New committees of Georgia educators reviewed the items for alignment with the curriculum, suitability, and potential bias or sensitivity issues. The review committees had the authority to accept, revise, or reject items. Accepted items were field tested by embedding the items into the currently operating CRCT. Field testing in this manner ensured that new test items were taken by a representative group of motivated students under standard conditions (GaDOE, 2009).

The modified Angoff method (Horn, Ramos, Blumer & Madaus, 2000; Tiratira, 2008) was used to determine the passing percentage, or cut score for the CRCT (GaDOE, n.d.). When calculating dichotomous cut scores (i.e. pass or fail), the modified Angoff method (Horn, et al; Tiratira, 2008) requires that subject-matter experts examine the content of each test item and then make a judgment regarding the probability that a minimally-qualified candidate would answer the item correctly from among eight choices (.5, .20, .40, .60, .75, .90, .95, or "Do not know."). Next, the sum total of each judge's rating of each item is computed. The sum totals from each judge are then averaged. This average becomes the recommended Angoff cut score. Because NCLB (2003) requires that states distinguish between students who do not meet standards, meet standards, and exceed standards, the Angoff method would be repeated with the subject-matter experts making a judgment about the probability that a candidate who met but did not exceeded expectations would answer an item correctly.

Once the Agnoff cut scores have been identified, the subject-matter experts are provided with information regarding the percentage of students who would not meet, meet, or exceed the standard based on their recommended score. At this point recommended cut scores may be revised (GaDOE, n.d.; Hall, Howerton, & Jones, 2008; Horn, et al, 2001). Once the subject-matter experts have finalized their cut scores, their recommendations are submitted to the Office of Student Achievement, the State Superintendent of Schools, and the State Board of Education for review and adoption (GaDOE, n.d.). Recommended cut scores may also be revised during the adoption process (Horn et al., 2001). No information regarding changes made to the recommended cut scores at the various stages in the process was available. Finally, an independent panel of experts in the field of educational measurement, Georgia's Technical Advisory Committee (TAC), reviewed all aspects of the test development and implementation process (GaDOE, n.d.).

While the procedures used to develop the CRCT were consistent with the recommendations of the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) for the development of criterion referenced tests (GaDOE, 2009) and similar to those used by other states (e.g. Virginia; Horn et al., 2001), they may result in potential limitations to the validity of the CRCT as a measure of student achievement. For example, the selection of content and the process for setting standards for the CRCT are based on judgments. Even when a well-conceived standard setting procedure is used, the result of the process is arbitrary. Resulting cut scores cannot be said to reflect any objective measure of competency with certainty unless other external evidence exists to support that conclusion (Horn et al., 2001). The fact that the subject-matter experts are provided with data regarding the percentages of students who were classified into each category further limits the objectivity of the cut-score by introducing a political element into the process. Experts may adjust recommended cut scores to increase or decrease the percentage of students within a given category (Hall, Howerton, & Jones, 2008). Finally, dividing students into more than two groups, as required by NCLB (2003), creates a dilemma. The majority of students who have received good instruction should have mastered the educational content measured by the test. Therefore, to identify students who exceed the standard in addition to students who met the standard, the test must contain more items of greater difficulty than would be necessary to make the distinction between students who are proficient and those who are not. As a result, students who are proficient may be misidentified as not meeting the standard as an artifact of test construction (Hall et al., 2008).

***Reliability*.** It is possible for test scores to be reliable but not valid. Reliability is the ability of the test to provide accurate and consistent interpretation of scores (Mason, 2007). Reliability is important because it affects the interpretation and confidence given to test results and ultimately the accuracy of decisions made about students (Hall, Howerton, & Jones, 2008). Reliable criterion referenced assessment occurs when item responses are consistent across all items in the domain (Horn et al., 2000). Nunnally (1975) stated that in applied settings “where important decisions must be made about humans on the basis of test scores, even a reliability of .90 is not be high enough.” (p. 10). Failure to meet high standards of reliability may result in large numbers of students being misidentified. For the 2009 First Grade Reading CRCT several reliability indices were reported. Internal consistency Cronbach’s alpha reliability coefficient was .89. The standard error of measurement (SEM) for student raw scores on the test was 2.24. Finally, a conditional standard error of measurement (CSEM), which expressed the degree of measurement error in scale score units was calculated using Hambleton and Swaminathan’s procedure and formula (as cited in GaDOE, 2009). The CSEM was 7 for the Meets Expectation cut score and 12 for Exceeds Expectations cut score.

Procedure

The pretest measures used in this study included the fall DIBELS ISF and LNF scores for kindergarten students and the fall DIBELS NWF scores for first grade students. The dependent variables for kindergarten students in this study were the spring DIBELS

LNF and NWF correct letter sound scores. The dependent variables for first grade students in this study were the spring DIBELS ORF score and the reading portion of the CRCT. The independent variable was participation in Headsprout.

After DIBELS administration was completed in the fall, teachers met by grade level at their school to determine which students would participate in Headsprout. Students at risk for reading failure at each grade level were divided into two groups by the grade level teachers at their schools based on fall DIBELS scores and teacher judgment. Grade level teachers used the scores from the ISF and LNF subtests (kindergarten) and the LNF and NWF subtests (first grade) to divide students into groups. The groups were students at risk for reading failure who participated in Headsprout and students at risk for reading failure who did not participate in Headsprout. The treatment group, students who participated in Headsprout, was comprised of the students at each school who were at greatest risk for reading failure based on the DIBELS data available in the fall of 2008 and teacher judgment. Because teachers met and considered all students at their schools by grade level for inclusion in the Headsprout program, some teachers had more students participating in Headsprout than did other teachers.

Data Collection.

The researcher was given access to the school district's data stored on the DIBELS website for the 2008-2009 school year. The data were downloaded from the website to an Excel spreadsheet using a school system computer. Data for kindergarten students identified as low risk on both the ISF and the LNF subtests of the DIBELS administered in the fall of 2008 was not included in this research. Data for first grade students identified as low risk on both the LNF and NWF subtests of the DIBELS administered in the fall of 2008 were not included in this research. Additionally, as described in the subjects section, 90 kindergarten students and 111 first grade students were not included in the research because they had been retained, were not enrolled in the school system for the entire study year, had previously participated in Headsprout, or did not appear to be at risk for reading failure. The resulting data pools included 373 kindergarten and 359 first grade students at risk for reading failure.

At the school system's request, the Headsprout Corporation provided the researcher with a copy of their student records from the 2008-2009 school year. Headsprout records included the following information for each student who participated in the program: name, teacher's name, school attended, gender, enrollment date, last episode completed, highest episode completed, and total time spent in the program reported in hours.

Data Collation. Once the data from the DIBELS website and the Headsprout Corporation were obtained, the researcher manually combined the information about students at risk for reading failure contained in each Excel spreadsheet into a new Excel spreadsheet and added student demographic data (gender, age in months, socio-economic status, and enrollment in special programs such as the Early Intervention Program (EIP) English for Speakers of Other Languages (ESOL) and Special Education and CRCT data (first grade only) to the data set.

Outliers. The data were examined through the Statistical Package for the Social Sciences (SPSS; IBM Inc., 2010) for outliers. In the kindergarten data set the fall LNF, winter ISF and LNF, and the spring LNF and NWF data had no outliers at the .001 level (Tabachnick & Fidell, 2007). The fall ISF variable had 7 outliers at the .001 level. These outliers were visually inspected. Based on visual inspection the cases that included these variables were not at risk on the fall LNF variable, were not at risk on the winter ISF and LNF variables, and were not selected by teachers to participate in Headsprout; therefore, the cases were removed from the data set because they did not appear to represent students at risk for reading failure. The removal of these 7 cases left a remaining kindergarten data set of 373 cases. Examination of the first grade data from the fall administration of the DIBELS NWF, spring administration of the DIBELS ORF, and CRCT revealed no outliers at the .001 level (Tabachnick & Fidell, 2007). Specific demographic information for these kindergarten and first grade students is presented in Table 2.

Missing data. The kindergarten and first grade data sets were inspected for missing data. Inspection of the kindergarten data for missing data revealed 41 instances of missing data for the spring LNF dependent variable and 35 instances of missing data for the spring NWF dependent variable. Inspection of the first grade data revealed 49 instances of missing data for the spring ORF dependent variable and no instances of missing data for the CRCT dependent variable. Further inspection of the data revealed that in both the kindergarten and first grade data sets, the majority of the missing data occurred in control group cases, most likely due to the fact that when schools were asked to collect spring data without assistance from a system wide team, they placed greater emphasis on collecting data about students in the treatment group than they did on collecting data about students in the control group.

This pattern of missing data, where there is a relationship between a variable in the data set and the likelihood of data to be missing, means that the data are missing at random (MAR; Baraldi & Enders, 2009). When data are MAR, traditional missing data techniques are inappropriate, and modern missing data analysis such as multiple imputation or maximum likelihood are recommended. Either of these techniques is appropriate for use with data that are MAR (Baraldi & Enders, 2009; Graham, 2009). Because SPSS (IBM Inc., 2010) provides a means for implementing multiple imputation to handle missing data, multiple imputation was used to account for the missing data in both the kindergarten and first grade data sets. Multiple imputation uses linear regression to impute, or predict, what the missing values would have been given the available data. (Baraldi & Enders, 2009; Tabachnick & Fidell, 2007).

Kindergarten data. Multiple imputation was used to account for missing data on the spring LNF and spring NWF dependent variables. 41 values on the spring LNF variable were imputed and 35 values on the spring NWF variable were imputed. The variables included in the imputation model (linear regression), in order of entry into the model, were Headsprout participation, age in months, socio-economic status, supplemental program participation, winter LNF score, winter NWF score, spring NWF score, and spring LNF score. The means and standard deviations for the original data set and each imputation are presented in Tables 5, 6, 7, and 8.

First grade data. Multiple imputation was used to account for the missing data in the spring ORF dependent variable. No data from the fall NWF variable or the CRCT variable were missing. The variables included in the imputation model (linear regression), in order of entry into the model, were Headsprout participation, age in months, socio-economic status, supplemental program participation, winter NWF score,

Table 5

K Spring LNF Multiple Imputation Descriptive Statistics for the Treatment Group

	N	Mean	SD	Med.	Mode	Min.	Max.	Skew	Kurt.
Original	117	35.66	15.24	37.00	38, 40	1	70	-.207	-.273
Imp 1	130	35.83	14.78	37.00	38, 40	1	70	-.168	-.596
Imp 2	130	35.74	15.77	37.50	38, 40	0	70	-.231	-.363
Imp 3	130	36.24	15.87	38.00	38, 40	1	70	-.141	-.427
Imp 4	130	35.85	15.30	38.00	38, 40	1	70	-.203	-.383
Imp 5	130	35.40	15.46	37.00	38, 40	1	74	-.101	-.213
Pooled	130	35.81							

Table 6

K Spring LNF Multiple Imputation Descriptive Statistics for the Control Group

	N	Mean	SD	Med.	Mode	Min.	Max.	Skew	Kurt.
Original	215	46.37	13.73	46.00	37	8	92	.303	.900
Imp 1	243	46.08	14.04	46.00	37	5	92	.249	.882
Imp 2	243	45.88	13.89	45.00	37	8	92	.253	.686
Imp 3	243	46.28	13.96	46.00	37	8	92	.240	.671
Imp 4	243	45.93	14.27	46.00	37	8	92	.141	.746
Imp 5	243	45.86	13.70	46.00	37	8	92	.281	.780
Pooled	243	46.01							

Table 7

K Spring NWF Multiple Imputation Descriptive Statistics for the Treatment Group

	N	Mean	SD	Med.	Mode	Min.	Max.	Skew	Kurt.
Original	126	21.33	14.27	20.00	14	0	69	.835	.768
Imp 1	130	21.40	14.37	20.00	14	0	69	.798	.596
Imp 2	130	20.97	14.19	19.00	14	0	69	.877	.828
Imp 3	130	21.17	14.12	19.50	14	0	69	.860	.850
Imp 4	130	20.92	14.24	19.00	14	0	69	.869	.805
Imp 5	130	21.12	14.14	19.50	14	0	69	.864	.846
Pooled	130	21.12							

Table 8

K Spring NWF Multiple Imputation Descriptive Statistics for the Control Group

	N	Mean	SD	Med.	Mode	Min.	Max.	Skew	Kurt.
Original	212	32.50	13.53	31.00	33	0	76	.772	.333
Imp 1	243	32.61	13.26	31.28	33	0	76	.771	.995
Imp 2	243	32.57	13.44	32.00	33	0	76	.668	.835
Imp 3	243	32.14	13.71	31.00	33	0	76	.696	.907
Imp 4	243	31.98	13.72	31.00	33	0	76	.696	.932
Imp 5	243	32.28	13.30	31.00	33	0	76	.719	.933
Pooled	243	32.32							

winter ORF score, spring NWF score, and spring ORF score. The means and standard deviations for the original data set and each imputation are presented in Tables 9 and 10.

Data Analysis.

The proposed data analysis statistic was ANCOVA. The first proposed covariate for each of the two outcome measures in both the kindergarten and first grade data sets was student age in months. This covariate was proposed because prior research has shown that student age is correlated with the acquisition of early literacy skills (Paris, 2005). Additionally, a pretest measure that prior research (see Table 4) suggested was correlated with each post-test measure was proposed for each posttest in the analysis to minimize the effects initial differences in reading skill between the control group and the treatment group. That is, for the kindergarten data set fall LNF scores were proposed as a covariate in the analysis of spring LNF scores and fall ISF scores were proposed as a covariate in the analysis of spring NWF scores. For the first grade data set, fall NWF scores were proposed as a covariate in the analysis of spring ORF scores and CRCT scores.

