The Effects of a Computer Assisted Reading Program on the Oral Reading Fluency and Comprehension of At-Risk, Urban First Grade Students

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Lenwood Gibson, Jr., M.S.

College of Education and Human Ecology

The Ohio State University

2009

Dissertation Committee:

Dr. Gwendolyn Cartledge, Advisor

Dr. Moira Konrad

Dr. Rajiv Ramnath
Copyright by

Lenwood Gibson, Jr.

2009
The current investigation studied the effects of a computer software program of the oral reading fluency and comprehension of eight, first grade students. These students were identified as being either “at-risk” or “some-risk” for reading failure on the DIBELS oral reading fluency winter benchmark assessment. A multiple probe design was used to determine if the Read Naturally software edition would increase the ORF, as measure as correct words per minute, on treatment and generalization probes. Also measured were the percentage of comprehension questions answered correctly on treatment probes and word retell on both treatment and generalization probes. The Read Naturally software edition consisted of a repeated reading sequence activity that the participants engaged in on a laptop computer. The sequence consisted of seven components: key words, one-minute cold reading, read along, practice reading, comprehension quiz, and pass timing. Treatment sessions lasted approximately thirty minutes and were conducted three to four times per week, for between 14 and 16 weeks. Pretest and posttest measure were collected to determine if the participants’ scores on these test would increase following the completion of the intervention.

The results of the intervention demonstrated the effectiveness of the Read Naturally program on the OFR and comprehension of the participants on treatment
probes. All of the participants were able to increase their scores on these measures over baseline levels; however, the results were not as evident for generalization probes. Many of the participants were not able to increase their ORF on generalization probes to benchmark levels. Due to this failure to increase reading fluency a second treatment phase was implemented with increase fluency requirements. During this second phase 5 out of 8 participants were able to increase their ORF to benchmark levels on generalization passages. Additionally, 5 out of 8 participants were able to decrease their DIBELS ORF risk status on winter benchmark assessments. These finding support research on the use of repeated reading strategies, but extend the research in several ways. First, the results suggest the computer assisted reading programs are a viable means to supplement classroom instruction. Second, they suggest that generalization might be promoted by increasing fluency goal criteria on practice passage reading. Finally, the results suggest that repeated reading strategies can be used successful with first grade students. Implications for classroom use, limitations, and directions for future research are presented.
Dedicated to my future children….

Although you are not here yet, I have worked so hard to be able to give you the life you deserve.
ACKNOWLEDGEMENTS

First and foremost I would like to thank the faculty at The Ohio State University PAES Special Education program. Through your guidance and support I was able to undertake this monumental task. I want to extend a special thank you to my adviser Dr. Gwendolyn Cartledge. She was the first female mentor that I have had but she is by far the best and most influential person in my academic career. Under her expert mentorship I have learned an enormous amount and dedicated myself to the improvement of disadvantaged children everywhere. I look forward to working with and learning from her for years to come.

Next, I would like to thank Dr. Moira Konrad for her guidance and suggestions through the processes of my candidacy and dissertation. I feel like I have become a stronger writer because of all of her feedback. I would also like to thank Dr. Rajiv Rammath for serving on my dissertation committee.

This study was made possible because of the cooperation of two schools in Columbus. I want to thank all of the teachers and administrator for allowing me to be a part of their students’ education. Starr Keyes was instrumental in providing interobserver agreement and for giving me her valuable input during our research meetings. I want to
thank my family for supporting me through this process, especially my mom and my sister.

Finally, I wanted to save the last paragraph for the most important person in my life, Jessica Rodriguez. Thank you for being there for me every step of the way. If not for your love and support, not to mention all the phone calls, I do not think I would have been able to make it out here. I know it has been a long and difficult road but it is almost over and I promise you will not regret standing by my side. I love you Sweet pea!
VITA

April 17, 1975.............................................. Born- Bristol, PA

May 23, 1999.............................................. B.A. in Psychology,
Eastern Connecticut State University
Willimantic, CT

September 10, 2002................................. M.S. Applied Behavior Analysis
Northeastern University
Boston, MA

2003-2006.............................................. Behavior Analyst
River Street School
Windsor, CT

2006-current ............................................ Graduate Research Associate
The Ohio State University

PUBLICATIONS


FIELDS OF STUDY

Major Field: Education

Specialization: Special Education and Applied Behavior Analysis
TABLE OF CONTENTS

ABSTRACT .................................................................................................................................................. iii
ACKNOWLEDGEMENTS ..........................................................................................................................vi
VITA ...........................................................................................................................................................viii
Publications .................................................................................................................................................viii
Fields of Study ........................................................................................................................................... ix
Major Field ................................................................................................................................................ ix
CHAPTER 1 Introduction .............................................................................................................................1
Reading Failure and Effect on Children ......................................................................................................1
    Presidential Response ............................................................................................................................ 2
    Trends in Reading Abilities: The Nation’s Report Card ........................................................................ 4
Reading Instruction .................................................................................................................................. 7
    Traditional reading instruction .............................................................................................................. 8
Can Computers Help? ................................................................................................................................. 9
Purpose of the study ................................................................................................................................... 11
Research Questions .................................................................................................................................. 12
CHAPTER 2 Literature Review ................................................................................................................... 13
The National Reading Panel Report .......................................................................................................... 13
    Components of Reading ......................................................................................................................... 14
Research Questions Three .................................................................150
   Word retell of treatment probes ...........................................150
Research Question Four .................................................................153
   Correct words per minute on generalization probes ..............153
   Word retell of generalization probes ...................................154
Research Questions Five ...............................................................160
   Letter/word identification .................................................160
   Word attack .................................................................161
   Passage comprehension ................................................162
   Summary of Results .........................................................162
Research Question Six ...............................................................162
Research Question Seven ...........................................................163
Research Question Eight ............................................................165
Limitations .............................................................166
   Failure to implement in a natural classroom environment .......166
   Lack of comprehension questions for generalization passages ....167
   Lack of standardized reading fluency measures ......................168
   Failure to assess generalization on actual classroom reading group...168
Implication for Classroom Use .................................................169
Additional Considerations..............................................................172
Directions for Future Research ...................................................173
REFERENCES ...........................................................................177

xvi
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Participant demographic information and DIBELS selection criteria scores</td>
<td>68</td>
</tr>
<tr>
<td>4.1 Baseline and intervention averages for school one participants; effect sizes and standard deviations included</td>
<td>105</td>
</tr>
<tr>
<td>4.2 Baseline and intervention averages for school two participants; effect sizes and standard deviations included</td>
<td>119</td>
</tr>
<tr>
<td>4.3 Woodcock Johnson III pretest and posttest scores for LWID, WA, and PC sub-tests for participants in school one</td>
<td>126</td>
</tr>
<tr>
<td>4.4 Winter and spring DIBELS ORF benchmark scores and risk statuses for school one students</td>
<td>127</td>
</tr>
<tr>
<td>4.5 Woodcock Johnson III pretest and posttest scores for LWID, WA, and PC sub-tests for participants in school two</td>
<td>131</td>
</tr>
<tr>
<td>4.6 Winter and spring DIBELS ORF benchmark scores and risk statuses for school two students</td>
<td>132</td>
</tr>
<tr>
<td>4.7 Social validity questionnaire results for the teachers of the participants</td>
<td>134</td>
</tr>
<tr>
<td>4.8 Social validity questionnaire results for the participants</td>
<td>136</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Treatment and generalization probe session data (CWPM) for school one participants</td>
<td>104</td>
</tr>
<tr>
<td>4.2 Treatment and generalization probe session data (retell) for school one participants</td>
<td>106</td>
</tr>
<tr>
<td>4.3 Treatment and generalization probe session data (percentage correct) for school one participants</td>
<td>107</td>
</tr>
<tr>
<td>4.4 Treatment and generalization probe session data (CWPM) for school two</td>
<td>118</td>
</tr>
<tr>
<td>4.5 Treatment and generalization probe session data (retell) for school two participants</td>
<td>120</td>
</tr>
<tr>
<td>4.6 Treatment and generalization probe session data (percentage correct) for school two participants</td>
<td>121</td>
</tr>
</tbody>
</table>
Learning how to read is arguably the most important academic skill for any individual to acquire. This is because reading is the basis on which many other academic activities are built (Flanagan, West, & Walston, 2004; Lyon, 1998; 2003), and students who fail to acquire adequate reading skills are much more likely to experience school failure. There is research to support the fact that students who perform poorly in reading during early primary grades are more likely to continue to have difficulties and may dropout of school altogether (Alexander, Entwisle, & Horsey, 1997; Finn & Rock, 1997; McIntosh, Flannery, Sugai, Braun, & Cockrane, 2008, Simner & Barnes, 1991). It is only logical then to focus on early reading instruction as a preventive strategy.

*Reading failure and effects on children.* The problem of illiteracy in the United States is so pronounced that the National Institute of Child Health and Human Development (NICHD) considered reading failure a public health problem (2001). This is because individuals who cannot read are much more likely to live in impoverished conditions (Askov, 1991; Dwyer, 1997) and have children who are at risk for reading failure as well. The phenomenon of intergenerational illiteracy involves children being raised by parents who cannot read and unable to foster their children’s reading abilities (Aikens & Barbarin, 2008; Askov, 1991; Cooter, 2006). This cycle of reading difficulties
and/or illiteracy only increases the need to identify students at high risk for reading failure early and provide them with sufficient instruction to prevent further problems as they grow older.

It is well documented that students must learn how to read prior to third grade if they are going to be successful throughout their school careers. In fact, students who are reading below grade level by the time they reach third grade will most likely never catch up to their grade level peers (Frances, Shaywitz, Stuebing, Fletcher, & Shaywitz, 2005; Juel, 1988). Once behind, these students are much more likely to stay behind as the school curriculum becomes increasingly more difficult (Carlson & Francis, 2002). It is quite probable that many of these students will either be placed in special education classes or develop behavior problems because of their difficulties with reading (Bennett, Brown, Boyle, Racine, & Offord, 2003; Goodman & Webb, 2006; Levy & Chard, 2001; Rivera, Al-Otaiba, Koorland, 2006; Scott & Shearer-Lingo, 2002; Torgesen, 2001).

*Presidential response.* The problems associated with reading failure in the United States became so alarming that the past two Presidential administrations have passed legislation to address reading instruction in the nation’s school systems. President Clinton passed the Reading Excellence Act in 1998 (PL 105-277), which provided $260 million to states to establish strong reading programs within their school districts. The funds were allocated for professional development and for instructional materials that were based on scientifically proven reading instruction. The concept of scientifically based reading research became very important to the development and implementation of reading programs across the nation.
To further the cause for reading reform in the United States, President G. W. Bush initiated the No Child Left Behind Act of 2001 (PL 107-110). This act included a piece of legislation, Reading First, which provided states with federal funding for professional development of kindergarten through third grade teachers. This professional development was used to inform teachers how to include scientifically based reading programs within their instruction. An important provision in the Reading First Legislation was the requirement of each state and local school district to report on the reading levels of their students. These reports were subject to an independent review process and states that failed to implement effective programs (i.e., students’ reading level did not increase) could lose their federal funding.

Both of these administrations not only made reading improvement a top priority for school districts, they passed legislation that mandated the use of scientifically validated reading programs. The criteria for these types of programs were strict and included the following guidelines:

A) means the application of rigorous, systematic, and objective procedures to obtain valid knowledge relevant to reading development, reading instruction, and reading difficulties; and

B) shall include research that-

i. employs systematic, empirical methods that draw on observation or experiment;
ii. involves rigorous data analyses that are adequate to test stated hypotheses and justify the general conclusions drawn;

iii. relies on measurement or observational methods that provide data across evaluators and observers and across multiple measures and observations; and

iv. has been accepted by peer reviewed journals or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review.

(“Reading and Literacy Grants to State Educational Agencies,” Title II (C) Sec. 2252 (5) [20 U.S.C 661a])

*Trends in Reading Abilities: The Nation’s Report Card*

The legislation of scientifically validated reading research required state and local educational agencies to implement such practices. Even before the enactment of the Reading Excellence Act and Reading First, the National Assessment for Educational Progress (NAEP) was reporting on the academic achievement of American students. These reports include data on reading, writing, and math. The primary concern of this section is to report on the data for reading.

Since 1992, the NAEP has been collecting and reporting on the overall reading abilities of fourth- and eighth-grade students across the country. These reports included data on reading scores and reading levels of students. For example, fourth-grade students needed to score a minimum of 238 points (out of a possible 500) on a standardized
reading assessment to be considered at the proficient level of reading. Students scoring lower than 238 but higher than 208 were considered to be reading at a basic level and any students that scored below 208 were reading below basic levels. If a student scored above 268 he/she was considered to be reading at an advanced level.

These reading levels and scores are adjusted for different grade levels, so the overall scores and achievement levels are different for eighth-grade students. In order for eighth-grade students to be reading on an advanced level they needed to score 323 or higher. To be proficient these students needed to score between 281 and 322. Basic readers needed to score between 243 and 280. Any eighth grade student scoring below 243 was considered to be reading below basic levels.

The data provided in the most recent report indicated that there were some statistically significant changes in the overall reading scores from 1992 to scores in 2007; however, these changes did not hold up when the data were disaggregated along certain student characteristics. The overall scores demonstrate that fourth- and eighth-grade students reading at or above proficient levels remained unchanged from 1992 through 2007 (e.g., eighth-grade students in the 90th percentile scored 305 in 1992 and 305 in 2007). For eighth-grade students reading at or below basic levels (indicated by a score of 280 or lower), there has been some improvement in their reading scores since 1998. There was an initial increase in all of the average scores at the basic and below level from 1998 to 2002, but this increasing trend leveled off over the remaining four years and scores returned to their 1998 level by 2007. That is, eighth grade students reading below basic levels (the 10th percentile) scored an average of 216 points in 1998. By 2002 this
average rose to 220 but then fell back to 217 by 2007. The exact reason for this trend is unclear but it is possible that there was an initial boost from the Reading Excellence Act, which then leveled off.

Although these scores seem to indicate that there were some improvements in the overall reading abilities of the nation’s students, a closer examination of these data reveals some troubling results. First, there is the achievement gap between White and minority students; specifically between White and Black and White and Latino students. In 1992, the gap between White and Black fourth-grade students was 32 points, where 71% of White students were reading at or above basic reading levels but only 32% of Black students were reading at or above basic. By 2007 this gap decreased to 27 points but 54% of Black fourth graders were still reading below what is considered basic compared to just 22% of White students.

These results are similar for eighth-grade students. In 1992, there was a 30 point gap between White and Black eighth grades and this gap only decreased by 3 points by 2007. The overall percentage of Black students in eighth grade that were reading below basic levels was 55 and although it decreased to 45% by 2007, the percentage was still well above the 16% of White eighth graders that were reading below basic level in 2007. The numbers for Latino students were very similar to Blacks; however, the percentage of Latinos reading below basic levels were actually slightly higher in every year of this report.

One possible explanation for these achievement discrepancies may be the prevalence of impoverished neighborhoods and schools for minority students. In 2007,
45% of fourth graders were eligible for free/reduced lunch, a poverty indicator, and these students scored an average of 205 points, which is below the cut off of 208 for basic reading skills. The remaining 54% of students that did not qualify for free/reduced lunch scored much better, obtaining an average score of 232 which was at the upper end of basic reading abilities. For states or districts in which the percentage of students who qualified for free/reduced lunch was higher, the scores on the reading assessment were substantially lower. For example, 66% of fourth grade students in Washington D.C. qualified for free/reduced lunch and these students scored an average of 188 points on the reading assessment. This is 20 points below what is considered basic reading skills.

As indicated by the NAEP’s biannual reporting, a large percentage of the fourth- and eighth-grade students in the United States are not reading proficiently. These overall results are a major concern, which drew the attention of the government in order to remedy the situation. Although there have been some noted reading improvements since the enactment of Reading Excellence and Reading First, many students still find themselves behind in the core skills they need to be efficient readers.

*Reading Instruction*

*Traditional reading instruction.* When it comes to reading, the problem of individualized instruction or active student engagement can become more pronounced. For many years, reading has been taught using a whole language approach in which students were not explicitly taught the different skills needed to be effective readers. Techniques such as sustained silent reading (SSR) and round-robin reading were used as instructional strategies.
Sustained silent reading requires students to select a book and read it quietly to themselves, often while teachers engage in other activities (e.g., preparing the next lesson, grading papers, etc.). The problem with this approach is that there is no way of assessing if students are actually reading the material. Also, even if students are “reading” the material, there is a possibility that struggling readers could be making mistakes that are going uncorrected. The role of direct feedback to correct errors is very important to any type of instruction, but it is especially important for early reading instruction (Heubusch & Lloyd, 1998; Wankaz, Vaughn, & Wexler, 2006; Welsch, 2006). This is because reading forms the basis of many other learning opportunities. Students who do not learn to read by the end of third grade are more likely candidates for reading failure and dropping out of school (Denti & Guerin, 1999).