Table 9

First Grade Spring ORF Multiple Imputation Descriptive Statistics for the Treatment Group

	N	Mean	SD	Med.	Mode	Min.	Max.	Skew	Kurt.
Original	169	40.75	22.64	38.00	52	4	113	.835	.501
Imp 1	181	40.97	22.37	38.00	52	4	113	.777	.432
Imp 2	181	40.61	22.07	38.00	52	4	113	.852	.651
Imp 3	181	40.70	22.25	38.00	52	4	113	.855	.359
Imp 4	181	40.91	22.09	38.00	52	4	113	.820	.589
Imp 5	181	40.21	22.22	38.00	52	4	113	.885	.627
Pooled	181	40.68							

Table 10

First Grade Spring ORF Multiple Imputation Descriptive Statistics for the Control Group

	N	Mean	SD	Med.	Mode	Min.	Max.	Skew	Kurt.
Original	141	46.40	29.20	40.00	38	0	138	.845	.286
Imp 1	178	47.22	27.31	43.00	38	0	138	.767	.412
Imp 2	178	45.18	26.47	40.00	38	0	138	.983	.897
Imp 3	178	46.72	27.51	41.70	38	0	138	.788	.379
Imp 4	178	46.38	26.93	41.00	38	0	138	.852	.683
Imp 5	178	45.26	27.11	39.62	38	0	138	.921	.723
Pooled	178	46.15							

CHAPTER 4

RESULTS

In this chapter, the data informing each research question are presented. The research questions for this study were: 1) Do Kindergarten students at risk for reading failure who participate in Headsprout in addition to their regular classroom reading instruction demonstrate different skill levels in reading as measured by the DIBELS LNF and NWF tasks than kindergarten students at risk for reading failure who do not participate in Headsprout? and 2) Do first grade students at risk for reading failure who participate in Headsprout in addition to their regular classroom reading instruction demonstrate different skill levels in reading as measured by the DIBELS ORF task and the CRCT than first grade students at risk for reading failure who do not participate in Headsprout?

Research Question 1 (Kindergarten)

In order to address the kindergarten research question posed in this study, two hypotheses, a null hypothesis and an experimental hypothesis, were considered. The null hypothesis was that kindergarten students at risk for reading failure who participated in Headsprout would perform similarly to kindergarten students at risk for reading failure who did not participate in Headsprout on spring DIBELS LNF and NWF measures. The experimental hypothesis was that kindergarten students at risk for reading failure who participated in Headsprout would have scores on spring DIBELS LNF and NWF measures that were different from those of kindergarten students at risk for reading failure who did not participate in Headsprout.

Prior to statistical analysis, the Kindergarten data were examined through SPSS (IBM Inc., 2010) for assumptions of normality of sampling distributions, homogeneity of variance, and homogeneity of regression slopes. Fall ISF scores were not normally distributed for either the treatment group (see Figure 1) or the control group (see Figure 2). The distribution of scores for the treatment group had a mean of 5.07, a median of 5, and modes of 3 and 5, resulting in a bimodal, positively skewed (.984) distribution, with the majority of the scores falling at or near zero, and a maximum score of 16. The distribution of scores in the control group was unimodal and positively skewed (.965). It had a mean of 5.86, a median of 5, and a mode of 6. The maximum score for this distribution was also 16.

The data for Fall LNF variable were not normally distributed for either the treatment (see Figure 3) or the control group (see Figure 4). The distribution of scores for the treatment group had a mean of 6.02, a median of 3, and a mode of 0, indicating a unimodal, positively skewed (1.71) distribution of scores, with a maximum score of 32. The distribution of scores in the control group was also unimodal and positively skewed (.884). It had a mean of 12.82, a median of 9, and a mode of 2. The maximum score for this distribution was 46.

The data for the spring LNF variable were normally distributed for both the treatment group (see Table 5) and the control group (see Table 6). The distribution of spring LNF scores for the treatment group had a pooled mean of 35.81, and the distribution of spring LNF scores in the control group had a pooled mean of 46.01.

Figure 1

Kindergarten Fall ISF Scores for the Treatment (HS: 1) Group

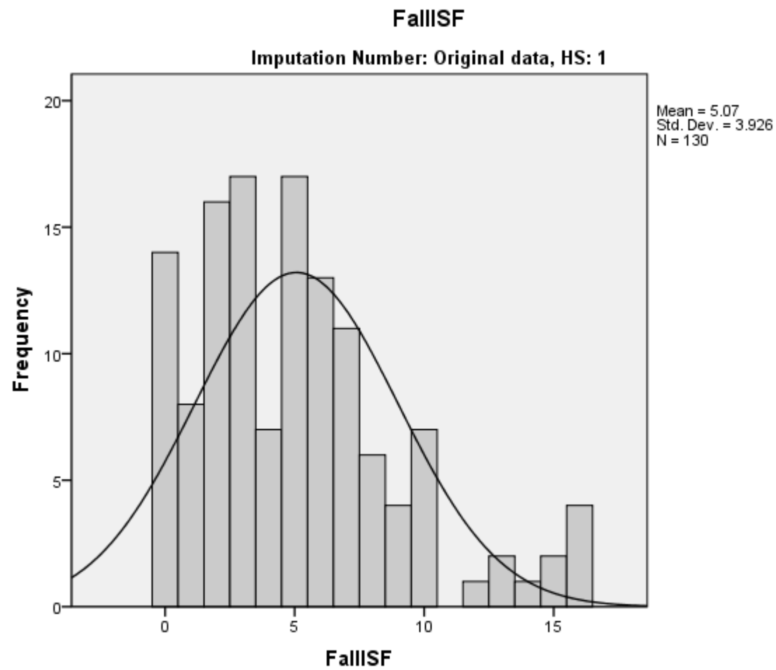


Figure 2

Kindergarten Fall ISF Scores for the Control (HS: 2) Group

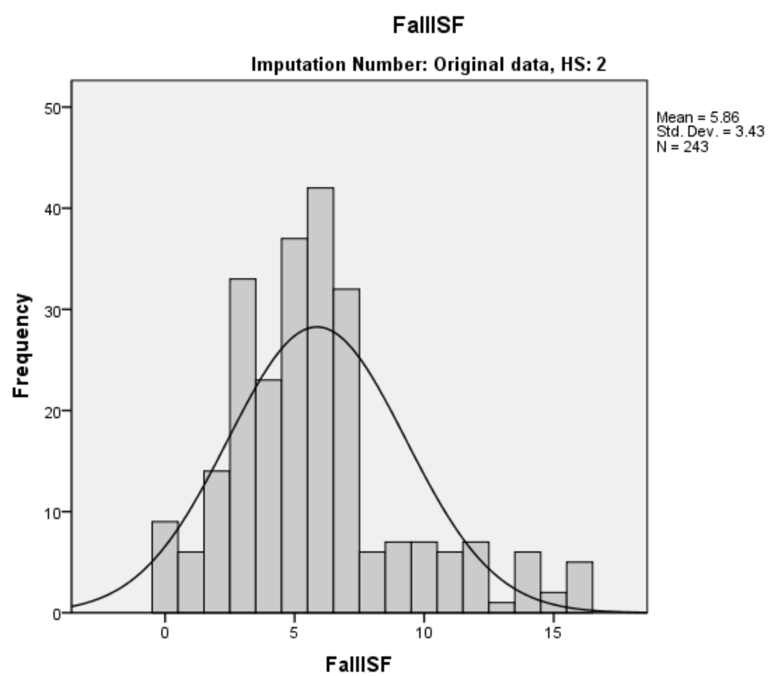


Figure 3

Kindergarten Fall LNF Scores for the Treatment (HS: 1) Group

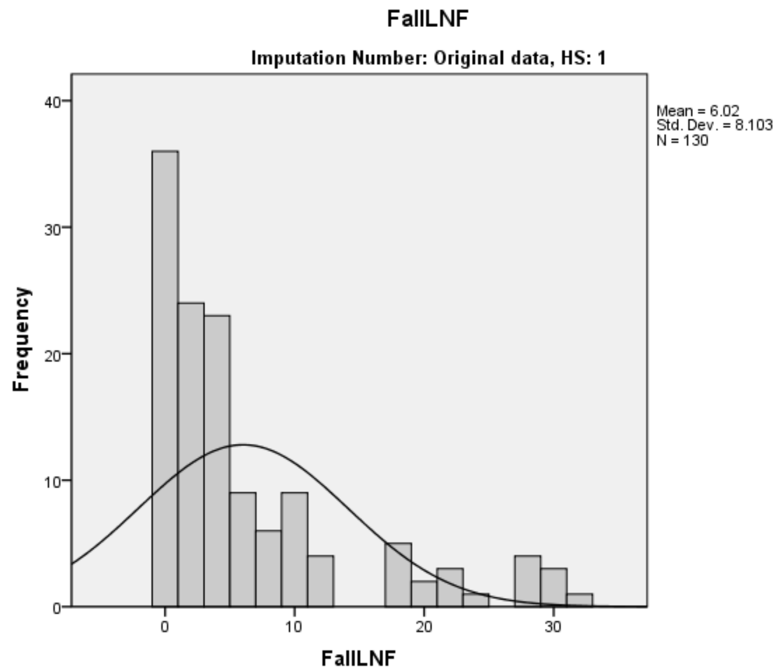
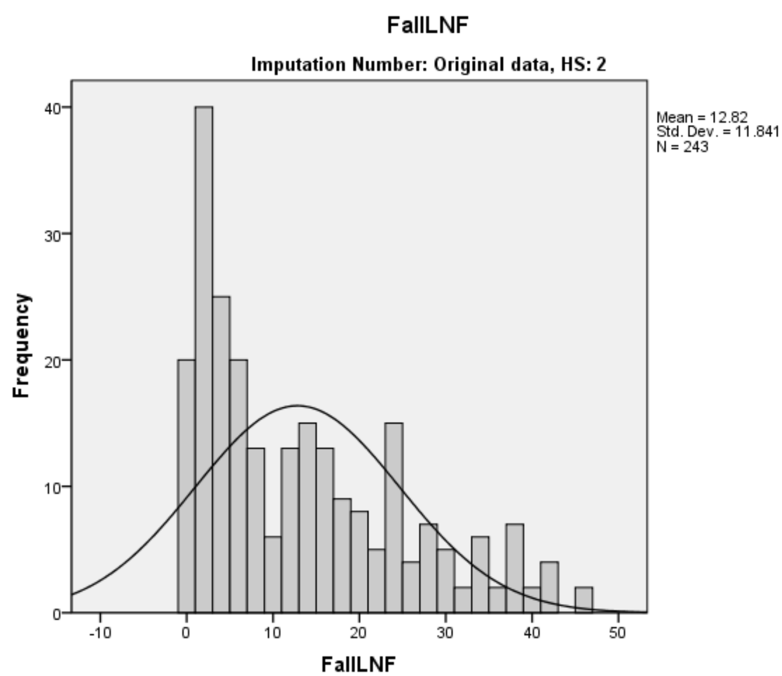


Figure 4

Kindergarten Fall LNF Scores for the Control (HS: 2) Group



Spring NWF scores were not normally distributed for the treatment group (see Table 7), but they were normally distributed for the control group (See Table 8). The distributions of spring NWF scores for the treatment group had a pooled mean of 21.12, the medians ranged from 19 - 20, and the modes were 14, resulting in unimodal, positively skewed (.798 - .869). The distributions of spring NWF scores for the control group had a pooled mean of 32.32, the medians ranged between 31 and 32, and the modes were 33.

Because most data were not normally distributed, and the number of subjects in the control group was roughly double the number of subjects in the treatment group, the proposed analysis using ANCOVA was not completed. Instead, the treatment group was redefined to include kindergarten students at risk for reading failure who completed at least 25 of the 40 Headsprout episodes recommended by the Headsprout Corporation for kindergarten students. One student was eliminated from the treatment group due to missing spring LNF and NWF data. The resulting in a treatment group included 51 kindergarten students who had completed at least 25 Headsprout episodes. A matched control group of kindergarten students who had not participated in Headsprout was selected from the pool of kindergarten students at risk for reading failure based on the following criteria: 1) Fall ISF score within 5 points, Fall LNF score within 5 points, age in months within 3 months, 4) an exact match on support program participation, and 5) an exact match on either socioeconomic status designation or race designation. The resulting matched data set included 51 kindergarten students in treatment group and 51 students in the control group. No data were missing from the matched data set.

Revised Analysis.

The final treatment group (matched treatment group) included 51 kindergarteners (31 male, 20 female), and the matched control group included 51 kindergarteners (28 male, 23 female). These students came from 10 schools within the district. The number of students from each school ranged from as few as 4 to as many as 15, and 76 of the students came from just 6 schools. Students completed an average of 1.6 Headsprout episodes each week. There was no significant difference between the groups in mean age (*t* (100) = .415, *p* = - .679): 65.82 months (SD = 3.08) and 65.57 months (SD = 3.11) for the matched treatment and control groups, respectively. The kindergarteners came from diverse socio-cultural backgrounds: White (29% treatment, 27% control), black (17% treatment, 14% control), Hispanic (6% treatment, 11% control), multiracial (2% treatment, 6% control), and Asian (2% treatment, 2% control). The students also participated in a variety of programs at their schools including an Early Intervention Program (17.6 % treatment, 17.6 % control) special education (9.8 % treatment, 9.8% control), and English for Speakers of Other Languages (5.9% treatment, 5.9% control). A few students participated in more than one of these support programs (5.9% treatment, 5.9% control). The percentage of treatment students who qualified for free or reduced lunch was not significantly different from the percentage of control students qualified for free or reduced lunch: 52.9 % treatment, 47.1 % control (X^2 = .044, p > .05). The treatment and control groups were also similar on both pretest measures, fall ISF and fall LNF. For ISF, *t* (100) = .425, p = .671. The treatment group had an ISF mean of 4.67 (SD = 3.70), and the control group had an ISF mean of 4.98 (SD = 3.73). For fall LNF, *t* (100)

= -.425, *p* = .815. The treatment group had an fall LNF mean of 8.67 (SD = 9.18), and the control group had an fall LNF mean of 9.08 (SD = 8.52).

The matched data set was examined through SPSS (IBM Inc., 2010) for assumptions of normality of sampling distributions and homogeneity of variance. Spring LNF scores and LNF change scores were normally distributed for both the treatment the control groups (see Figures 5-8) based on results of Kolmogorov-Smirnov, treatment group spring LNF *D* (51) = .065, *p* = .200; control group spring LNF *D* (51) = .089, *p* = .200; treatment group LNF change *D* (51) = .093, *p* = .200; and control group LNF change *D* (51) = .092, *p* = .200. The assumption of homogeneity of variance was not met for any of the spring LNF data based on the results of Levene's test. For spring LNF scores, *F* (1, 100) = 5.58, *p* = .02) and for spring LNF change scores, *F* (1, 100) = 194.879, *p* =.001, indicating that the variances between the treatment and control groups were significantly different.