Another technique that is used to teach reading is round-robin. This is where a group of students are required to take turns reading aloud from the same material. This allows the teacher to engage a group of students at the same time and provide feedback for mistakes. Although this technique may provide more opportunities for students to respond and receive feedback, there are also disadvantages for struggling readers. For instance, students who are not confident in their reading ability may try to predict when they are going to be called on and only “practice” reading that section of text. This could severely limit their ability to obtain the information presented in the rest of the text.

An even more serious problem is the development of escape and avoidance behaviors associated with reading failure. Students that are having difficulties with reading may become anxious in round-robin style reading instruction. Reading under
these conditions will quickly become an aversive task to them and they are likely to engage in disruptive behavior to escape the classroom environment. There is evidence in the research to suggest that reading problems are correlated with behavior problems in students (Gest & Gest, 2005; Jones, Lignugaris/Kraft, & Peterson, 2007; McIntosh, Chard, & Horner, 2006).

*Can Computers Help?*

It is clear that reading instruction in public school systems can pose a challenge for even the highest quality teacher. It is not surprising that only between 31% and 33% of the nation’s students were reading at or above proficient levels in fourth and eighth grade (National Assessment for Educational Progress, 2007). With conditions of overcrowding, weak instructional control, and non-evidence based teaching practices, researchers continue to look for ways to improve the educational process. One possible solution is the introduction and improvement of technology and more specifically the use of computers as a supplemental educational tool.

A serious problem faced by educational systems, especially in densely populated areas, is the high student to teacher ratio. In many classrooms across the country, teachers have 25 or 30 students and no form of assistance. Under these conditions, teachers are left to their own devises to structure their daily routines. Often times, this may involve large group or whole class lessons, worksheet activities, and large amounts of self-instruction (i.e., students are instructed to read material and answer question on their own). While many students are able to absorb instructional material through these methods and acquire an adequate education, there are plenty who do not. These students
soon find themselves behind their peers and they often struggle to catch up. The consequence of such a system can result in special education placement and the risk of school failure.

A key feature of computer technology is the ability to provide individualized instruction to students of all ability levels. There is evidence to support that learning is enhanced by opportunities to interact with and respond to instructional material (Huby, 2001; Malanga & Sweeney, 2008; Pratton & Hales, 1986). The more a student actively engages with the subject matter, the better chances that he/she will acquire the skills to master it. Through active engagement students are exposed to response opportunities where feedback can be provided for correct and incorrect responding. Computer technology is one way to engage students in responding because many programs require the user to actively interface with the equipment when operating a program. This is accomplished through actions such as mouse clicks, typing on the key board, using touch-screen, and voice recognition technology. So the use of a computer will almost automatically increase opportunities for students to interact and respond to instructional material.

**Computer Assisted Instruction for Reading.** Computer technology has flourished over the past 30 years. As this technology continues to grow, it is becoming increasingly integrated into the schools and classrooms across the country. Computer assisted instruction can be a viable and valuable teaching tool. Computers can be successfully used to teach many academic subjects, such as math and reading (Fuchs, Fuchs, & Hamlet, 2006; Leonard, Davis, & Silder, 2005). With the rapid growth of computer
technology, it is reasonable to assume that it will become an ever increasing component in the education of school children across the country. But how does computer technology fit into the everyday classroom situations?

Although there are a number of computer software programs available that were designed to improve the reading skills of students, many of these programs have not been investigated empirically. Some of these programs may only be minimally effective or not effective at all. There is also a paucity of research on the use of computer programs to increase the oral reading fluency (ORF) of young learners. So, although there is the potential for computer programs to be used as supplemental intervention for struggling readers, more empirical research is needed to examine specific software and determine its overall utility.

**Purpose of the study**

This study examined the use of a computer software reading program named Read Naturally (RN) to determine its effectiveness across several reading sub-skills. Specifically, the study examined if the RN software can be used to increase the ORF and comprehension of urban first-grade students identified as at-risk for reading failure. An additional focus was to determine if the ORF and comprehension skills generalized to untaught, standardized reading assessments.

**Research Questions**

1) What effect will the Read Naturally computer software program have on the oral reading fluency of urban, first-grade students at risk for reading failure?
2) What effect will the Read Naturally computer software program have on the number of comprehension questions answered correctly by the participants?

3) What effect will the Read Naturally computer software program have on the oral re-tell of the participants?

4) What effect will the Read Naturally computer software program have on the ORF of generalization passages (i.e. Aimsweb ORF progress monitoring measures)?

5) What effect will the intervention have on the reading scores of the three Woodcock-Johnson reading sub-tests as administered pre and post the intervention?

6) What effect will the intervention have on the pretest and posttest scores on the DIBLES ORF benchmark assessment?

7) How will teachers rate the effects of the intervention on the participants ORF and comprehension?

8) Will the participants respond favorably to the Read Naturally program, as measured by a questionnaire administered at end of the study?
Chapter 2: Literature Review

This chapter contains a review of the literature on instructional methods to increase oral reading fluency (ORF) and computer assisted instruction (CAI) in reading. The review on ORF instructional strategies includes research on repeated readings, error correction procedures, and multi-component interventions. Research on the early use of computer based reading instruction and programs specifically designed to increase ORF (i.e. Read Naturally) make up the CAI literature reviewed.

The National Reading Panel (NRP) report

When it comes to reading instruction and what is needed for an individual to be a good reader, there has been long standing divisions among the academic community. On one side of the debate, teaching professionals argue for the use of the whole language approaches, which are based on the position that reading develops naturally and therefore does not emphasize the need for explicit instruction in phonemic awareness skills (Drecktrah & Chiang, 1997; Goodman, 1989). The other side contends that reading is not a natural phenomenon, but man made and thus there is a need for explicit instruction in early reading skills such as phonemic awareness and systematic phonics. Explicit instruction involves an instructional paradigm that is teacher centered and emphasizes sequentially teaching reading skills that build upon each other (Carnine, Silbert, & Kameenui, 1990; Drecktrah & Chiang, 1997).
Although the debate over the two approaches still continues to this day, over a
decade ago the U.S. government took steps to determine the most effective approaches to
teach reading to the nation’s school children. In 1997 the U.S. Congress commissioned a
panel of experts to examine and present a report on the most effective, research based
methods to teach students to read. This panel was made up of a group of experts from the
different areas of reading, teacher education, and scientific inquiry. Their task was to
conduct an in-depth analysis of the research literature on reading instruction and
presented report to Congress on the most effective methods for teaching reading to
children. The report presented to Congress on April 13, 2000 and was entitled: *Teaching
Children to Read: An Evidence-Based Assessment of the Scientific Research Literature
on Reading and Its Implications for Reading Instruction* (National Institute of Child
Health and Human Development, 2000).

This report indentified five vital skills that every student should acquire in order
to become an effective reader. These skills are phonemic awareness, the alphabetic
principle, vocabulary, reading fluency, and reading comprehension. Each skill builds on
the previous and the combination of all five is important for the overall success in reading.
The following section provides a brief overview of each component and its importance
for developing into a strong reader.

*Components of Reading*

*Phonemic awareness.* This is a basic reading skill and usually one of the first set
of skills taught in early literacy instruction (DeBaryshe & Gorecki, 2007; Jule, 1998).
Phonemic awareness involves the ability to identify and manipulate units of spoken
language. Language can be broken down into small units and each of these units produce specific sounds. These sounds can be combined to make an entire word which is named blending, for example the sounds /C/-/A/-/T/ blend together to make the word cat. Words can also be broken apart to make smaller units, called segmenting. For example the word cat can be split into the three letter sounds /C/-/A/-/T/, (Ehri, 2000). This may not develop easily in all children and needs to be explicitly taught to those students at the highest risk for reading failure (Bursuck & Damer, 2007). Phonemic awareness can be taught using a wide variety of classroom activities and there is a growing amount of research to support the instructional effectiveness of these strategies (Cihon, Gardner III, Morrison, & Paul, 2008; Hindson, Byrne, Fielding-Barbsley, Newman, Hine, & Shankweiler, 2005; Koutsoftas, Harmon, & Gray, 2009; Thornton & Vinzant, 2000; Yeh & Connell, 2008).

Alphabetic principle. Another key component of reading instruction is the alphabetic principle. Since the English language is based on a standard set of symbols or letters, students need to acquire the ability to distinguish what sounds each symbol or letter makes. To complicate this process the English language is filled with inconsistencies when specific letter combinations are put together. These inconsistencies may make it more difficult for some students to learn how to read compared to others (Spencer, 2002). Despite the inconsistency, the use of phonics instruction has been shown successful in teaching students the relationship between written (graphemes) letters/words and the sounds that the make (phonemes) (Mesmer, 2005; Villaume & Brabham, 2003). This is largely due to the fact that there are more consistent
relationships between the written and spoken words in the English language (Bursuck & Damer, 2007).

*Reading fluency.* The ability to read a passage of connected text with speed, accuracy, and proper expression is the definition of fluent reading. This skill is very important because students that read fluently are much more likely to comprehend what they read. There is a positive correlation between reading fluency and reading comprehension (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Martens, Eckert, Begeny, et al., 2006; Reis, McCoach, Coyne, Schreiber, Eckert, & Gubbins, 2007). Researchers speculate that this is because once students have the ability to read at a fluent level, they are much more likely to concentrate on the meaning of the words/text as opposed to trying figure out how particular word is read.

*Vocabulary.* As students advance through grade levels in school, they are faced with increasingly difficult instructional material. There comes a point in time that they are no longer taught how to read but are expected to gather information from reading (Harlaar, Dale, & Polmin, 2007; Spor, 2005). An important aspect of reading ability is to understand the meaning of the words that are presented in text. There is evidence that a child’s vocabulary or the number of words they know the meaning of, is directly related to their school trajectory. That is, the larger the vocabulary, the more successful a student will be in school and ultimately in life (Hart & Risely, 1992). Although daily activities such as conversing, reading, and writing are believed to contribute to vocabulary acquisition, there are direct instruction strategies shown effective in increasing the
vocabulary of struggling students (Fore, Boon, & Lowrie, 2007; Lubliner & Smetana, 2005).

Comprehension. The overall goal of reading is the ability to understand what one reads. Durkin (1993) defines comprehension as “intentional thinking during which meaning is constructed through interactions between text and reader”. Reading comprehension is the pinnacle of reading and it is why a person engages in reading in the first place. It is the culmination of all of the sub-skill and the area of reading that gets the most attention by researchers, teachers, and parents alike. There is a plethora of research on reading comprehension and strategies to improve it in struggling readers (Antoniou & Souvignier, 2007; Nelson & Manset-Williamson, 2006; Reis, Eckert, McCoach, Jacobs, & Coyne, 2008; Thames, Reeves, Kazelskis, York, Boling, Newell, & Wang, 2008). The link between reading fluency and comprehension is particularly important because the more fluent a student reads; the more material he/she will cover and the more knowledge will be acquired.

Oral Reading Fluency

As previously stated, reading is the ability to read a passage of connected text with speed, accuracy, and with proper expression. Oral reading fluency (ORF) is the rate at which a student reads a particular passage aloud. Many times ORF rates are measured using novel or unfamiliar passages and one-minute timed tests. In recent years the development of measurement tools such as the Dynamic Indicator of Basic Early Literacy Skill (DIBELS) (Good & Kaminski, 2002) have gained credibility in the academic arena as being reliable assessments for measuring the reading abilities of students (Burke &
Hagan-Burke, 2007; Hagan-Burke, Burke, & Crowder, 2006); both of these studies compared the DIBELS sub-tests to a norm referenced, standardized test of reading and the results indicated that the DIBELS produced similar outcomes to the standardized assessment. These types of fluency measurements are becoming more widely accepted as reliable indicators of reading abilities and predictors of future reading performance (Burns, Dean, & Klar, 2004; Deno, 2003; Fuchs & Fuchs, 2004). Many times these measures are used as screening tools to identify struggling students and monitor their progress as they are exposed to various academic interventions.

Results of research have indicated a strong correlation between ORF and overall reading ability (Baker, et al., 2008; Fuchs, et al., 2001; Good, Simmons, & Kame`enui, 2001; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008). A recent study conducted by Schilling, Carlisle, Scott, and Zeng (2007), used scores from the DIBELS to predict the reading scores on an end of the year, state mandated reading achievement test. The participants in this study included first, second, and third graders who attended 44 different schools across 9 districts in a Midwestern state. These students were given the DIBELS ORF sub-tests at three different times during the school year. In the spring, the same students were administered a statewide reading assessment test. The results of the DIBELS oral reading fluency subtest were predictive of how the students scored on the statewide reading test conducted at the end of the year. Specifically, students who were considered at high risk for reading problems as indicated by the DIBELS also scored below grade average for reading on the end of the year assessment. Study findings were
that a third of the students who were low risk for reading problems on the DIBLES ORF sub-test also scored at or above grade level on the end of the year assessments.

Decoding and Oral Reading Fluency

The ability to accurately identify words is important to the overall success of reading. When a beginning reader encounters a novel word, they will generally attempt to apply the rules learned in phonic instruction to decode the word. This can be accomplished by sounding out individual letter sounds, clusters of sounds, or recognizing the entire word. A key element for decoding is the speed and accuracy with which a student can read words. This concept, called automaticity, was introduced by LaBerge and Samuels (1974) and contends that word decoding becomes increasingly automatic in experienced readers. When students are successful at this skill, they are able to spend less time trying to figure words out and more time reading a given piece of text. Although it seems reasonable to assume that students who can decode are also fluent readers, there are those who fail to develop adequate reading fluency even though they exhibit decoding skills.

A study conducted by Harn, Stoolmiller, and Chard (2008), examined the relationship between nonsense word fluency (NFW) and oral reading fluency. This study used the DIBELS NWF and ORF sub-tests with a large number (938) of first grade students in a Pacific Northwest school district. The NWF sub-test was administered three times throughout the year, in the fall, winter, and spring and the ORF sub-test was given in the winter and spring.
The overall results demonstrated that for many of the participants, when their NFW scores increased so did their ORF scores, but there were some participants where this did not hold true. A closer analysis of this group revealed that while they engaged in decoding, it was considered either a sound-by-sound approach or a sound-recoder approach. In the sound-by-sound approach, students sounded out each letter of a word and in the sound-recoder approach; they sounded out each letter and then put all of the sounds together to complete the word. Approximately one third of the participants in this study were classified as sound-by-sound or sound-recoders. These participants scored significantly lower on the spring ORF tests than did the participants that read non-sense words as a unit. In fact, 70% of the students in the sound-by-sound or sound recoders groups failed to reach the spring benchmark on the DIBLES ORF sub-test whereas only 5% of the participants that read words as a unit failed to reach ORF benchmarks.

More evidence exists to support the notion that not all individuals who exhibit decoding skills automatically make improvements in ORF (Kourea, 2007; Reading & Van Deuren, 2007). For instance, Kourea provided explicit instruction of phonemic awareness activities and a supplementary oral reading fluency activity to first grade students identified as at-risk for reading failure. Throughout the course of the study, students were administered the DIBELs nonsense word fluency, phoneme segmentation fluency, and oral reading fluency subtests to monitor their progress. The results of this study revealed that participants at the greatest risk for reading failure (i.e., the treatment group that had the lowest scores on progress monitoring assessments) reached benchmark on NWF and PSF sub-tests but failed to reach benchmark for ORF at the end of the year.
Even though these participants reached benchmark for the sub-tests that are associated with good decoding skills, they did not improve their oral reading fluency enough to no longer be considered at risk.

An interesting note in this study was the role of the fluency building activity. The investigator included a fluency activity that was specifically developed to target oral reading fluency in addition to the phonemic awareness activities. The author speculated that enough emphasis may not have been given to this portion of the intervention because it only consisted of a 5-minute activity. It was recommended that future research focus on expanding the fluency activity to determine if it would be effective in increasing the ORF scores of at-risk students.

The combined results of these studies point to the need to provide explicit and systematic instruction to increase oral reading fluency. Although teaching decoding skills is important and will likely increase the ability of many students to read connected text with speed and accuracy, it may not be enough for students with persistent reading deficits. Even with instruction on phonemic awareness skills that produce increases in decoding, it is still possible for struggling readers to fail to develop grade level ORF abilities. Decoding skills are a corner stone of an efficient reader but instruction on this ability may not be enough to decrease the overall risk level of struggling students. Even with the supplemental fluency activity provided by Kourea (2007), students exhibiting persistent reading difficulties did not improve sufficiently, indicating the need to focus directly on increasing oral reading fluency, especially in light of the potential impact it has on reading comprehension.
Oral Reading Fluency and Comprehension

The overall goal of reading is for the reader to understand and therefore obtain information about what is being read. Reading comprehension is the primary concern for many educators because it is the key indicator that students are obtaining the information they need to perform adequately on academic tasks. Researchers have extensively focused on reading comprehension and strategies used to improve it (Antoniou & Souvignier, 2007; Berninger, Abbott, Vermeulen, & Fulton, 2006; Gabl, Kaiser, Long, & Roemer, 2007; Kim et al., 2006; Thames et al., 2008). In these studies, a strong link between reading comprehension and reading fluency was found. In addition, numerous supporting studies indicate a direct relationship between improved fluency and reading comprehension (Buchanan, 2006; Cates, Thomason, & Havey, 2007; Fuchs et al., 2001; Klauda & Guthrie, 2008; Therrien & Hughes, 2008).