Spring NWF scores were not normally distributed for either the treatment group (see Figure 9) based on the results of Kolmogorov-Smirnov, Spring LNF *D* (51) = .169, *p* = .001, or the control group (see Figure 10) based on the results of Kolmogorov-Smirnov, Spring LNF D (51) = .138, *p* = .014. The distribution of scores for the treatment group had a mean of 25.1, a median of 22, and a mode of 21, resulting in a unimodal, positively skewed distribution, with a minimum score of 2 and a maximum score of 69. The distribution of scores for the control group had a mean of 32.47, a median of 31, and a mode of 22, resulting in a unimodal, positively skewed (.750) distribution, with a minimum score of 11 and a maximum score of 65. The assumption of homogeneity of variance was met for the spring NWF data based on the results of Levene's test,

Figure 5

Kindergarten Spring LNF Scores for the Matched Treatment (HS: 1) Group

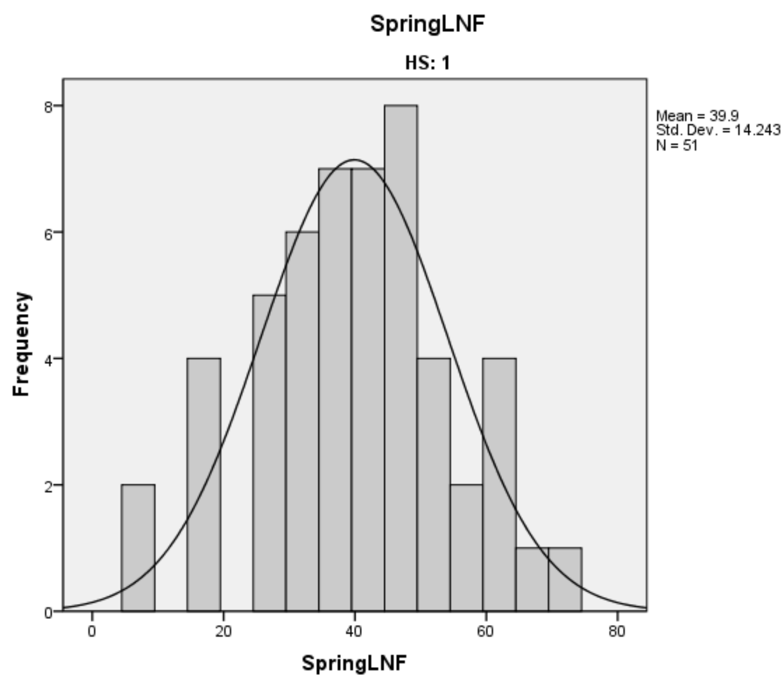


Figure 6

Kindergarten Spring LNF Scores for the Matched Control (HS: 2) Group

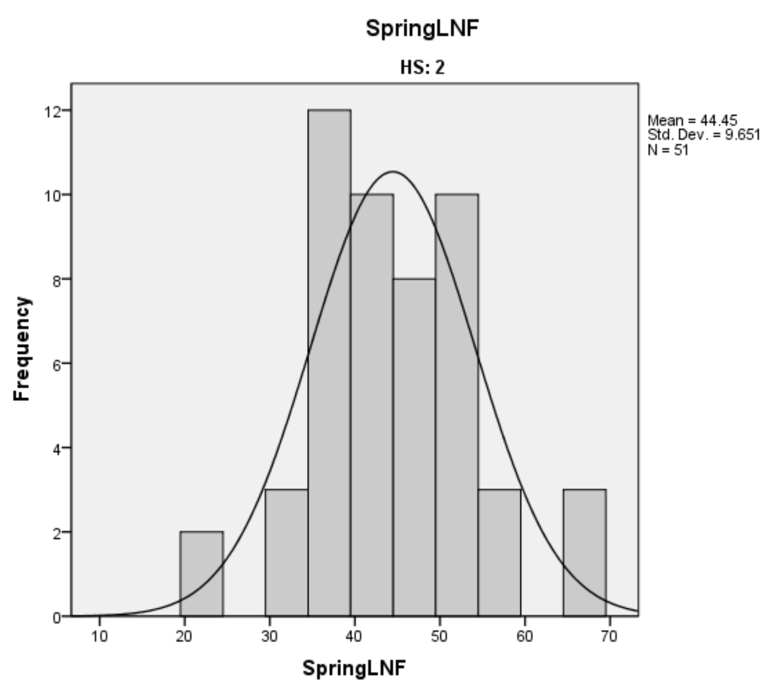


Figure 7

Kindergarten LNF Change scores for the Matched Treatment (HS: 1) Group

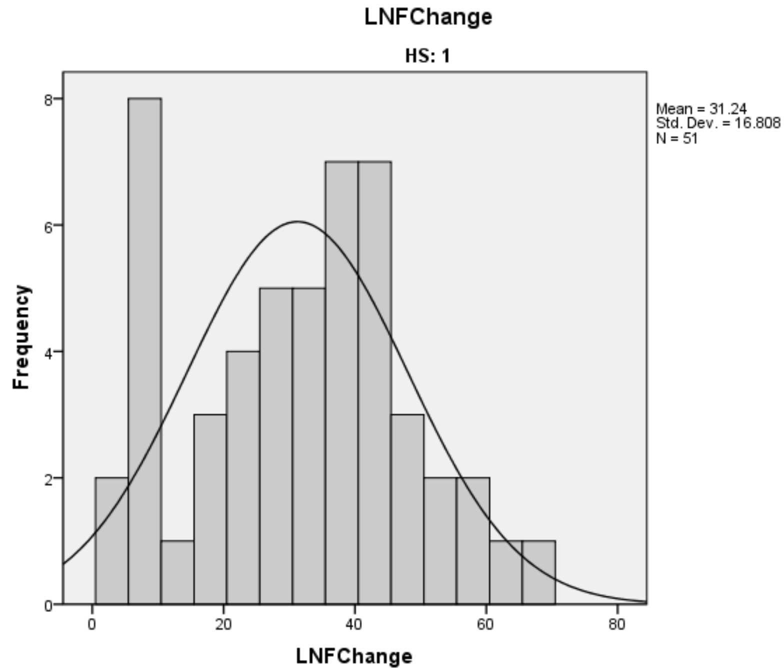


Figure 8

Kindergarten LNF Change Scores for the Matched Control (HS: 2) Group

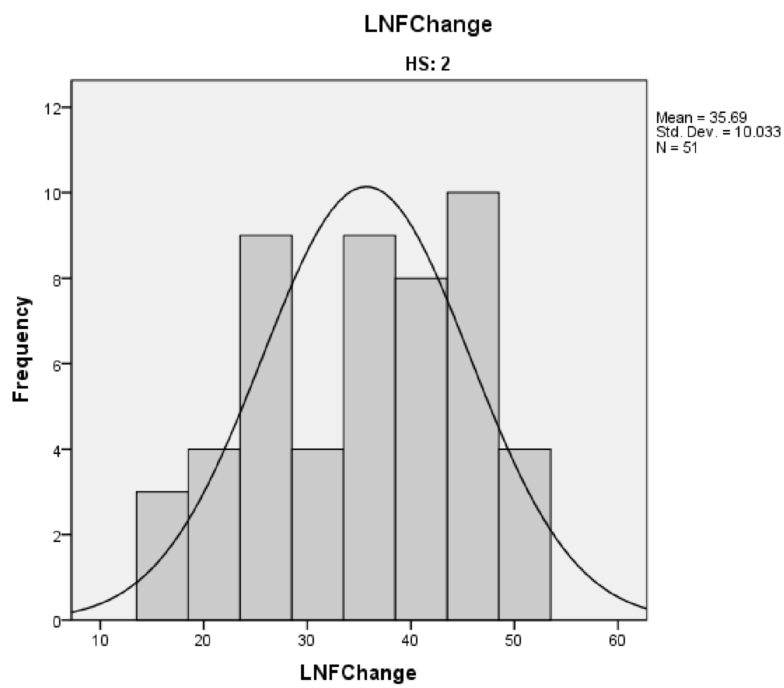


Figure 9

Kindergarten Spring NWF Scores for the Matched Treatment (HS: 1) Group

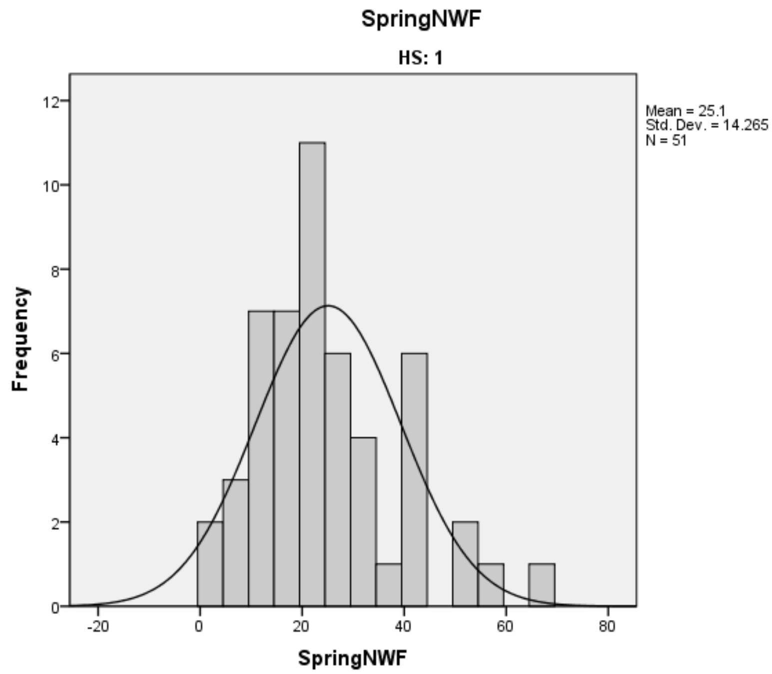
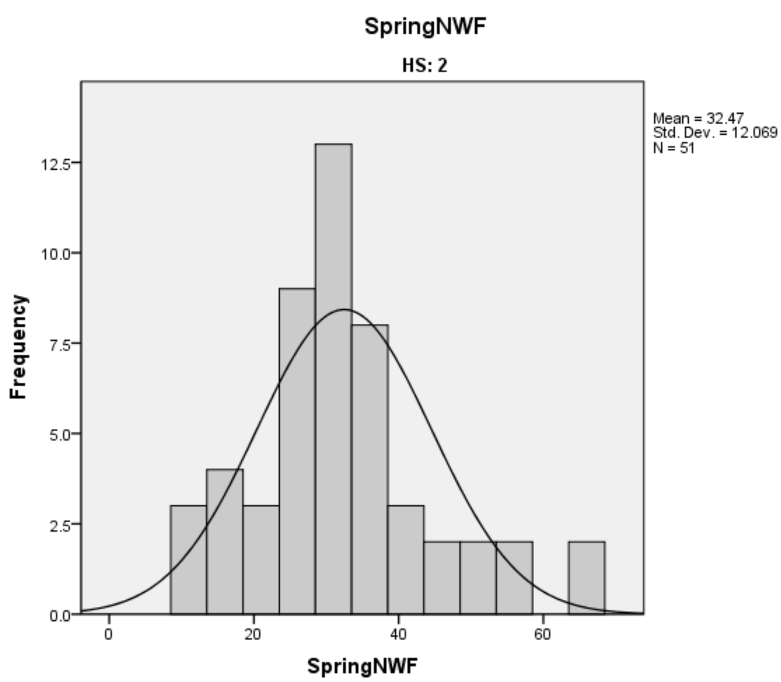


Figure 10

Kindergarten Spring NWF Scores for the Matched Control (HS: 2) Group



F (1, 100) = 1.84, *p* = .178), indicating that the variances between the treatment and control groups were not significantly different.

All data were analyzed using independent samples *t*-tests because *t*-tests are robust to violations of normality when group sizes are relatively equal (Tabachnick & Fidell, 2007), and the control and treatment groups in this data set were exactly equal. The more stringent value of *t* provided in SPSS for use with data that do not meet the assumption of homogeneity of regression slopes was used where appropriate.

Spring LNF, LNF Change, and NWF Results. Table 11 presents mean pretest (Fall LNF and Fall ISF), posttest (Spring LNF and Spring NWF), and change (LNF Change) scores for students in the treatment and control groups. Independent samples *t*-tests were performed to determine whether there were any significant differences between the groups at pretest. No significant differences between the groups were found on either fall ISF scores (*t* = -.425 (100); *p* = .671) or fall LNF scores (*t* = -.235 (100); *p* = .815).

Independent samples *t*-tests were also conducted to compare the spring LNF and spring LNF scores of kindergarten students in the treatment group with those of students in the control group. LNF change scores were also calculated for each group, and the change scores were compared using an independent samples *t*-test. There were no significant differences between groups on Spring LNF scores (*t* = -1.888 (100); *p* = .062) or LNF change scores (*t* = -1.624 (100); *p* = .108). There was a significant difference between groups for spring NWF scores in favor of kindergarten students in the control group (*M* = 32.47, *SD* = 12.70; *t* = -2.818 (100); *p* = .006) as compared to those in the

Table 11

Kindergarten Mean Pretest, Posttest, and Change Scores for the Matched Groups

	Treatment (N = 51)		Control (N = 51)	
	Mean	SD	Mean	SD
Fall LNF	8.67	9.18	9.08	8.52
Fall ISF	4.67	3.70	4.98	3.74
Spring LNF	39.90	14.24	44.45	9.65
Spring NWF	25.10	14.27	32.47	12.07
LNF Change	31.24	16.81	35.69	10.03

treatment group (*M* =25.10, *SD* = 14.27). The magnitude of the difference between the means was small (*eta squared* = .073).

Based on LNF results for this sample of kindergarten students at risk for reading failure, the null hypothesis that kindergarten students at risk for reading failure who participated in Headsprout would perform similarly to kindergarten students at risk for reading failure who did not participate in Headsprout on the LNF measure cannot be rejected. When NWF results for this sample of kindergarten students are considered, however, the null hypothesis that kindergarten students at risk for reading failure who participated in Headsprout would perform similarly to kindergarten students at risk for reading failure who did not participate in Headsprout on spring DIBELS NWF measures can be rejected, and a comparison of the means of the two groups suggests that the kindergarten students who did not participate in Headsprout performed better than those who did. The magnitude of the difference between the means was small (*eta squared* = .073).

Research Question 2 (First Grade)

In order to address the first grade research question posed in this study, two hypotheses, a null hypothesis and an treatment hypothesis, were considered. The null hypothesis was that first grade students at risk for reading failure who participated in Headsprout would perform similarly to first grade students at risk for reading failure who did not participate in Headsprout on the spring DIBELS ORF measure and the CRCT. The treatment hypothesis was that first grade students at risk for reading failure who participated in Headsprout would have scores on the spring DIBELS ORF measure and the CRCT that were different from those of first grade students at risk for reading failure who did not participate in Headsprout.