An earlier study investigated the differences that may exist between fluency and comprehension scores for readers who either engaged in context-free reading versus reading in context (Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003). Context-free reading tasks consisted of reading lists of words that were not in an organized format; whereas reading tasks that were in context consisted of reading connected text passages. Participants were asked to read both from a word list (context-free) and a connected text passage. The words used for the list came from the same passage the participant received for the context reading phase. The participants were also administered a standardized reading assessment that measured comprehension.
The results of this study revealed a strong correlation between reading speed and comprehension; however, the correlation was higher for context reading than for list reading. That is, context reading was a better predictor of reading comprehension abilities than was word list reading. Although the results of this study indicated that context reading fluency was an important predictor of overall reading skills, the authors also pointed out that this largely depended on the reading level of the student. For example, students that were identified as slow readers on the connected text passages also exhibited low abilities to read from the word list. The authors surmised that reading in context is the intermediate step or link between list word reading and reading comprehension. Also, reading in context uses the same processes as comprehension and therefore an individual is more likely to understand what he/she is reading when it is in context as opposed to just a list.

Recently, more researchers have focused on the link between reading fluency and reading comprehension to provide empirical evidence of the positive relationship. A study conducted by Riedel (2007) aimed to compare the results of the DIBELS benchmark sub-tests to comprehension and vocabulary scores of a group of urban first grade students. The main purpose of this study was to determine if the DIBELS sub-tests (NWF, PSF, ORF, letter naming fluency, and retell fluency) were accurate predictors of end of the year reading comprehension scores for first and second grade students. A large sample of urban, primarily African American, first-grade students were selected for this study and were administered the DIBELS benchmark sub-tests at the beginning, middle,
and end of the year. A standardized reading assessment was given to the same group of students at end of first and second grade.

The results of the first grade comparison revealed that NWF and LNF were better predictors of comprehension deficits than was PSF for the initial (i.e., beginning of the year) assessment. By the end of the year assessment, ORF became the strongest predictor of comprehension problems with over 80% accuracy for predicting deficits. This sub-test even out-scored the retell fluency, which was designed to test for comprehension. The results of the second grade comparison were very similar with the ORF sub-test having the highest accuracy in predicting reading comprehension deficits in the participants. The results of this study support the use of oral reading fluency measures to identify and predict reading deficits in young readers.

Oral Reading Fluency Instruction

Repeated Readings

Repeated readings refers to procedures in which students read the same passage until they are able to read it with the speed and accuracy to meet a certain criterion (e.g., 40 correct words per minute). This intervention was first introduced (Samuels, 1979) to increase an individual’s ability to read fluently. Samuels conducted a study in which he had students read a short passage to a teaching assistant who recorded the speed and accuracy of the reading. Each student was asked to continue rereading the passage four consecutive times. After the fourth reading the number of words read during a minute timing was determined by subtracting the number of error for the total number of words read. The results of this study demonstrated that many of the students improved their rate
of reading following practiced re-readings while simultaneously decreasing the number of errors that they made.

Samuels believed that one benefit of this type of intervention was the fact that it was easy to implement (i.e., teaching assistant, parents, and even other students could record rates), which would free up time for a teacher to engage in other instructional activities. Also, the intervention an audio recording while reading the story or using the teaching assistant/parent/peer as a tutor for more difficult or unknown words. Samuels contended that the use of repeated readings could be a valuable supplemental intervention for struggling readers.

*Early repeated reading research.* Early researchers used repeated reading strategies to test their viability as a way to increase the ORF of many readers (Dowhower, 1987; Koskinen, & Blum, 1986; Moyer, 1982; Rashotte & Torgesen, 1985, Rasinski, 1989; Swain & Allinder, 1996). These researchers were concerned with determining how effective repeated reading strategies were; as well as, the critical aspects of the strategies. Rashotte and Torgesen (1985) investigated the use of repeated reading with students diagnosed with learning disabilities. These authors were concerned with the lack of information about the types of reading passages that were being used in repeated reading studies. Specifically, they investigated the amount of word overlap and similarity in context between reading passages had any effect of the reading speed, accuracy, and comprehension of the participants. The authors also examined if the use of non-repetitive reading was just as effective as repeated reading for increasing fluency. A third condition
consisted of the participants reading several different stories during a session as opposed to reading the same story repeatedly.

Each student was exposed to all three conditions and the general procedures for each condition were similar. Students were given grade level reading passages and instructed to read it four times in a session (with the exception of condition three). At the conclusion of the first reading, they were presented four comprehension questions and then one question following the remainder of the readings (i.e., one question after readings 2, 3, and 4). In condition one, students were instructed to read one of seven stories that were all unrelated and did not contain a high number of overlapping words. In condition two, students were instructed to read one of seven stories that were highly related and contained many overlapping stories. In the final condition, students were instructed to read four different stories in one session, none of which were repeated.

The results of the first comparison (repeated reading stories that shared word overlap and context versus those that did not) demonstrated that there was a significant difference in ORF for the word overlap/high context stories. In other words when students read stories that had very similar words and meaning, they demonstrated a higher rate of speed and accuracy than when the stories were different. The authors indicated that these results lend support to the LaBerge and Samuels (1974) theory of automaticity in that the more times readers are exposed to the same words the faster and more automatic they recognize them.

Although there was a difference between the repeated reading conditions, there was not a significant difference between condition 1 and condition 3 and the difference
between conditions 2 and 3 only approached significance. This suggested that the use of the same story in repeated reading strategies might not be as important as simply providing more opportunities to read because there was no statistical difference between the conditions in which the repeated reading was implemented and the condition in which the participants read four different stories only one time each.

The authors speculated that these results may have differed from Samuels (1979) study because of the differences in the procedures. In this study, participants read the same story for a specified number of times (i.e., 4 times per session) whereas in the 1979 study the participants read the same story until they reached a specific criterion. The authors also noted that the participants indicated that they enjoyed the repeated reading sessions and regardless of how they performed requested to continue with the sessions even after the study was concluded.

Dowhower (1987) conducted a study that focused on the use of repeated reading with a group of second grade transitional readers. These participants were considered transitional because they were reading at or above grade level in one area of reading (i.e., word decoding) but below grade level in another area (ORF). In the case of this study, participants were selected if they were reading below 50 correct words per minute on 200 word passages but they were able to accurately decode at least 85% of the words in the same passage.

A total of 19 participants were selected for the study and were randomly assigned to one of two conditions. The first condition was called assisted repeated reading in which each participant listened to an audio recording of the passage and was instructed to
read along. They continued to reread the passage until the reached the criterion of 100 words per minute. In the second condition (i.e. the unassisted) participants were instructed to read the passages on their own and did not receive any assistance for unknown words. They needed to read the passage until they reached the 100 word per minute criterion as well.

The training procedures consisted of five separate steps. First, the participants were given an initial test that consisted of an unfamiliar, 200 word passage. There were a total of five different passages and data were collected on the rate (words read per minute), accuracy (number of correct words read), comprehension (number of correct recall questions), and prosodic measures. Following the initial test, participants were either introduced to the assisted or the unassisted condition. In each of these conditions, the participants were given a passage pre-test, which used the same dependent variables as the initial test. Then the participants engaged in the assisted or the unassisted repeated reading conditions. Once a participant met criterion of 100 words per minute, they were given a passage post-test to ensure that they could read the treatment story to criterion. If they demonstrated criterion reading on the passage post test, they were given a passage transfer test that consisted of different but similar stories that were read in the training sessions. Each participant repeated the training sequence with five different stories. Finally, they were given a passage post-test which was used as an overall assessment of reading gains.

The results of this study yielded several findings. First, there were significant increases in reading rate (fluency), accuracy, and comprehension scores for both the
assisted and the unassisted groups. All of the participants began below grade level on reading rate, accuracy of word reading, and comprehension. Following the repeated reading intervention, both groups made significant gains and improved their overall reading to grade level (according to standardized reading assessments). There was also an improvement in how the participants read (i.e., the prosodic reading abilities). They went from being word-by-word readers to reading increasingly longer word segments with fewer pauses and proper intonation. Another finding involved the transfer skills of both groups. This involved reading similar but novel stories at the end of each training sequence. The goal was to determine if the participants would be able to generalize the reading skills that they had acquired to a novel passage. The analysis indicated that both groups were able to maintain similar reading rates from post-training reading stories to the transfer (novel) story readings. When comparing the combined average scores on the initial test to the final test there were considerable gains for reading rate, accuracy, and comprehension.