Prior to statistical analysis, the first grade data were examined through the Statistical Package for the Social Sciences (IBM Inc., 2010) for assumptions of normality of sampling distributions, homogeneity of variance and homogeneity of variance. Fall NWF scores were normally distributed for the treatment group (see figure 11) but not for the control group (see figure 12). The distribution of scores for the treatment group had a mean of 15.74 and a standard deviation of 7.30, a median of 16, and a mode of 23. The resulting distribution was unimodal with a slight negative skew (-.218). It had a maximum score of 35. The distribution of scores in the control group was unimodal and positively skewed (.703). It had a mean of 22.10 and a standard deviation of 11.57. The median was 20, and the mode was 17. The maximum score for this distribution was 55.

The data for the spring ORF variable (see Tables 10 & 11) were not normally distributed for either the treatment group or the control group. The distribution of spring ORF scores for the treatment group had a pooled mean of 40.68. Standard deviations for the original data and the multiple imputation data sets ranged from 22.07 to 22.64. All score distributions were unimodal and positively skewed, with median scores of 38 and modal scores of 52. The distribution of spring ORF scores for the control group had a pooled mean of 46.15. Standard deviations for the original data and the multiple imputation data sets ranged from 26.47 to 29.20. All score distributions were unimodal and positively skewed, with median scores that ranged from 39.62 to 43 and modal scores of 38.

Figure 11

First Grade Fall NWF Scores for the Ttreatment (HS: 1) Group

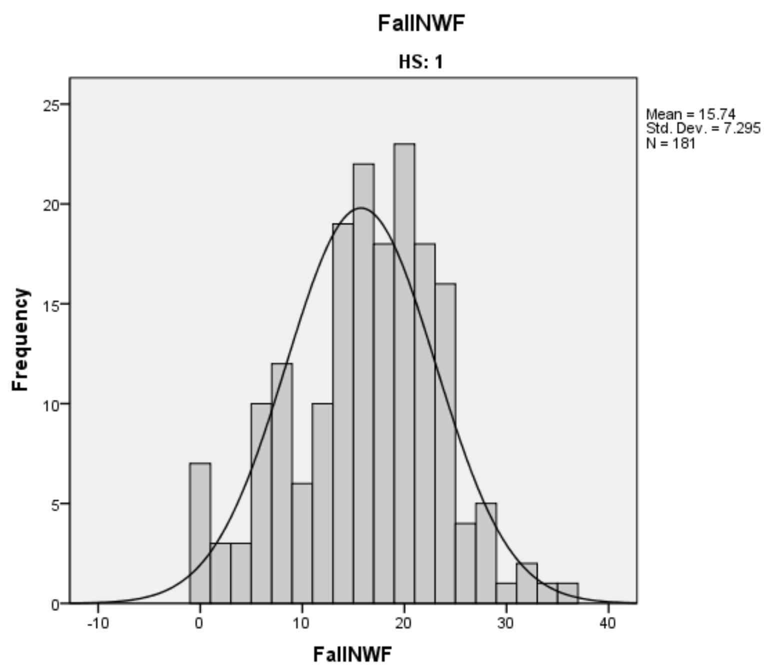
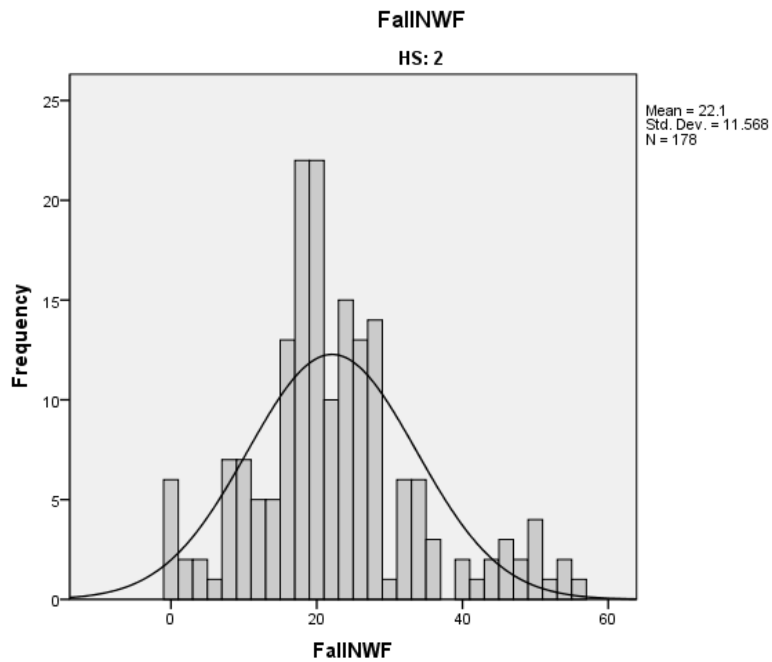


Figure 12

First Grade Fall NWF Scores for the Control (HS: 2) group

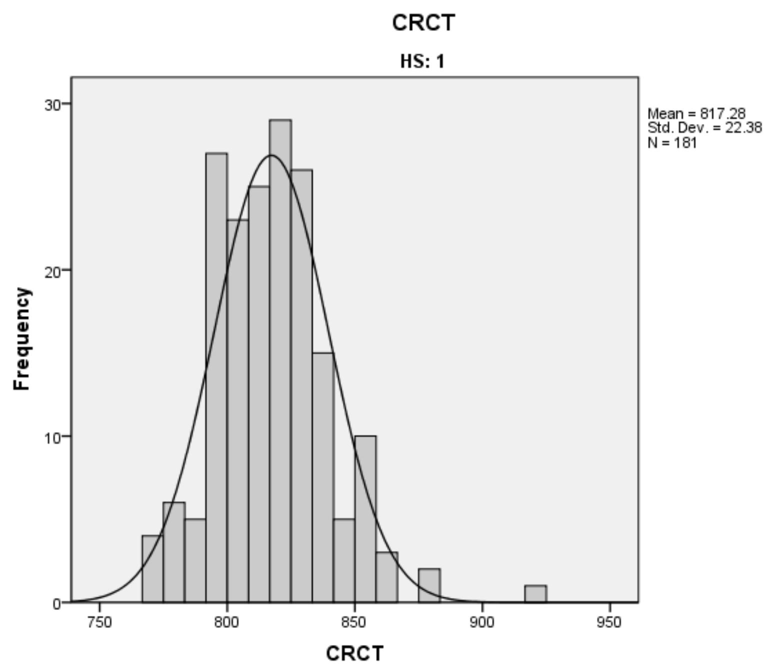


CRCT scores were not normally distributed for either the treatment group or the control group (see figures 13 and 14). The distribution of CRCT scores for the treatment group had a mean of 817.28, a median of 817, and a mode of 827, resulting in a positively skewed (.703), markedly peaked (1.88), unimodal distribution. The distribution of spring CRCT scores for the control group had a mean of 825.66, a median of 820, and a mode of 817, resulting in a positively skewed (.879), peaked (1.34) unimodal distribution.

The assumption of homogeneity of variance was not met for any of the data based on the results of Levene's test. For fall NWF scores, F (1, 357) = 18.35, p = .00; spring ORF scores, F (1, 308) = 10.17, p = .002; and CRCT scores, F (1, 357) = 12.07, p = .001,

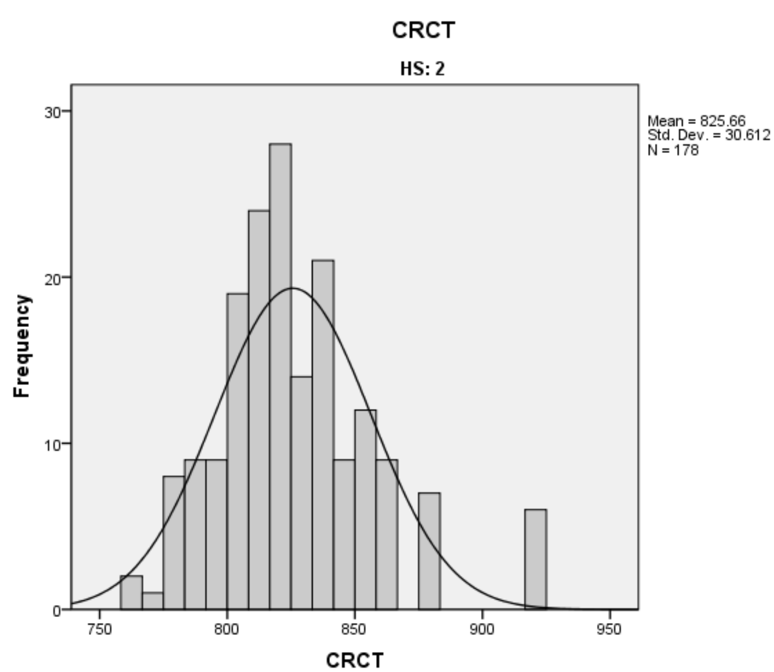
Figure 13

First Grade CRCT Scores for the Treatment (HS: 1) Group



F*igure 14*

First Grade CRCT Scores for the Control (HS: 2) Group



indicating that the variances between the treatment and control groups were significantly different for all variables. Therefore, the proposed analysis using ANCOVA was not completed.

The proposed covariates, age in months and fall NWF scores, were tested to ensure that they were independent of the independent variable. That is, were the subjects' age in months and fall NWF scores roughly equal across the treatment and control groups? Results of ANOVAs indicated that age in months was not significantly different across the treatment and control groups ($p > .05$) for both spring ORF scores and CRCT scores, indicating that age in months was an appropriate covariate to include in the analysis. Fall NWF scores were significantly different across the treatment and control groups ($p < .01$) for both spring ORF scores and CRCT scores, indicating that fall NWF scores were not an appropriate covariate to include in an ANCOVA.

Because all the data were not normally distributed, and the variable, fall NWF score, intended to control for initial early reading skill differences among students was different across the treatment and control groups, the proposed analysis using ANCOVA was not completed. Instead, the treatment group was redefined to include first grade students at risk for reading failure who completed at least 64 (80%) of the 80 Headsprout episodes. One student was eliminated from the treatment group due to missing spring NWF data. The resulting in a treatment group included 79 first grade students who had completed at least 64 Headsprout episodes. A matched control group of first grade students who had not participated in Headsprout was selected from the pool of first grade students at risk for reading failure based on the following criteria: 1) Fall ISF score within 5 points, Fall LNF score within 5 points, age in months within 3 months, 4) an exact match on support program participation, and 5) an exact match on either socioeconomic status designation or race designation. Three students in the treatment group could not be matched, and they were removed from the data set. The resulting matched data set included 76 first grade students in treatment group and 76 first grade students in the control group. No data were missing from the data set.

Revised Analysis.

The final sample of first grade students consisted of two groups of students. The treatment group consisted of 76 students (49 boys, 27 girls), and the control group consisted of 76 students (44 boys, 32 girls).). These students came from 13 schools within the district. The number of students from each school ranged from as few as 6 students at a school to as many as 30 students from a school, and 111 of the students came from just 7 schools. Students completed an average of 1.6 Headsprout episodes each week. There was no significant difference (t = -.829 (150), p = .35) between the groups in mean age: 78.28 months (*SD* = 3.35) and 77.76 months (*SD* = 3.45) for the treatment and control groups, respectively. The students came from diverse socio-cultural backgrounds, including: white (65.8% treatment, 60.5% control), black (18.4% treatment, 28.9% control), Hispanic (5.3% treatment, 3.9% control), multiracial (9.2% treatment, 5.3% control), and American Indian/ Alaskan Native (1.3% treatment, 1.3% control). The students also participated in a variety of programs at their schools including an Early Intervention Program (EIP; 31.6% treatment, 31.6% control) special education (6.6% treatment, 6.6% control), and English for Speakers of Other Languages (ESOL; 1.3% treatment, 1.3% control). The percentage of treatment group and control group students who qualified for free or reduced lunch was roughly equal (46.1% treatment, 43.4%

control; X^2 =.027, p = .87). The treatment and control groups were also similar on the pretest measure, NWF. For NWF, t (150) = -.384, p = .702. The treatment group had an NWF mean of 16.36 (SD = 7.00), and the control group had an NWF mean of 16.80 (SD = 7.38).

The data set was examined through SPSS (IBM Inc., 2010) for assumptions of normality of sampling distributions and homogeneity of variance. Spring ORF scores CRCT scores were not normally distributed for the treatment group (see Figure 15) based on results of Kolmogorov-Smirnov, treatment group spring ORF D (76) = .124, p = .006. The distribution spring ORF scores for the treatment group had a mean of 46.14 and a standard deviation of 23.21. The median score was 44 and the modal scores of 45 and 52, resulting in a bimodal, positively skewed (.909), and peaked (.903) distribution. They were normally distributed for the control group (see figure 16), spring ORF D (76) = .064, p = .200.

CRCT scores were normally distributed for the treatment group (see figure 17), D (76) = .100, p = .060; but not for the control group (see figure 18), D (76) = .122, p = .007. Control group CRCT scores had a mean of 822.78 with a standard deviation of 27.61. The median score was 817 and the modal score was 840, resulting in a unimodal, positively skewed (.810) and peaked (.921) distribution. The assumption of homogeneity of variance was met for the spring ORF data based on the results of Levene's test, F (1, 150) = .075, p = .784), but not for CRCT scores, F (1, 150) = 8.67, p = .004, indicating that the variances between the treatment and control groups were not significantly different for spring ORF scores but were significantly different for CRCT scores.

Figure 15

First Grade Spring ORF Scores for the Matched Treatment (HS: 1) Group

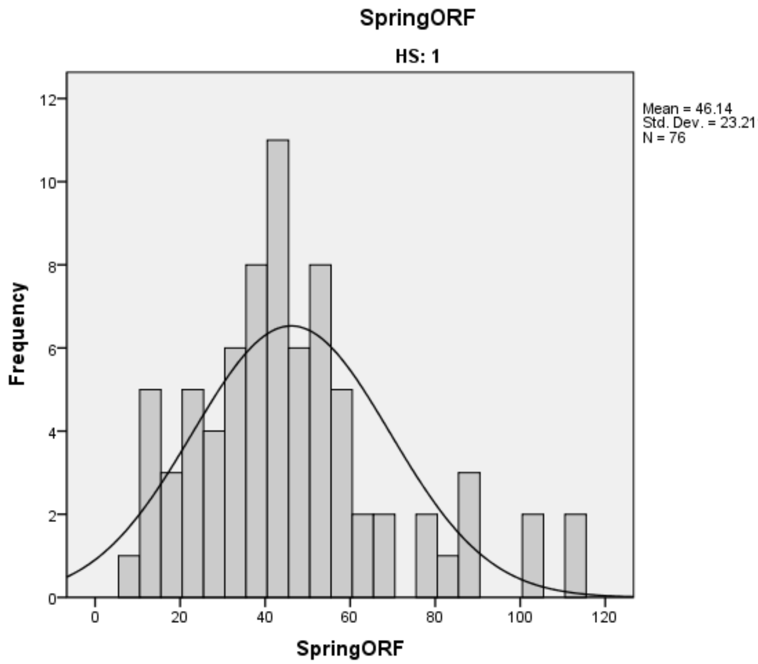


Figure 16

First Grade Spring ORF Scores for the Matched Control (HS: 2) Group

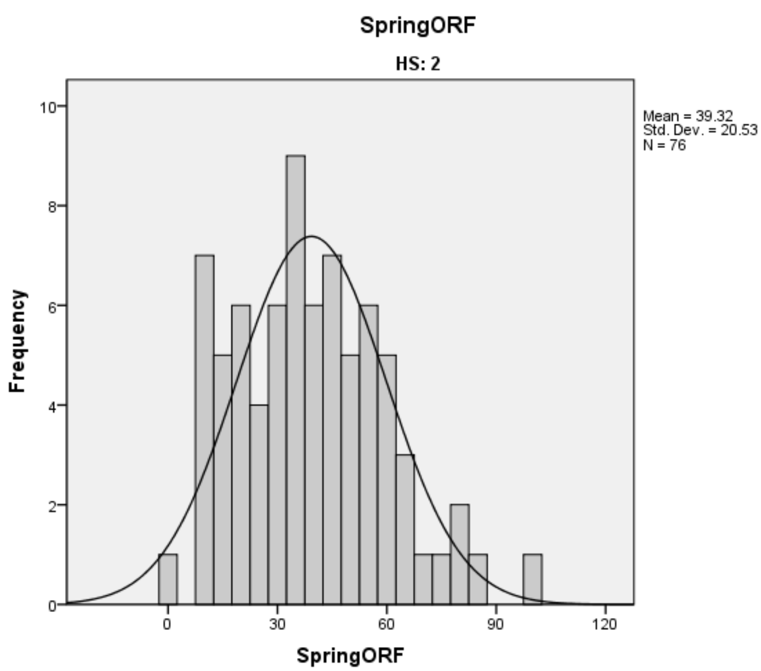


Figure 17

First Grade CRCT Scores for the Matched Treatment (HS: 1) Group

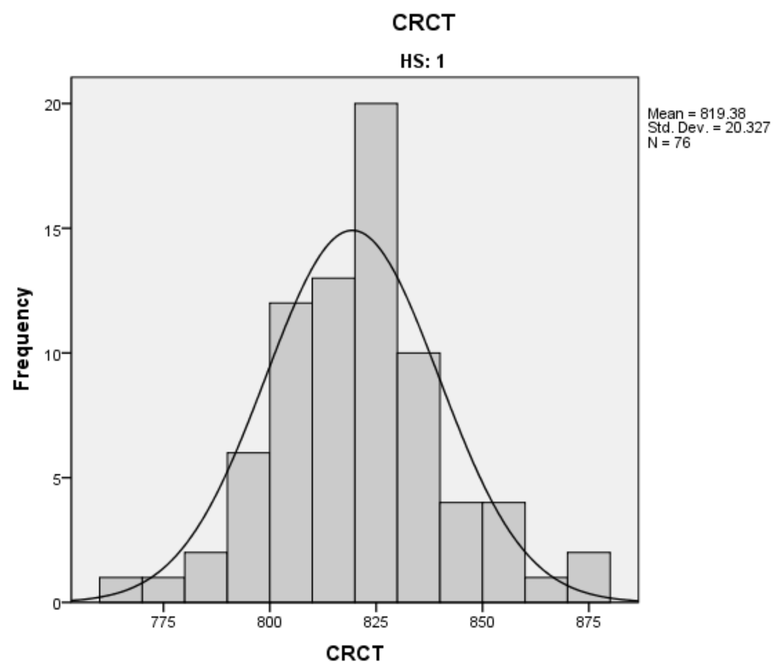
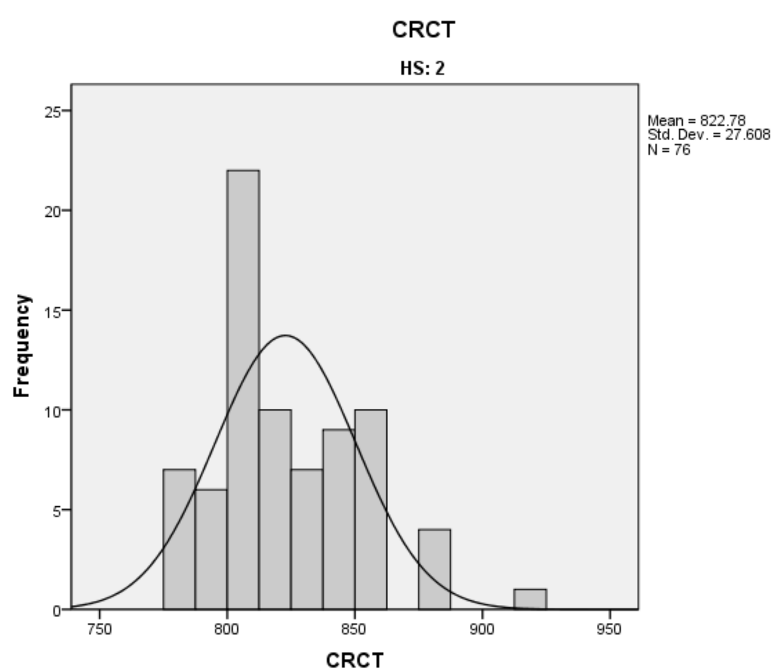


Figure 18

First Grade CRCT Scores for the Matched Control (HS: 2) Group



All data were analyzed using independent samples *t*-tests because *t*-tests are robust to violations of normality when group sizes are relatively equal (Tabachnick & Fidell, 2007), and the control and treatment groups in this data set were exactly equal. The more stringent value of *t* provided in SPSS for use with data that do not meet the assumption of homogeneity of regression slopes was used where appropriate.

Spring ORF and CRCT Results. Table 12 presents mean pretest (Fall NWF) and posttest (Spring ORF and CRCT) scores for students in the treatment and control groups. Independent samples *t*-tests were performed to determine whether there were any significant differences between the groups at pretest in terms of age, gender, socio-economic status, mean fall LNF scores, and mean fall NWF scores. No significant differences between the groups were found.

Independent samples *t*-tests were also conducted to compare the spring ORF and CRCT scores of first grade students in the treatment group with those in the control group. There was no significant differences between groups on Spring ORF scores (t = 1.92 (150); p = .057). There was not a significant difference between groups on CRCT scores (t = -.863 (150); p = .390). Based on results for this sample of first grade students at risk for reading failure, the null hypothesis that first grade students at risk for reading failure who participated in Headsprout would perform similarly to first grade students at risk for reading failure who did not participate in Headsprout on the spring DIBELS ORF measure and the CRCT cannot be rejected.

Table 12

Mean Pretest and Posttest Scores for Matched First Grade Students

	Treatment (N = 76)		Control (N = 76)	
	Mean	SD	Mean	SD
Fall NWF	16.36	7.00	16.80	7.38
Spring ORF	46.14	23.21	39.32	20.53
CRCT	819.38	20.33	822.78	27.61

CHAPTER 5

DISCUSSION

This applied research, which analyzed an existing data set collected by a school system, examined the benefit of providing Headsprout as a standard protocol, Tier 2 intervention across the school system to supplement the general education program's balanced literacy curriculum for kindergarten and first grade students at risk for reading failure. Headsprout, a type of CAI, provides a planned curriculum of instruction in phonemic awareness, phonics, reading fluency, vocabulary, and comprehension that conforms to what is currently known about effective reading instruction (FCRR, 2003). For both kindergarten and first grade students, comparisons were made between treatment group students at risk for reading failure who used Headsprout in addition to their regular classroom balanced literacy instruction and control group students at risk for reading failure who did not use Headsprout. The treatment and control groups at each grade level were similar in terms of student age, socioeconomic status, and race. There were no significant differences between treatment and control groups at either grade level on pretest measures of early literacy skills.

Impact of Headsprout on Early Reading Skills

The kindergarten students included in this study were identified as at risk for reading failure based on fall DIBELS ISF and LNF scores, and teacher judgment. Both the treatment and control groups of students made similar gains over the school year on the DIBELS LNF measure. The mean spring LNF scores for both the treatment and control groups fell within the low risk category based on the DIBELS end of year criterion score for LNF. In contrast, on the end of year DIBELS NWF measure, the control group significantly outperformed the treatment group. The effect size associated with the difference was small. Mean end of year DIBELS NWF scores for both the treatment and control groups fell within the low risk category based on the DIBELS end of year criterion score for NWF.

The first grade students included in this study were identified as at risk for reading failure based on fall DIBELS LNF and NWF scores, and teacher judgment. Both treatment and control groups made gains over the school year. At pretest, mean DIBELS fall NWF scores for both the treatment and control groups fell within the some risk category based on DIBELS beginning of the year criteria. At the end of the school year, DIBELS spring ORF scores for the treatment group fell within the low risk category and control group DIBELS spring ORF scores improved to within .68 point of the low risk category based on the DIBELS end of year criteria. Mean CRCT scores for both the treatment and control groups surpassed the cut score of 800 that indicates a student met the standard on the CRCT.

This research found no consistent patterns of significant differences between treatment and control group students at posttest. Both treatment and control group students improved their early literacy skills enough that they were at low risk for reading failure by the end of the study year. Participation in Headsprout did not appear to provide added benefits to kindergarten or first grade students at risk for reading failure beyond the benefits of classroom instruction using a balanced literacy approach. In fact, there was evidence from the DIBELS NWF task that kindergarten students who did not participate in Headsprout had better early literacy outcomes than those who participated in Headsprout. The results of the present study are consistent with the findings of recent, large scale randomized experiments that investigated the effectiveness of Headsprout with first grade students (e.g. Campuzano et al., 2009; Dynarski et al., 2007) and found that participating in Headsprout did not appear to provide meaningful educational benefit to students. The findings of the current study are surprising given that the students included in this study who participated in Headsprout received as much as 26 hours of additional reading instruction beyond that provided by the reading curriculum in use within the school system.

There are several potential explanations for the finding that Headsprout did not appear to benefit the students at risk for reading failure in this sample. First, Headsprout is designed as a reading curriculum that happens to be delivered to students via the internet rather than as a reading intervention. Both CLT (Kalyuga et al., 2003; Pollock, et al., 2002) and research in the area of early reading intervention (Cooke et al., 2010; McIntyre et al., 2005; Wanzek & Vaughn, 2007) suggest that effective intervention for students at risk for reading failure may need to be qualitatively different from the type of balanced literacy instruction provided to typically developing readers. Specifically, Cavanaugh et al., (2004) suggest that effective reading intervention should be individualized to the student, delivered individually or in small groups, occur several times a week for a minimum of 15 minutes at a time, and last across several months. While Headsprout, when implemented as intended by the publisher, meets four of these conditions (e.g. it is delivered individually, it occurs several times a week, it provides 20 minute episodes, and it lasts over several months), it does not specifically tailor instruction to meet individual student needs in two important ways: 1) Each student begins with the first Headsprout episode regardless of their existing reading skill level, and 2) Headsprout does not provide additional instruction for students who do not exhibit skill mastery on the benchmark assessments that are administered after each five episodes completed. The only option for remediation is for the student to repeat the same series of five episodes.

Secondly, while a great deal of planning and program evaluation occurred during Headsprout's development (Layng et al., 2003; Layng et al. 2004b), and the program has been recognized as meeting the NRP's (2000) guidelines for evidence based instruction in early reading (FCRR, 2003) the instructional approach it employs for teaching the relationships between letters and sounds is best described as analogy, or onset-rime phonics. The reading curriculum utilized by the school system that provided the data for this research employed a synthetic phonics approach when teaching about the relationships between letters and sounds. The majority of the supplemental programs favored by individual schools also employed a synthetic phonics approach to instruction. This miss-match between the instructional approach used for general classroom instruction and the instructional approach used by the Headsprout program may have contributed to the Headsprout's apparent lack of impact on the early literacy skills of students at risk for reading failure in this research. CLT (Kalyuga et al., 2003; Pollock, et al., 2002) implies that using two different instructional approaches at the same time when teaching early reading skills to struggling readers could result in confusion about early reading skills rather than increased mastery of the skills.

Students at risk for reading failure who participated in Headsprout in this study did so outside of their regularly scheduled reading instruction. While this approach has the advantage of increasing the total amount of reading instruction that the students received, it also meant that there was an opportunity cost associated with participation in Headsprout. That is, the students who participated in Headsprout did not receive some other type of instruction because they were participating in Headsprout. This may have impacted the results of this study in two ways. First, it appears that teachers may have opted not to send students to participate in Headsprout because they were missing other instruction. The treatment group students in this study completed, on average, just 1.6 episodes a week. The Headsprout Corporation recommends that students complete three Headsprout episodes each week. Research in the area of reading intervention suggests that students participate in an intervention several times a week (Cavanaugh et al., 2004). This suggests that even if Headsprout is effective as a reading intervention, the students at risk for reading failure in this sample may not have participated in Headsprout with enough frequency for the program to have been effective. In other words, Headsprout, as implemented within this school system may not have been an intensive intervention (Cavanaugh et al., 2004). Secondly, in the case of kindergarten students in particular, it is possible that students who did not participate in Headsprout received other small group or individual intervention in the area of early reading skills during the times that Headsprout was offered. This is because in the school system where this research occurred, each kindergarten classroom has both a teacher and a paraprofessional. Having two adults in each classroom, may have increased the ability of the adults to provide meaningful amounts of remedial instruction that was specifically targeted to individual student needs.

Finally, there is an assumption, supported by some early research in the area of CAI, that providing the amounts of practice in early reading skills that students at risk for reading failure via computer inherently results in levels of student engagement required for learning to occur (Hall et al., 2000; Sivin-Kachala & Bialo, 1998; Soe et al., 2000). Since that research was published, the availability of computers for students to use both at home and at school has increased dramatically (Gray, Thomas, & Lewis, 2010), and the sophistication of the software that students use at home is often greater than the sophistication of the software available to them at school (Lovell & Phillips, 2009). Therefore, the conventional wisdom that CAI is inherently engaging to students may no longer be true. It is possible that even well designed CAI like Headsprout, is not inherently motivating enough to overcome the frustration that students who are at risk for reading failure often experience while trying to master early reading skills.