The results of this study not only provided support for the use of repeated reading procedures to increase the reading skills of struggling students, it also extended the literature in two important ways. First, it used novel stories to determine if the participants’ reading abilities would generalize. One possible criticism of repeated reading strategies is that it uses the same passages over and over and therefore may not be a true measure of reading fluency. By using novel passages to test for generalization, it can be argued that the performance on novel passages is a more accurate measure ORF. Another way that this study extended the literature was to use prosodic measures. Oral
reading fluency is often described as the speed and accuracy of an individual’s reading but this definition also included reading with meaning. This measure is often overlooked in lieu of focusing on speed and accuracy. This study introduced analyses to directly measure the prosodic aspect of reading such as intonation and the length of pauses between words.

Refining repeated reading procedures. After the reliability of repeated reading was confirmed, researchers have since sought to refine strategies to increase ORF and comprehension (Buchanan, 2006; Cates, Thomason, & Havey, 2007; Fuchs et al., 2001; Huang, Nelson, & Nelson, 2008; Klauda & Guthrie, 2008; Meyer & Felton, 1999; Therrien, 2004; Therrien & Hughes, 2008; Valleley & Shriver, 2003; Yurick, Robinson, Cartledge, Lo, & Evans, 2006). In one recent study, (Yurick et al., 2006) conducted a series of investigations using peer mediated, repeated reading procedures to increase the oral reading fluency rates and comprehension of the participants. This study consisted of three experiments to investigate the effects of the peer mediated, repeated reading intervention. The initial experiment included eight, fifth-grade students who were considered at risk for reading problems (i.e., they performed below grade level on a standardized test). A multiple baseline across participants design was used to determine the effects of two conditions on the reading and comprehension scores of the participants.

In the first condition, which served as a baseline, participants were given a 200-word reading passage and instructed to read it to themselves for a total of 10 minutes. This condition was call sustained silent reading (SSR) and was used to mimic typical classroom reading practice. Following the completion of the 10 minute SSR, each
participant was instructed to read the passage aloud to the experimenter and then asked five comprehension questions. The participants were not told they were being timed and they were allowed to complete the entire passage but data were collected on only the first one minute of the aloud reading.

Prior to the second condition being implemented, students were paired together and provided with training sessions. These sessions consisted of guided practice for the role of the listener, the role of the reader, and then independent practice. Once the pair was able to demonstrate the entire procedure without any prompting, they were instructed to engage in the repeated reading condition. This condition was called paired repeated reading.

This repeated reading condition consisted of partner reading in which a pair of students took turns reading paragraphs of an assigned passage. The students were previously trained on a correction procedure if their partner made an error in reading. This correction procedure consisted of stopping the reader if an error was made. The listener instructed the reader to “stop and sound it out,” then to read a group of words that included the missed word, and finally to either say the group of words three times fast or say the group forward and backward. The investigator and an assistant monitored the groups and provided help as needed. If the pair completed the passage prior to the 10 minutes elapsing, they were instructed to start back at the beginning of the passage.

After the 10-minute session was over, each student was asked to read the passages aloud for one minute. Following the one minute timed reading, the participant counted the number of words read in the first minute and plotted it on a graph. The participants
continued to use the same grade level passage until they reached the criterion of 180 words per minute with 10 or fewer errors. Once they reached this criterion, each participant was asked five comprehension questions associated with the passage. If they answered all of the questions correctly they were moved to the next grade level passage. The passages began one grade level below the actual grade the students were in (e.g., 4th-grade level) and progressed to what was considered young adult level passages. Following the completion of a grade passage, participants were rewarded with tangible items for their progress.

The results of this first experiment indicated that the peer mediated, repeated reading procedure increased the ORF, accuracy, and comprehension of the participants over the sustained silent reading. All of the participants were reading well below grade level ORF rates for 5th grade during baseline measures. The participants averaged approximately 85 words per minute on the 4th-grade level reading passages used for the SSR in baseline. Once the paired reading condition was implemented, all but one of the participants were able to advance to the 6th-grade reading level passages and six of the participants advanced to the young adult reading passages. There was an average increase of 68 words per minute for the participants in the paired reading condition over the SSR condition. There were similar gains on the comprehension measures where participants were averaging between 1.1 and 3.8 correct answers in baseline and this increased to 5.0 correct answers for most of the participants in the paired reading passages.

Experiments 2 and 3 were conducted to strengthen the results of the first experiment. These experiments used the same general procedures with different groups of
students. The procedures were adjusted slightly for improvement; however, they followed the same format (i.e., SSR followed by peer mediated repeated reading). A key difference in experiment two was the use of third grade students. Also there was the implementation of generalization probes to test for ORF and comprehension on untrained reading passages. The differences for experiment 3 were the use of fourth-grade students and the way the repeated reading training occurred. In experiment 3, a pull-out, small group model was used in which instruction took place outside of the classroom. The results of these experiments were similar to the first experiment in that the repeated reading procedures increased oral reading fluency rates over the sustained silent reading.

*Error correction procedures.* Another area of research on repeated reading strategies focused on error correction procedures. Systematic procedures for correcting reading errors were developed and studied for their effectiveness (Begney, Daly III, & Valleley, 2006; Begeny & Silber, 2006; Eckert, Dunn, Ardoin, 2006; Martens et al., 2007; Nelson, Alber, & Gordy, 2004; Pany & McCoy, 1988). A study published in 2004 examined a systematic error correction procedure to examine its effects on reducing reading errors and increasing correct words per minute (Nelson, Alber, & Grody, 2004). Four second-grade students diagnosed with learning disabilities were exposed to a range of treatment options to see if they would increase their ORF and decrease errors. These treatment options consisted of baseline, a systematic error correction procedure, error correction procedure paired with repeated reading, and the use of previously practiced material with the combination of error correction and repeated reading.
During baseline, each participant was given a reading passage and instructed to read it for five minutes. Following the five minutes, the participant was instructed to read the same passage from the beginning for one minute while the experimenter recorded the number of correct and incorrect words. The next condition was the systematic error correction procedure. In this condition, each participant was instructed to read a different passage for five minutes, but every time he/she made a mistake the experimenter stopped them and implemented the correction. This correction procedure consisted of the experimenter stopping the student and identifying the incorrect word, having the participant repeat the word correctly, and then read the sentence again. At the end of the five minute training session, the experimenter reviewed each missed word with the participants by pointing to the word and asking them to read it. If they still made an error, the experimenter provided the correct pronunciation and had the participant read the word again. The error correction plus repeated reading condition was similar to the previous condition except the error correction procedure only lasted for three minutes. Following the three minute reading, the participants were instructed to engage in a repeated reading exercise. During the repeated reading portion of this condition, participants engaged in three one-minute readings of the same passage. The final timed reading served as the reading in which the number of correct and incorrect responses was recorded. In the final condition, the same exact procedure was used as the previous condition with the exception of the material. In order to control for the increasing difficulty level of the passages through the study, the final condition used passages from the original baseline due to their decreased difficulty level.
The results of this study yielded several findings. First, there was an overall increase in correct words per minute when comparing baseline to all the other conditions. The largest overall increase in ORF was following the implementation of the repeated reading plus error correction condition with an increase of nearly 20 correct words per minute. The second finding involved the decrease in errors. The incorrect words per minute decreased on average across all treatment conditions, compared to baseline levels. The largest decrease in errors was in the final condition which used previously read passages. This was likely due to several factors. For one, participants were engaged in a combination of the error correction procedures and the repeated reading. As demonstrated in the other conditions, both of these treatment strategies were effective in increasing correct words per minute and decreasing the number of mistakes. When they were combined with reading material that the participants were previously exposed to, fewer errors were exhibited. The results of this study demonstrated that if response errors are corrected immediately and systematically they can be practically eliminated. The results also provided support for the use of efficient yet effective procedures to correct reading.

More recently, another correction procedure has been investigated as another possible way to increase ORF and to reduce reading errors (Begeny, Daly III, & Valleley, 2006). This procedure is called phrase drill correction and was previously used in research as a way to reduce reading errors (O’Shea, Munson, & O’Shea, 1984). Phrase drill correction involves having a student read a selected passage while the instructor records the words that are read incorrectly. Following the completion of the reading passage, the instructor has the student read a phrase or series of phrases that contain the
incorrect words. The student was then instructed to re-read the passage to determine if he/she could accurately read the missed words.