In contrast to the finding of Campuzano et al. (2009), and Dynarski et al. (2007), and the current study, other researchers (e.g. Headsprout, 2007; Clarfield, 2006; Clarfield and Stoner, 2005) have reported that students benefited from using the program. It is difficult to evaluate the practical implications of the findings reported in these studies for a variety of reasons. Technical reports of research conducted at individual schools and published by the Headsprout Corporation (2007) on their website, reported statistically significant differences in standardized test scores favoring students who participated in Headsprout at two separate elementary schools. In New York City, kindergarten and first grade students at risk for reading failure who completed at least 70 of the 80 available Headsprout episodes outperformed kindergarten and first grade students who did not participate in Headsprout by 3 to 8 months as measured in grade equivalent scores. In Los Angeles, first grade students who participated in Headsprout outperformed first grade students who did not participate in Headsprout by 5.7 Normal Curve Equivalent points on a standardized test of reading. The Headsprout Corporation did not report the effect sizes associated with the differences, which makes it difficult to evaluate the practical impact that Headsprout had on student achievement.

Clarfield and Stoner (2005) used single subject methodology to demonstrate that three students with ADHD demonstrated less off task behavior and made greater gains in reading fluency when using Headsprout as compared to when they participated in small group and independent reading work. These findings are limited by the small number of participants and the fact that the intervention occurred outside the classroom during non-instructional time. It is not known whether the gains in reading that these students made were specific to the Headsprout program, or simply due to the fact that the students received supplemental instruction in reading. Finally, Clarfield (2006) demonstrated that kindergarten students with no risk factors (N = 16), kindergarten students at risk for reading problems (N = 11), kindergarten students at risk for behavior problems (N = 8), and kindergarten students at risk for both reading and behavior problems (N = 7) who participated in Headsprout made greater gains in reading than students who participated in Lexia (N = 18, 9, 8, 9, respectively), but not all gains reached statistical significance and effect sizes were small. No comparison to students who did not receive CAI was made. While independent reviews (FCRR, 2003) of Headsprout report that the program was designed to conform to what is known about best practices in early reading instruction, it would be difficult to conclude that Headsprout has an educationally meaningful, positive effect on student achievement in reading based on currently available research.

Implications

The use of CAI as a standard protocol Tier 2 intervention has often been cited as a promising real-world application of CAI within school schools because of its apparent potential to address often cited barriers to the successful implementation of RTI such as lack of intervention resources and lack of resources for instruction (Response to Intervention Adoption Survey, 2010). Providing standard protocol, supplemental instruction at the Tier 2 level that requires less direct instruction from teachers, like the Headsprout program was intended to do in this study, is one of the most common applications of CAI in schools (Cheung & Slavin, 2011). Other programs, like PLATO and WERP, are also examples of supplemental CAI designed to provide additional instruction to students at their assessed levels of need in the area of reading (Campuzano et al., 2009; Dynarski et al., 2007). Similar to the results of this research, a recent meta-analysis by Cheung and Slavin (2011) that included 57 evaluations of supplemental CAI programs in reading, revealed a combined effect size of only .11, which is considered small. Further, when the impact of supplemental CAI on student reading achievement was analyzed by student grade level, studies of kindergarten students had a combined effect size of only .15 and studies with elementary aged students had a combined effect size of only .10. Cheung and Slavin (2011) concluded that "the types of supplementary computer assisted instruction programs that have dominated the classroom use of education technology in the past few decades are not producing educationally meaningful results in reading" (p. 15). The types of CAI commonly in use with kindergarten and first grade students at risk for reading failure do not appear to be fulfilling their initial promise as effective Tier 2 standard protocol interventions.

This lack of educational effect may be inherent to the types of CAI currently available in the area of early reading instruction and intervention. However, it is also important to consider the issue of treatment integrity when evaluating the effectiveness of any intervention, including CAI. Within the body of research regarding CAI for early reading (e.g. Campuzano et al., 2009; Macaruso & Rodman, 2011), as with the current study, a consistent theme is lack of sufficient time for students within the intervention when the CAI is implemented without significant levels of support from outside personnel, such as researchers. Often few students within the potential sample of treatment groups students spend the amount of time using the CAI recommended by its publisher, and few students within the treatment groups of these studies spend the amount of time using the CAI that would meet the definition of an intensive intervention (Cavanaugh, Kim, Wanzek & Vaughn, 2004). In other words, for some reason, when teachers are given control of CAI programs, they are not choosing to have students participate in CAI. One potential reason for this is lack of appropriate professional learning opportunities in the area of CAI in general and with specific CAI programs in particular. For example, in the current study, Headsprout provided only four hours of professional development about Headsprout, and this training was attended by only one or two representatives from each elementary school within the system. In effect, teachers were asked to have their students, during instructional time, participate in a program when the only information they had about the program was its name. Similarly, in the Dynarski, et al. (2007) study, a majority of first grade teachers asked to implement CAI in their classroom reported that they felt unprepared to use the program after the single day of training that was provided.

It is also important to recognize the difference between programs and interventions that are designed to reflect what is known about effective early reading instruction and programs and interventions that have demonstrated significant and meaningful positive impact on the reading achievement of students in research studies. Headsprout provides an example of the potential pitfall of assuming that programs that should be effective are effective. An independent review of Headsprout by the FCRR (2003) concluded that "the content and design of Headsprout Early Reading reflect scientific research with an abundance of instructional strategies in phonemic awareness, phonics, fluency, vocabulary and comprehension" (p. 4). On a cautionary note, FCRR did indicate that "whether use of the program produces gains in reading that re independent of, or in addition to, the gains that might result from classroom instruction is yet to be determined in studies that employ appropriate control groups" (p. 4). The currently available research evidence suggests that Headsprout does not deliver on its promise.

Future Directions

While recent research suggests that supplemental CAI for young students at risk for reading failure, as currently implemented, does not produce meaningful educational results (Slavin & Cheung, 2011), information available in the existing literature does provide information regarding practical ways to potentially increase the effectiveness of CAI. For example, Cheung and Slavin (2011) reported that CAI was most effective when it was an integral part of the reading curriculum rather than an addition to an existing curriculum. This finding is consistent with some of the earliest research on CAI in which Atkinson and his colleagues (1970) concluded that CAI could not replace a theory-driven curriculum taught by a teacher, but it could serve as a useful supplement to the curriculum if it was complementary to the curriculum. The Canadian program, ABRACADABRA (Comaskey et al., 2009; Hipps et al., n.d.; Savage et al., 2009), represents one promising line of research in this area. Additional investigation of both instructional software and intervention software directly tied to balanced literacy curricula represents a promising area for future research.

Further investigation in the area of treatment integrity is also warranted. Given that one of the strongest arguments in favor of CAI for students at risk for reading failure is its potential to provide effective intervention without requiring large amounts of resources in terms of teacher time, it is especially important to understand why teachers are not implementing CAI with their students. While it is plausible that providing teachers extensive professional learning regarding the integration of CAI into the reading curriculum and the implementation of CAI with students may increase the likelihood that CAI will have positive effects on student achievement (Cheung & Slavin, 2011; Macaruso & Rodman, 2011; Macaruso & Walker, 2008), it is not clear that this is true. There may be other, not yet identified, reasons that teachers frequently choose not to implement CAI with their students.

The relationship between CAI and student engagement also warrants further investigation. Changes in the amount and type of technology that many students routinely access outside of the school setting may have changed previously identified positive relationships between technology use and student engagement. Additionally, as technology continues to evolve, the term CAI no longer refers only to software programs, but also includes new forms of technology such as interactive white boards, which may fundamentally change the way technology is implemented in schools (Cheung & Slavin, 2011). The findings of current research in the area of CAI caution that these innovations should not be accepted as substantially beneficial to student learning without supporting evidence generated through research.

Limitations

The main goal of this research was to investigate the use of CAI in typical classroom settings. Like all research, this study is not without its limitations, which are related to the decision to utilize an existing data set collected by the school system itself. Consequently, the final sample sizes of the treatment and control groups represent a relatively small proportion of the students at risk for reading failure actually enrolled in the Headsprout program during the study year. For example, in the first grade sample, approximately 359 first grade students at risk for reading failure were enrolled in the Headsprout program for the full study year, but only 79 first grade students completed at least 80% of the program episodes. While reduced numbers of participants resulted in reduced statistical power, it was important to ensure that students in the treatment group participated in a sufficient number of Headsprout episodes to achieve an educational effect if one was present. Another potential limitation of this study is the fact that students at risk for reading failure were not randomly assigned to treatment and control groups, and in fact, the students at greatest risk were assigned to the treatment group. This is common practice within school systems where resources are often limited. Efforts were made to account for this difference in group assignment by closely matching control group students to treatment group students based on objective criteria.

Conclusion

The skill of reading with comprehension is a central issue in education across the country. The US government has emphasized the need for increased literacy rates among school children by creating reading initiatives such as the No Child Left Behind Act (2001). The national goal that every child should be able to read by the end of third grade has encouraged schools to actively search for effective reading interventions for diverse populations of students. As educators continue to search for interventions to improve reading outcomes for their students, it is especially important to avoid the pitfall of assuming that an intervention that logic dictates should work, does work. Despite the fact that the kindergarten and first grade students in this study who participated in Headsprout did not appear to benefit from the program beyond the benefit provided through general classroom instruction, it is important to note, that many of the students at risk for reading failure in both the treatment groups and the control groups demonstrated increased reading skill on post-test measures. In fact, many students were able to improve their early literacy skills enough to move from the DIBELS some risk category at the beginning of the school year to the DIBELS low risk category by the end of the school year. This is an affirmation of the balanced literacy curriculum and effective teaching in the public school system that provided the data for this study.

References

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education, 23,* 300-316.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing.* Washington, DC: AERA Publications.

Association for Educational Communications and Technology. (1977). Educational technology: Definition and glossary of terms (Vol. 1). Washington, DC: Association for Educational Communications and Technology.

Atkinson, R.C., Fletcher, J. D., Chetin, H. C., & Stauffer, C. M. (1970). Instruction in initial reading under computer control: The Stanford project (Technical Report No. 158). Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, CA. Retrieved from: http://suppes-corpus.stanford.edu/techreports/IMSSS_158.pdf

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E., & Thomas Beck, C. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review, 37*, 18-37.

Baraldi, A. N. & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*, 5-37. doi: 10.1016/j.jsp.2009.10.001

Barker, T. A. & Torgesen, J. K. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. *Journal of Educational Computing Research, 13*, 89-103. doi: 10.2190/TH3M-BTP7-JEJ5-JFNJ

Bauseman, K. L., Cassady, J. C., Smith, L. L., & Stroud, J. C. (2005). Kindergarten literacy Achievement: The effects of the PLATO integrated learning system. *Reading Research and Instruction, 44*, 49-60.

Beck, I., Farr, R., Strickland, D., Flor Ada, A., Brechtel, M., McKeown, M., Roser, N., & Yopp, H. (2003). *Harcourt Trophies: A Harcourt Reading/ Language Arts Program*. Orlando: Harcourt.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing Words to Life: Robust Vocabulary Instruction Solving Problems in the Teaching of Literacy*. New York: The Guilford Press.

Bergan, J. R., & Kratochwill, T. R. (1990). *Behavioral Consultation and Therapy*. New York: Plenum Press.

Berninger, V. W., Vermeulen, K., Abbott, R. D., McCutchen, D., Cotton, S., Cude, J., Dorn, S., & Sharon, T. (2003). Comparison of three approaches to supplementary reading instruction for low-achieving second-grade readers. *Language, Speech, and Hearing Services in Schools, 34*, 101-116. doi: 0161-1461/03/3402-0101

Blachowicz, C. L. Z., Bates, A., Berne, J., Bridgman, T., Chaney, J., & Perney, J. (2009). Technology and at risk learners and their classrooms. *Reading Psychology, 30*, 387-411. doi: 10.1080/02702710902733576

Blok, H., Oostdam, R., Otter, M., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. Review of Educational Research, 72, 101-130. Retrieved from: http://www.jstor.org/stable/3516075

Burke, M. D. & Hagan-Burke, S. (2007). Concurrent criterion-related validity of early literacy indicators for middle of first grade. *Assessment for Effective Intervention*, *32*, 66-77. doi: 10.1177/15345084070320020401

Burke, M. D., Hagan-Burke, S., Kwok, O., & Parker, R. (2009). Predictive validity of early literacy Indicators from the middle of kindergarten to second grade. *The Journal of Special Education*, *42*, 209-226. doi: 10.1177/0022466907313347

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research implemented models. *Journal of Psychoeducational Assessment*, *23*, 381-394. doi: 10.1177/073428290502300406

Campuzano, L., Dynarski, M., Agodini, R., Rall, K., & Pendleton, A. (2009). *Effectiveness of Reading and Mathematics Software Products: Findings from two student cohorts*. Washington, D.C.: Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee

Carlson, J. F., Romhild, A., McCormick, C., Chin, K, & Geisinger, K. F. (2010). *Evaluating reading tests for the state of Florida: Dynamic Indicators of Basic Early Literacy Skills, 6th Edition (DIBELS) and Florida Assessments for Instruction in Reading (FAIR)*. Lincoln, NE: Buros Center for Testing.

Cassady, J. C. & Smith, L. L. (2003) The impact of a reading-focused integrated learning system on phonological awareness in kindergarten. *Journal of Literacy Research*, *35*, 947-964. doi: 10.1207/s15548430jlr3504_2

Cassady, J. C., & Smith, L. L. (2005). The impact of a structured integrated learning system on first grade students' reading gains. *Reading and Writing Quarterly, 21*, 361–376. doi: 10.1080/10573560591002277

Cavanaugh, C. L., Kim, A., Wanzek, J. & Vaughn, S. (2004). Kindergarten reading interventions for at risk students: Twenty years of research. *Learning Disabilities: A Contemporary Journal, 2*, 9-21.

Chall, J. (1976). The great debate: Ten years later, with a modest proposal for reading stages. Paper presented at the Conference on Theory and Practice of Beginning Reading Instruction University of Pittsburg, Learning Research and Development Center. Retrieved from: http://www.eric.ed.gov/PDFS/ED155617.pdf

Chambers, B., Abrami, P., Slavin, R., & Madden, N. (2011). A three-tier model of reading instruction supported by technology. *International Journal of Innovation and Learning, 9*, 286-297.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332. Retrieved from: http://ro.uow.edu.au/edupapers/128

Chera, P., & Wood, C. (2003). Animated multimedia talking books can promote phonological awareness in children beginning to read. *Learning and Instruction, 13*, 33-52.

Cheung, A. C. K., & Slavin, R. E. (2011). The Effectiveness of Education Technology for Enhancing Reading Achievement: A Meta-Analysis. Best Evidence Encyclopedia (BEE). Baltimore: Center for Data-Driven Reform in Education (CDDRE). Retrieved from: www.bestevidence.org

Clarfield, J. (2006). Examining the efficacy of two computerized reading programs for kindergarten students at risk for reading and behavior problems (Doctoral dissertation). Retrieved from: ProQuest Dissertations & Theses. (Publication No. AAT 3216946).

Clarfield, J., & Stoner, G. (2005). The effects of computerized reading instruction on the academic performance of students identified with ADHD. *School Psychology Review, 34*, 246-254.

Comaskey, E., Savage, R., & Abrami, P. (2009). A randomized efficacy study of web based analytic and synthetic programmes among disadvantaged urban kindergarten children. *Journal of Research in Reading, 32,* 92-108. doi: 10.1111/j.1467-9817.2008.01383.x

Compton, D. L., Appleton, A. C., & Hosp, M. K. (2004). Exploring the relationship between text-leveling systems and reading accuracy and fluency in second-grade students who are average and poor decoders. *Learning Disabilities Research & Practice, 19*, 176-184.

Cooke, N. L., Kretlow, A. G., & Helf, S. (2010). Supplemental reading help for kindergarten students: How early should you start? *Preventing School Failure, 54*, 137-144. doi: 10.1080/10459880903492924

Cusumano, D. L. (2007). Is it working? An overview of curriculum based measurement and its uses for assessing instructional, intervention, or program effectiveness. *The Behavior Analyst Today, 8,* 24-34.

Desrochers, M. N., & Gentry, D. G. (2004). Effective use of computers in instruction. In Moran, D. J. (Eds.) *Evidence-Based Educational Methods* (pp. 127-138). St. Louis: Elsevier Science & Technology 127-138.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., Means, B., Murphy, R., Penuel, W., Javitz, H., Emery, D., & Sussex, W. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort.* Washington, D.C.: Institute of Education Sciences. Retrieved from: http://ies.ed.gov/ncee

Ehri, L. C. (2004). Teaching phonemic awareness and phonics: An explanation of the National Reading Panel Meta-analyses. In P. McCardle & V. Chhabra (Eds.), *The Voice of Evidence in Reading Research*. Baltimore: Paul H. Brooks Publishing.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9,* 167-188.

Englert, C. S., Zhao, Y., Collings, N., & Romig, N. (2005). Learning to read words: The effects of internet-based software on the improvement of reading performance. *Remedial and Special Education, 26*, 357-371.

Erickson, G. C., Foster, K. C., Foster, D. F., Torgesen, J. K., & Packer, S. (1992). Daisy Quest I and II. [Computer software]. Scotts Valley, CA: Great Wave Software.

Erickson, G. C., Foster, K. C., Foster, D. F., Torgesen, J. K., & Packer, S. (1993). Daisy's Castle [Computer software]. Scotts Valley, CA: Great Wave Software.

Fast ForWord Language (2002) [Computer software]. Oakland, CA: Scientific Learning Corporation.

Fletcher, J. M. & Vaughn, S. (2009). Response to intervention models as alternatives to traditional views of learning disabilities: Response to the commentaries. *Child Development Perspectives, 3*, 48-50.

Florida Center for Reading Research. (2003). *Headsprout Early Reading*. Retrieved from: http://www.fcrr.org/FCRRreports/PDF/headsprout_final.pdf

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice, 16,* 203-212.

Foshay, R. (2002). An Overview of the research base of PLATO (Technical Paper No. 12). Bloomington MN: PLATO Learning Inc. Retrieved from http://www.plato.com/media/Technical-White%20Papers/1/An%20Overview%20of%20the%20Research%20Base%20of%20PLATO%20Software.pdf

Foster, K., Erickson, G., Foster, D., Brinkman, D., & Torgesen, J.K. (1994). Computerized assisted instruction in phonological awareness: Evaluation of the DaisyQuest program. *Journal of Research and Development in Education, 27*, 126-137.

Frey, B. B., Lee, S. W., Tollefson, N., Pass, L., & Massengill, D. (2005). Balanced literacy in an urban school district. *Journal of Educational Research, 98*, 272-280.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (2008). Making secondary intervention work in a three-tier responsiveness to intervention model: Findings from the first grade longitudinal reading study of the National Research

Center on Learning Disabilities. *Reading and Writing: An Interdisciplinary Journal, 21*, 413-436. doi: 10.1007/s11145-007-9083-9

Fuchs, D., & Fuchs, L. S. (2005). Peer Assisted Learning Strategies: Promoting word recognition, fluency, and reading comprehension in young children. *The Journal of Special Education, 39*, 34-44.

Fuchs, D., & Fuchs, L. S. (2005). Peer Assisted Learning Strategies: Promoting word recognition, fluency, and reading comprehension in young children. *The Journal of Special Education*, *39*, 34-44.

Fuchs, D., Mock, D, Morgan, P. L., & Young, C. L. (2003). Responsiveness to intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice, 18*, 157-171.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*, 188-192.

Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children, 61*, 15-24.

Fuchs, L. S., & Fuchs, D. (2009). On the importance of a unified model of responsiveness to intervention. *Child Development Perspectives, 3*, 41-43.

Georgia Department of Education (2007). CRCT Content Descriptions- Reading. Retrieved from www.doe.k12.ga.us

Georgia Department of Education Office of Assessment and Accountability (2008). *Georgia Student Assessment Program Student Assessment Handbook.* (2008-2009). Retrieved from: http://www.doe.k12.ga.us/DMGetDocument.aspx/Student%20Assessment%20Ha ndbook%202008-2009.pdf?p=6CC6799F8C1371F6E6F831068DAF4D53C727B818AEEBF7BBAE804ABFD452AA7A&Type=D

Georgia Department of Education Office of Assessment and Accountability (2008). *What Georgia Educators Need to Know About Georgia's Testing Program*. Retrieved from: http://www.dekalb.k12.ga.us/instruction/testing/files/TestingNewsletter2008GADOE.pdf

Georgia Department of Education, Office of Curriculum, Instruction, and Assessment (2009). *An assessment and accountability brief: Validity and Reliability for the 2009 CRCT*.

Georgia Department of Education Testing Division (2006). *What Georgia Educators Need to Know About Georgia's Testing Program*. Retrieved from: http://public.doe.k12.ga.us/DMGetDocument.aspx/Final%20-%20testing%20newspaper.pdf?p=4BE1EECF99CD364EA5554055463F1FBBF5D074D5FB1F2CAEB3B63B3ECB220CDD26C2114F3C57D8D2040FD88C56B817B3&Type=D

Georgia Performance Standards, GPS; O.C.G.A. § 20-2-281. (2008).

Governor's Office of Student Achievement. (2009). Retrieved from: http://www.gaosa.org/FindASchool.aspx?TabRequested=District&FromSection=score&PageReq=106&CountyId=638&SY=2010

Glover, T. A., & DiPerna, J. C. (2007). Service Delivery for Response to Intervention: Core components and directions for future research. *School Psychology Review, 36*, 526-540.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills (6th ed.)*. Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from: http://dibels.uoregon.edu/

Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, L., Smith, S., & Flindt, N. (2004). *Technical Adequacy and Decision Making Utility of DIBELS* (Technical Report No. 7). Eugene, OR: University of Oregon.

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

Graham, J. W. (2009). Missing data: Making it work in the real world. *Annual Review of Psychology, 60*, 549-576. doi: 10.1146/annurev.psych.58.110405.085530

Gray, L., Thomas, N., and Lewis, L. (2010). *Teachers' Use of Educational Technology in U.S. Public Schools: 2009* (NCES 2010-040). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hagan-Burke, S., Burke, M. D., & Crowder, C. (2006). The convergent validity of the Dynamic Indicators of Basic Early Literacy Skills and the Test of Word Reading Efficiency for the beginning of first grade. *Assessment for Effective Intervention, 31*, 1-15. doi: 10.1177/073724770603100401

Hall, J. D., Howerton, L., & Jones, C. H. (2008). Achievement testing in the No Child Left Behind era: The Arkansas benchmark. *Research in the Schools, 15*, 64-76.

Hall, T., Hughes, C., & Filbert, M. (2000). Computer assisted instruction in reading for students with learning disabilities: A research synthesis. *Education and Treatment of Children, 23*, 173-193.

Harn, B. A., Stoolmiller, M., & Chard, D. J. (2008). Measuring the dimensions of alphabetic principle on the reading development of first graders: The role of automaticity and unitization. *Journal of Learning Disabilities, 41*, 143-157.

Hasselbring, T. S. (1986). Research on the effectiveness of computer-based instruction: A review. *International Review of Education, 32*, 313-324

Hasselbring, T. S. (2001). A possible future of special education technology. *Journal of Special Education Technology, 16*, 15-21.

Headsprout. (2005). [Internet-based computer software]. Seattle, WA: Headsprout Corporation.

Headsprout. (2007).Results count: Outcome data and case studies. Seattle, WA: Headsprout.

Hintz, J., Ryan, A. & Stoner, G. (2003). Concurrent validity and Diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review, 32*, 541-556.

Hipps, G., Abrami, P., Savage, R., Cerna, N., & Jorgensen, A. (n.d.). ABRACADABRA: The research, design and development of web-based early literacy software. Montreal, Quebec, Canada: Centre for the Study of Learning and Performance Concordia University. Retrieved from:

http://grover.concordia.ca/abracadabra/promo/en/download/Hipps_ABRACADABRA_The%20research.pdf

Hodgkinson, H. (1991). Reform versus Reality. *Phi Delta Kappan, 73,* 8-16.

Hook, P. E., Macaruso, P., & Jones, S. (2001). Efficacy of Fast ForWord training on facilitating acquisition of reading skills by children with reading difficulties—A longitudinal study. *Annals of Dyslexia*, *51*, 75-96.

Horn, C., Ramos, M., Blumer, I., Madaus, G., Lynch, P. S., & Lynch, C. A. (2000). *Cut Scores: Results May Vary*. Boston: The National Board on Educational Testing and Public Policy (Monograph). Retrieved from http://www.bc.edu/research/nbetpp/statements/M1N1.pdf.

Houghton Mifflin School publishers website Retrieved from: http://www.hmhschool.com/store/ProductCatalogController?cmd=Browse&subcmd=LoadDetail&level1Code=1&frontOrBack=F&division=S01&&division=S01&sortEntriesBy=SEQ_NAME&sortProductsBy=SEQ_TITLE&cmain=003399&cfaded=99CCFF

Howell, R. D., Erickson, K., Stanger, C., & Wheaton, J. E. (2000). Evaluation of a computer-based program on the reading performance of first grade students with potential for reading failure. *Journal of Special Education Technology*, *15*, 5-14.

Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. § 1400 *et seq*.

Juel, C. (1998). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.

Kalyuga, S., Ayers, P., Chandler, P., Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31.

Kamil, M. L. (2005). Reading Instruction for Low-Achieving and At-Risk Students (Publication Series No. 3). Philadelphia: Institute of Education Sciences . Retrieved from: http://www.temple.edu/lss/pdf/publications/pubs2005-3.pdf

Kamil, M. L. (2008). *Elements of Successful Reading Instruction*. Center on Innovation and Improvement. Lincoln, IL: Academic Development Institute. Retrieved from: http://www.centerii.org/techassist/solutionfinding/resources/SuccReadInstruct.pdf

Kaminski, R., Cummings, K. D., Powell-Smith, K. A., & Good, R. H., III. (2008). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) for formative assessment and evaluation. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology V: Vol. 4* (pp. 1181-1204). Bethesda, MD: National Association of School Psychologists.

Kaminski, R. A. & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25*, 215-227.

Kaminski, R. A., & Good, R. H. (1998). Assessing early literacy skills in a problem solving model: Dynamic Indicators of Basic Early Literacy Skills. In M. R. Shinn (Ed.), *Advanced applications of Curriculum-Based Measurement* (pp. 113-142). New York: Guilford.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95*, 3-21.

Kulik, J. (2003). *Effects of Using Instructional Technology in Elementary and Secondary Schools: What Controlled Evaluation Studies Say: Final Report*. Arlington, VA: SRI International.

Kulik, J., & Kulik, C. C. (1987). Computer based education: What 200 evaluations say. Paper presented at the Annual Convention of the Association for the Educational Communications and Technology, Atlanta, GA, Feb. 26-March 1 1987. Retrieved from: http://www.eric.ed.gov/PDFS/ED285521.pdf

Kulik, C. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior*, *7*, 75-94.

Language for Learning. (2005). Columbus, OH: McGraw-Hill.

Layng, T. V. J., Twyman, J. S., & Strikeleather, G. (2003). Headsprout Early Reading: Reliably teaching children to read. *Behavioral Technology Today, 3*, 7-20.

Layng, T. V. J., Twyman, J. S., & Strikeleather, G. (2004a). Engineering discovery learning: The contingency adduction of some precursors of textual responding in a beginning reading program. *The Analysis of Verbal Behavior, 20*, 99-109.

Layng, T. V. J., Twyman, J. S., & Strikeleather, G. (2004b) Selected for success: How Headsprout Reading Basics teaches beginning reading. In D. J. Moran, (Ed.), *Evidence-Based Educational Methods* (pp. 171-197).Saint Louis: Elsevier Science & Technology Books.

Loeb, D., Gillam, R., Hoffman, L., Brandel, J., & Marquis, J. (2009). The effects of Fast ForWord language on the phonemic awareness and reading skills of school age children with language impairments and poor reading skills. *American Journal of Speech Language Pathology, 18*, 376-387.

Lovell, M., & Phillips, L. (2009). Commercial software programs approved for teaching reading and writing in the primary grades: Another sobering reality. *Journal of Research on Technology in Education, 42*, 197-216.

Lyon, G. R. (1997). Hearing on literacy: Why kids can't read. In Testimonies to Congress: 1997-2002. Covington, LA: Center for the Development of Learning. MacArthur, C., Ferretti, R., Okolo, C., & Cavalier, A. (2001). Technology applications for students with literacy problems: A critical review. *The Elementary School Journal, 101*, 273-301.

Macaruso, P., Hook, P., & McCabe, R. (2006). The efficacy of computer based supplementary phonics programs for advancing reading skills in at risk elementary students. *Journal of Research in Reading, 29*, 162-172. doi: 10.1111/j.1467-9817.2006.00282.x

Macaruso, P., & Rodman, A. (2011). Efficacy of computer-assisted instruction for the development of early literacy skills in young children. *Reading Psychology, 32*, 172-196. doi: 10.1080/02702711003608071

Macaruso, P., & Walker, A. (2008). The efficacy of computer-assisted instruction for advancing literacy skills in kindergarten children. *Reading Psychology, 29*, 266-287. doi: 10.1080/02702710801982019

Maris, J. (2005). Validation of the Learning Machine. Proceedings of the 2005 Informing Science and IT Education Joint Conference June 16 -19 Flagstaff, AZ. Retrieved from: http://informingscience.org/proceedings/InSITE2005/P01f90Mari.pdf

Mason, E. J. (2007). Measurement issues in high stakes testing. *Journal of Applied School Psychology, 23*, 27-46. doi: doi:10.1300/J370v23n02_03

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly, 40*, 148-182. Retrieved from: http://www.jstor.org/stable/4151678

Mathes, P. G., Torgesen, J. K., & Allor, J. H. (2001). The effects of Peer-Assisted Literacy Strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. American Educational Research Journal, 38, 371-410. Retrieved from: http://www.jstor.org/stable/3202463

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43-52.