Begeny et al. (2006) used an alternating treatment design to determine which of three treatment approaches would be most effective to increase the ORF of participants. The three treatment approaches were repeated reading alone, repeated reading with phrase drill error correction, and a reward condition where the student was rewarded for faster readings. In the repeated reading alone condition, the participant was required to read a passage two times in a row, before being checked for ORF. Oral reading fluency was assessed by having the participant read the same practice passage for one minute and calculating the number of words read correctly. The phrase drill condition was similar to the repeated reading condition with the exception of a correction procedure. During the first phrase drill reading, the experimenter recorded any words that the participant missed. Following the reading, the experimenter reviewed each missed word by presenting the correct pronunciation to the participant and had him read the word in a short phrase three times. After the correction procedure the participant was instructed to read the passage again to test for ORF. During the reward condition the participant was instructed to read a passage and then given a different passage and instructed to read it as well. If the participant read more correct word on the second passage, he was granted access to a high preference tangible reward. There was also a baseline condition in which the participant was simply asked to read a passage for one minute to check for ORF.

The results of this study indicated that all three treatments increased the ORF of the participant over baseline; however the repeated reading and the phrase drill correction
conditions were associated with the highest ORF rates. The participant performed better in these two conditions than in the reward and baseline conditions. Although the reward condition was not as successful as the other two conditions, the use of rewards did increase ORF over baseline. It should be noted that the reading rates for the reward condition was actually lower than baseline for the last session. In fact, with the exception of baseline, ORF rates were on a decreasing trend for all of the conditions throughout this study. This may have been due to an increase in difficulty level as the experiment progressed. This difficulty level increased by a quarter to a half a grade level for each condition from session to session. In terms of reducing the number of incorrect words read per passage, the phrase drill condition produced the fewest errors of all four conditions.

Repeated reading and generalization. A very important consideration when it comes to ORF and comprehension strategies is how well they translate to unpracticed reading material. Much of the more current repeated reading research includes checks for generalization (Ardoin, McCall & Klubnik, 2007; Begeny, Daly III & Valleley, 2006; Dowhower, 1987; Therrin & Kubina, 2007; Martens, et al., 2007; Vandenberg, Boon, Fore, & Bender, 2008; Welsch, 2007). Ardoin, McCall and Klubnik (2007) published a study that investigated two different types of repeated readings. The purpose of this study was to determine the differential effects of a standard repeated reading phase compared to a multiple exemplar repeated reading phase on generalization.

Prior to either phase in baseline, the participant was given a novel passage and was asked to read it. The same passage was given to the participants again following the
intervention to determine if there was any increase in ORF. In the standard repeated reading phase, participants were exposed to a pre-selected training passage that corresponded with the generalization passage. They were then introduced to a repeated reading procedure. During this procedure the experimenter first read the training passage to the participants and then they were allowed to read the same passage four times. After the first reading, the participants were exposed to a correction procedure in which the experimenter presented all of the errors made during the reading. That is, the experimenter presented each word that was missed, provided the correct pronunciation of the word, had the participant repeat the word correctly. Then the participants were instructed to repeat a 2 to 3 word phrase that contained the missed word. During the subsequent three readings if any words were missed, the experimenter implemented a word level correction procedure. During this correction procedure, each missed word was presented to the participant and they were instructed to segment and blend the word. Following the fourth and final reading, each participant was given the generalization passage again and asked to read it. They were able to earn tangible rewards if they beat their pre-intervention ORF rate or if they read more than 100 words within a minute with only one mistake.

The alternate phase in this study was called the multiple exemplar phase. Participants were first instructed to read the generalization passage to establish a baseline. Following this baseline reading, they were given one passage and exposed to the same procedures as the standard repeated reading phase. That is, the experimenter read the passage once to the participants and then instructed the participant to read the same
passage twice. After the second reading, the participants were given another (different) story and the same procedures were followed. The correction procedures in the multiple exemplar phases were identical to the standard phase. The key difference between the standard repeated reading phase and the multiple exemplar phase was the number of stories used and the number of times each story was read. In the standard phase, one story was read four times and in the multiple exemplar phase two stories were each read twice.

The results of this study demonstrated that both intervention phases increased the oral reading fluency rate of the participants but the standard repeated reading phase produced overall higher rates of ORF when compared to the multiple exemplar phase. In addition to this finding, it was demonstrated that the participants were able to increase their ORF on the generalization passage. That is, following the repeated reading intervention, the ORF rates were higher on the generalization passage than they were on the same passage prior to the intervention.

As a follow-up to the 2007 study, Ardoin, Eckert, and Cole (2008) conducted another study that compared a standard repeated reading procedure with a multiple exemplar procedure. During the repeated reading phase, participants were instructed to read a passage three times in a row. The experimenter timed them for one minute to check for ORF, but allowed the participants to finish reading the entire story. Immediately following the repeated reading, participants were given one of two generalization passages to read. One of the passages contained high word overlap and the other passage contained medium word overlap. Oral reading fluency measures were collected on both passages for all of the participants.
The second phase was a multiple exemplar strategy. During this treatment participants were instructed to read three variations of the same story. The experimenters selected a story and rewrote it a couple of times by rearranging sentences, words, and replacing words with synonyms or antonyms. Participants read the stories one time each and then were immediately given a high word overlap or medium word overlap, generalization passage. Oral reading fluency passages were collected on both generalization passages.

The results of this study were similar to the 2007 Adorin et al., study in that the repeated reading strategy produced higher rates of ORF on the practice passages. The results of the generalization probes were less clear. Participants had higher rates of ORF on generalization passages that contained medium word overlap but there were no significant differences between the treatments on the high overlap passages.

Multi-component interventions. As demonstrated in several of the previously discussed studies, repeated reading interventions may be enhanced by using them in combination with other strategies such as error correction procedures, peer tutoring, and use of positive reinforcement. Recently, several studies have used multi-component approaches to determine if they would be effective in increasing ORF of struggling readers (Alber-Morgan, Ramp, Anderson, & Martin, 2007; Martens et al., 2006). In the Martens study a group of second- and third-grade students were assessed on their ORF as well as standardized reading assessments. These students’ scores were also compared to a group of control students who did not receive any supplemental reading instruction (i.e., they received only their classroom reading activities).
The treatment group participants were exposed to a training sequence that consisted of several different components. First these participants were given pre-training assessment passages that corresponded to their current level. They were asked to state their goal for reading the passage (which was 100 correct words per minute) and told if they reached the goal they would earn a ticket towards a prize. The participants were then instructed to read the passage and although they were allowed to read the entire passage, data was only collected during the first minute. If the participant reached the goal, they were given a ticket and presented with another, more difficult passage. If they failed to reach the goal, that passage was used as the training passage and the participants preceded to the training session.

During training sessions participants were exposed to the multi-component, fluency building activity. The first component was the use an error correction procedure. This study used a phrase drill correction of the errors made on the pre-training passage. The participants were instructed to read a short phrase three times, which contained the missed words from the pre-training passage. The next component was called listening passage preview and it consisted of the experimenter reading the passage to participants at a fluent rate as they followed along. Another component was a repeated reading strategy in which the participant and the experimenter took turns reading the passage two times each. Following the training session, the participants engaged in a post-training assessment that was identical to the pre-training assessment in that ORF was assessed on the training passage.
An additional component of this study was to test for retention. This was done by having the participants reread the training passage two days after the post-training assessment data was collected on it. Participants were required to reach a specific criterion on this retention check before they could progress to the next passage level. The passages progress from pre-K level all the way through fourth grade level. In order to test for generalization, reading probes were administered prior to and after untrained passages. These probes consisted of untrained passages on 2nd, 3rd, and 4th grade levels. In addition to the generalization probes, a pre and post standardized reading assessment was given as the basis for between group comparisons.

The results of this study indicated that the multi-component, repeated reading intervention was effective in the following ways. First, the treatment group participants for both grade levels (i.e., second and third grade) made more significant gains on the untrained generalization passages and the standardized test scores compared to the control group students. Specifically, the second grade treatment group participants made significant increases on two of the three grade level probes, whereas the control group participants made significant increases on only one of the probes. The third grade treatment group outperformed their control group counterparts on all three of the generalization probes.