McBride, J. R., Ysseldyke, J., Milone, M., & Stickney, E. (2010). Technical adequacy and cost benefit of four measures of early literacy. *Canadian Journal of School Psychology*, *25*, 189-204. doi: 10.177/0829573510363796

McCook, J. E. (2006). *The RTI Guide: Developing and Implementing a Model in Your Schools*. Horsham, Pennsylvania: LRP Publications.

McIntyre, E., Jones, D., Powers, S., Newsome, F., Petrosko, J., Powell, R., & Bright, K. (2005). Supplemental instruction in early reading: Does it matter for struggling readers? *Journal of Educational Research, 99,* 99-107.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton. D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71*, 445-463.

Mitchell, M. J., & Fox, B. J. (2001). The effects of computer software for developing phonological awareness in low-progress readers. *Reading Research and Instruction, 40*, 315-332.

Moats, C. (2004). The professional development of reading teachers. In The Voice of Evidence in Reading Research, P. McCardle & V. Chhabra (Eds.). Baltimore: Paul H. Brooks Publishing.

Murphy, R., Penuel, W., Means, B., Korbak, C., Whaley, A., & Allen, J. (2002). E-DESK: A review of recent evidence on the effectiveness of discrete educational software. Arlington, VA: SRI International. Retrieved from: http://ctl.sri.com/publications/downloads/Task3_FinalReport3.pdf

National Association of State Directors of Special Education (2008). Response to Intervention: Blueprints for implementation. Retrieved from: http://nasdse.org/Portals/0/DISTRICT.pdf

National Center for Educational Statistics. (2003). The Nation's Report Card: Reading 2003 Major Results. Retrieved from: http://nces.ed.gov/nationsreportcard/reading/results2003.

National Center for Education Statistics (2005). *National Assessment of Adult Literacy: A First look at the literacy of America's Adults in the 21st century.* Washington, D.C.: U.S. Department of Education Institute of Education Sciences NCES 2006-470.

National Center for Educational Statistics. (2009). The Nation's Report Card: Reading 2009. National Assessment of Educational Progress at Grades 4 and 8. Washington, D.C.: National Center for Educational Statistics. Retrieved from: http://nces.ed.gov/nationsreportcard/pdf/main2009/2010458.pdf

National Reading Panel (2000). Teaching Children to Read: An Evidence-based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction. Washington, DC: US Department of Health and Human Services. Retrieved from http://www.nationalreadingpael.org.

Nelson, (2008). Beyond correlational analysis of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A classification validity study. *School Psychology Quarterly, 23*, 542-552. doi: 10.1037/a0013245

No Child Left Behind Act of 2001, 20 U.S.C. § 6319 (2008). Retrieved from: http://www2.ed.gov/nclb/landing.jhtml

Nunnally, J. C. (1975). Psychometric theory 25 years ago and now. *Educational Researcher, 4*, 7-14. Retrieved from: http://www.jstor.org/stable/1175619

Oregon Reading First Center (2006). Review of Comprehensive Reading Programs. Retrieved from http://oregonreadingfirst.uoregon.edu/downloads/instruction/curriculum_review/core_review/trophies2003_core_review.pdf

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1-4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1-8.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*, 184-202. doi:10.1598/RRQ.40.2.3

Paris, S. G., Carpenter, R. D., Paris, A. H., & Hamilton, E. E. (2005). Spurious and genuine correlates of children's reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), Children's Reading Comprehension and Assessment (pp.131-160). Mahwah, NJ: Lawrence Erlbaum Associates.

Patterson, W. A., Henry, J. J., O'Quinn, K., Ceprano, M. A., & Blue, E. V. (2003). Investigating the effectiveness of an integrated learning system on early emergent readers. Reading Research Quarterly, 38, 172-207. Retrieved from: http://www.jstor.org/stable/4151782

PLATO Beginning Reading (2001) [Computer software]. Bloomington, MN: PLATO Learning.

Pollock, E., Chandler, P., Sweller, J. (2002). Assimilating complex information. Learning & Instruction, 12, 61-86.

Powers, S. & Price-Johnson, C. (2006). *Evaluation of the Waterford Early Reading Program in Kindergarten 2005-06*. Tucson, AZ: Creative Research Associates. Retrieved from: http://www.eric.ed.gov/PDFS/ED501576.pdf

Pressley, M. (2001). *Effective Beginning Reading Instruction*. Executive Summary and Paper Commissioned by the National Reading Conference. Chicago, IL: National Reading Conference. Retrieved from: http://www.nrconline.org/publications/pressleywhite2.pdf

Pressley, M. (2005). Reading Instruction That Works: The Case for Balanced Teaching, Third Edition. New York: The Guilford Press.

Quinn, D. W. & Quinn, N. W. (2002). *Grades 1-8, Apache Junction Unified School District 43, Apache Junction, Arizona*. PLATO Evaluation Series. Bloomington, MN: PLATO Learning Inc. Retrieved from: http://www.eric.ed.gov/PDFS/ED470285.pdf

Reading First Impact Study Final Report (2008). Institute of Education Sciences. Retrieved from: http://www.eric.ed.gov/PDFS/ED503344.pdf

Reschly, D. J. (2005). Learning disabilities identification: Primary intervention, secondary intervention, and then what? *Journal of Learning Disabilities, 38*, 510-515.

Response to Intervention Adoption Survey. (2010). Townson, MD: Spectrum K12 School Solutions. Retrieved from: http://www.spectrumk12.com/about_us/about_us_overview

Reading Mastery Classic. (2003). Columbus, OH: McGraw Hill.

Riedel, B. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly, 42*, 546-567.

Ritchey, K. D. (2008). Assessing letter sound knowledge: A comparison of letter sound fluency and nonsense word fluency. *Exceptional Children, 74*, 487-506.

Ritchey, K. D. & Goeke, J. L. (2006). Orton-Gillingham and Orton-Gillingham-based reading instruction: A review of the literature. The Journal of Special Education, 40, 171-183. Retrieved from: http://www.jstor.org/stable/20068319

Rouse , H., & Fantuzzo, J. (2006). Validity of the Dynamic Indicators of Basic Early Literacy Skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review, 35*, 341-355.

Rouse, C. E. & Krueger, A. B. (2004). Putting computerized instruction to the test: A randomized evaluation of a 'scientifically based" reading program. *Economics of Education Review, 23*, 323-338. doi: 10.1016/j.econedurev.2003.10.005

Rosenfield, S. A. & Gravois, T. A. (1996). *Instructional Consultation Teams: Collaborating for Change*. New York: The Guilford Press.

Rumelhart, D. E. (1994). Toward an interactive model of reading. In R. B. Ruddell, M. R. Ruddell & H. Singer (Eds.). *Theoretical Models and Processes of Reading, 4th Edition* (pp. 864-894). Newark: DE International Reading Association.

Rupley, W. H., Blair, T. R., & Nichols, W. D. (2009). Effective reading instruction for struggling readers: The role of direct/ explicit teaching. *Reading & Writing Quarterly, 25*, 125-138. doi: 10.1080/10573560802683523

Savage, R. S., Abrami, P., Hipps, G., & Deault, L. (2009). A randomized controlled trial study of the ABRACADABRA reading intervention program in grade 1. *Journal of Educational Psychology, 101*, 590-604. doi: 10.1037/a0014700

Savage, R. S., Erten, O., Abrami, P., Hipps, G., Comaskey, E., & van Lierop, D. (2010). ABRACADABRA in the hands of teachers: The effectiveness of a web-based literacy intervention in grade 1 language arts programs. *Computers & Education, 55,* 911-922. doi:10.1016/j.compedu.2010.04.002

Scarborough, H. S. & Brady, S. A. (2002). Toward a common terminology for talking about speech and reading: A glossary of the "phon" words and some related terms. *Journal of Literacy Research, 34*, 299-336.

Schilling, S. G., Carlisle, H. F., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement?. *The Elementary School Journal, 107*, 429-448. Retrieved from: http://www.jstor.org/stable/10.1086/518622

Sedita, J. (2005). Effective vocabulary instruction. *Insights on Learning Disabilities, 2*, 33-45.

Shaywitz, S. E. & Shaywitz, B. A. (2004). Neurobiologic basis for reading and reading disability. In P. McCardle & V. Chhabra (Eds.). *The Voice of Evidence in Reading Disability* (pp. 417-442). Baltimore: Paul H. Brooks Publishing.

Shinn, M. R., & Shinn, M. M. (2002). *AIMSWeb Training Workbook: Administration and Scoring of Reading Curriculum Based Measurement (R-CBM) for Use in General Outcomes Measurement*. Retrieved from http://www.aimsweb.com.

Sivin-Kachala, J. & Bialo, E. (1998). Report on the effectiveness of technology in the schools 1990-1994. ERIC Document Reproduction Service No. ED371726. Retrieved from: http://www.eric.ed.gov/PDFS/ED371726.pdf

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing Reading Difficulties in Young Children*. Washington, D.C.: The National Academy Press.

Soe, K., Koki, S., & Chang, J. (2000). Effect of computer-assisted instruction (CAI) on reading achievement: A meta-analysis. Pacific Resources for Education and Learning (PREL) ERIC Document Reproduction Service No. ED443079. Retrieved from: http://www.prel.org/products/Products/effect-cai.htm

Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *The Journal of Special Education, 36*, 223-233.

Stanovich, K. E. (1980). Toward an Interactive-Compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*, 32-71. Retrieved from: http://www.jstor.org/stable/747348

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*, 795-819. doi: 10.1002/pits.20113

Strong, G. K., Torgerson, C. J., Torgerson, D., & Hulme, C. (2011). A systematic meta-analytic review of evidence for the effectiveness of the Fast ForWord language intervention program. *The Journal of Child Psychology and Psychiatry*, *53*, 224-235. doi: 10.1111/j.1469-7610.2010.02329.x

Sweet, R. W. (2004). The big picture: Where we are nationally on the reading front and how we got here. In The Voice of Evidence in Reading Research, P. McCardle & V. Chhabra (Eds.). Baltimore: Paul H. Brooks Publishing.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics, Fifth Edition.* Boston: Pearson Education, Inc.

Telzrow, C. F., McNamara, K., & Hollinger, C. L. (2000). Fidelity of problem-solving implementation and relationship to student performance. *School Psychology Review, 29*, 443-461.

Tiratira, N. L. (2008). Cutoff scores: The basic Angoff method and the item response theory method. *The International Journal of Educational and Psychological Assessment, 1*, 27-35.

Torgesen, J. K. (2009). The response to intervention instructional model: Some outcomes from a large-scale implementation in reading first schools. *Child Development Perspectives, 3,* 38-40.

Tracey, D. & Young, J. (2007). Technology and early literacy: The impact of an integrated learning system on high-risk kindergartners' achievement. *Reading Psychology, 28*, 443-467. Retrieved from: http://www.eric.ed.gov/PDFS/ED491554.pdf

Troia, G. A. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. Reading Research Quarterly, 34, 28-52. Retrieved from: http://www.jstor.org/stable/748268

Troia, G. A. & Whitney, S. D. (2003). A close look at the efficacy of Fast ForWord Language for children with academic weaknesses. *Contemporary Educational Psychology, 28*, 465-484. doi: 10.1016/S0361-476X(02)00045-0

Twyman, J. S., Layng, T. V. J., Strikeleather, G., & Hobbins, K. A. (2004). A non-linear approach to curriculum design: The role of behavior analysis in building an effective reading program. In *Focus on Behavior Analysis in Education*, W. L.

Heward, T. E. Heron, N. A. Neef, S. M. Peterson, D. M. Sainato, G. Y. Cartledge, R. Gardner, L. D. Peterson, S. B. Hersh, & J. C. Dardig (Eds.). Baltimore: Paul H. Brooks Publishing.

VanDerHeyden, A. M., Witt, J. C., & Gilberson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45*, 225-256. doi:10.1016/j.jsp.2006.11.004

Vaughn, S. & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice, 18*, 137-146.

Vaughn, S., Wanzek, J., Woodruff, A. L.. & Linan-Thompson, S. (2007). Prevention and early identification of students with reading disabilities. In D. Hager (Ed.), *Evidence-based reading practices for response to intervention*. Baltimore: Paul H. Brookes Publishing.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45*, 2-40.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R. S., Pratt, A., & Denckla, M. B. (1996). Cognitive profiles of difficult to remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*, 601–638.

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities, 39*, 175-169.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36*, 541-561.

Waterford Early Reading Program (2003) [Computer software]. Pearson Digital Learning.

What Works Clearinghouse. (2009). Headsprout Early Reading. Institute for Education Sciences. Retrieved from: http://ies.ed.gov/ncee/wwc/pdf/wwc_headsprout_102709.pdf

What Works Clearinghouse. (2007). Beginning Reading. Institute for Education Sciences. Retrieved from: http://ies.ed.gov/ncee/wwc/pdf/BR_TR_08_13_07.pdf

Wild, M. (2009). Using computer-aided instruction to support the systematic practice of phonological skills in beginning readers. *Journal of Research in Reading, 32*, 413-432. doi: 10.111/j.1467-9817.2009.01405x

Wilson, H. A. & Atkinson, R. C. (1967). Computer based instruction in initial reading: A progress report on the Stanford Project (Technical Report No. 119). Stanford, CA: Institute for Mathematical Studies in the Social Sciences. Retrieved from: http://suppes-corpus.stanford.edu/techreports/IMSSS_119.pdf

Witt, J. (2002). Screening to enhance equitable placement (STEEP). Retrieved from http://www.joewitt.org.

Wood, C., Pillinger, C., & Jackson, E. (2010). Understanding the nature and impact of young readers' literacy interactions with talking books and during adult reading support. *Computers & Education, 54*, 190-198. doi:10.1016/j.compedu.2009.08.003

Zhao, Y., Englert, C. S., Chen, J., Jones, S. C., & Ferdig, R. E. (2000). The development of a web-based literacy learning environment: A dialogue between innovation and established practices. *Journal of Research on Computing in Education, 32*, 435-454.