The second finding pertains to the within subject analysis of the immediate and retention gains on the training passages. Both second and third grade treatment groups made substantial progress in terms of increases in fluency and progression through the grade level training passages. Second grade students gained an average of almost 23
words per minute and 21 words per minute on the immediate and retention passages respectively. They also progressed through over two grade levels worth of passages during the course of the study. Similar gains were found for the third grade students, who made gains of over 24 words per minute on the immediate passages and 26 words per minute on the retention checks. These students also progressed through nearly 3 grade levels worth of passages during the study. A third finding demonstrated increases in pre-training fluency levels and decreases in the number of times a training passage had to be used before criterion was met. As participants progressed through the grade level passages, their pre-training ORF scores increased and the amount of time they spent on the passage increased.

This finding indicates that as the participant became more fluent readers, their overall reading abilities may have improved. If their increases were solely due to the training procedures (i.e. the repeated reading), the average pre-training scores would have likely remained unchanged or decreased as the material become more difficult. Since the pre-training scores actually increased as the study progressed and since the number of times to reach criteria became fewer, it can be suggested that the participants’ reading abilities became stronger over the course of the study.

**Summary of Repeated Reading Research**

Over the past three decades, repeated reading strategies have been developed and refined to provide beneficial interventions to promote increases in oral reading fluency. Since the original study by Samuels in 1979, researchers have expanded on simple repeated reading procedures to help increase comprehension and generalization. The use
of error correction and multi-component procedures helped to establish repeated reading as one of the lead strategies to increase ORF of struggling readers.

Early research was conducted to determine if the use of a repeated reading procedure could be replicated with various groups of students. These groups consisted of students with and without reading difficulties and students from various grade levels. The majority of the published research determined that repeated readings were an effective strategy to increase oral reading fluency and comprehension.

The success of the early research sparked a bevy of new studies on the topic of repeated reading and oral reading fluency. Researchers turned their attention to more fine-grained analyses of procedures and became increasingly concerned about generalization of skills to unpracticed reading material. Error correction, peer mediated, and multi-component procedures were designed and tested for their effectiveness. Researchers also zeroed in on at-risk populations in order to determine the best approaches to remediate reading difficulties. The overall results of much of this research supported the success of repeated reading strategies as effective interventions.

Implications for Classroom Use

Even though the success of repeated reading strategies has proven successful in the research literature does not automatically translate to classroom practice. Though many of the procedures are relatively brief and easy to implement there are some possible factors that may pose problems to classroom teachers. First are the time and personnel constraints of a typical classroom environment. Teachers are likely to have specifically scheduled times for reading instruction throughout the day. Although, a selected repeated
reading procedure may only take a brief period of time to implement the teacher may not have to time or resources to do so. Considering that much of the research was conducted using individual or small group instruction, the viability of classroom implementation can be called into question.

The second possible problem with classroom implementation involves treatment integrity. Most procedures call for very specific steps to be implemented in order for the treatment to be effective. It has not yet been determined if these procedures would be effective if they were not implemented with a high degree of fidelity. Often classroom schedules need to be flexible to accommodate for unexpected situations that often arise in a school day. While teachers may be skilled at adapting to these changes, the consistency and accuracy of instruction may suffer. The use of repeated reading procedures has held up well under experimental conditions, but it still needs to be determined how well they hold up under more typical classroom conditions too.

Computer Assisted Instruction for Reading

Introduction of Computer Technology

*Early computer use.* Research on the use of computers as teaching tools date as far back as the early 1970’s (Ellis, 1974; Fletcher, 1972). Fletcher and Atkinson (1972), reported on the effects of a computer assisted reading program that was developed at Stanford University in the late 1960’s. This program used computers to develop individualized reading instruction for kindergarten through third grade students. The result of this study indicated that students who used the computer assisted instruction scored significantly higher on all but one of the comparison post tests.
Microcomputers and early software programs. Although there was an increased interest in the use of computer technology throughout the 1970’s it was not until the introduction of micro-computers in the early and mid 1980’s that this technology began to gain more attention from researchers and school professionals alike. One major roadblock in the use of computers for individual teaching opportunities was their high cost (Majer, 1973; Marsh, 1983). With the introduction of the microcomputer and the reduced cost of the technology, there was an increase in the use of computer assisted instruction for reading. The promise of a technology as great as computers was touted as the solution to many educational problems. One researcher of this time period predicted that computers could help eliminate 90% of reading failure in the years ahead (Glasser, 1981). While such a statement has not exactly lived up to its billing, the use of computer technology has become an important part of the everyday lives of many school children.

Early research on computer assisted instruction tended to focus on the usability of software programs as opposed to their effectiveness. Several studies conducted in the mid 1980’s provided overviews of software programs that were designed to teach reading skills (Bradley, 1984; Candler & Johnson, 1984). Other studies published in the mid and late 1980’s focused on the implementation of computer based reading interventions with students (Bass, Ries, & Sharpe, 1986; Casteel, 1989; Gore, Morrison, Maas, & Anderson, 1989).

The study published by Bass et al., (1986), compared the reading and math scores of low achieving fourth through sixth grade students. These students were split into experimental and control groups and each was administered several standardized
measures for math and reading. All of the students in the study received their regular classroom instruction but students assigned to the experimental group participated in between 10 and 15 hours of additional computerized reading/math instruction. It should be noted that the study did not provide any information on the specifics of the computerized instruction. Students in the control group received supplementary classroom instruction in addition to their typical classroom instruction but this instruction was not computer-based.

The outcomes of the statistical analyses performed revealed that there were mixed results across the groups at different levels. At the fourth grade level both, the microcomputer and the traditional classroom instructional group improved their scores on the reading and math post-tests. At the fifth- and sixth-grade levels, results were also mixed. Both groups in fifth grade showed increased scores for the reading tests but were split on the math test (i.e. the traditional group scored higher). Similar results were reported for the sixth grade groups but the computer group outperformed the traditional group on the reading measures, whereas the reverse was true for the math tests.

Although the overall results of this study were mixed, it can be determined that the microcomputer instruction performed just as well as the traditional classroom instruction at all levels except one and it surpassed the traditional instruction in reading instruction for sixth grade students. One limitation of this study was the fact that there were not any details regarding the content and operation of the computer assisted instruction. So, although it provided evidence about the validity of using computers in the
classroom as instructional aides, it falls short of guiding school professionals on how to
implement similar programs.

CAI and diverse populations. With the expansion of computer based learning
firmly established in the research literature and the ever growing evolution of the
microcomputer by the late 1980’s and early 1990’s, researchers began to study the
effectiveness of computers for more diverse populations. A population that researchers
became interested in was early primary school students in preschool through first grade.
The use of supplemental instruction can be the difference between success and failure for
some students. So, with the increase in CAI programs as supplemental programs, it
seemed beneficial to study their effect with young children.

A potential obstacle to using CAI programs to teach reading to young children is
their ability to understand how to interact with the technology. To compound this
problem, children in the most need for supplemental instruction come generally from
very low SES backgrounds. Due to this fact, they are very unlikely to have access to a
computer at home and possibly even in school. So, the question of how well young,
disadvantaged children would fair using computer to learn reading was very relevant
during the late 1980’s and early 1990’s. To answer these questions, Gore (1989),
conducted an experiment with 14, five-year-old, preschool students from disadvantaged
backgrounds.

This study used a pre/post-test single group design and administered a
standardized reading test at the beginning and the end of the school year. Each child was
taught specific pre-reading skills via a collection of commercialized computer software
programs. These skills included rhyming, letter recognition and visual matching. In order to control for classroom instruction that could promote an increase in reading ability, the preschool teachers were directed to avoid using drill and practice techniques for the specific skill being taught by the CAI programs. The secondary purpose of this study was to determine if the participants would be able to operate the computers with minimal assistance. In order to test for this, the experimenters provided very basic instructions on the first day of the intervention and then allowed the children to follow the instructions provided by the software itself. It was noted that if the students needed any assistance they were encouraged to ask for it.

The results of this study were very promising. When pre and post tests scores were compared, they indicated that students were closer together in knowledge base and pre-reading skills at the end of the study compared to the beginning. The results of the ANOVA were significant for the pre and post test measures. Individual t-Tests were conducted for each sub-skill and the results indicated that there were significant differences on two of the four skills and scores for a third skill were approaching significance. As for the secondary purpose of determining if the children would be able to operate the software independently, it was reported that there was very minimal need for assistance for any of the participants.

The results of this study are important for a couple of reasons. First, they provided more evidence of the value and effectiveness of computer assisted instruction. In this study, teachers were specifically instructed not to provide any additional instruction on the skills being taught by the computer. From the results it is clear that the participants
were able to improve their reading skills with only the CAI. Second, the results also provided some proof that students, even very young children, can learn to operate computers with minimal assistance. Taken together, these results indicated that teachers may be able to allocate some of their teaching time on more pressing areas and allow students to engage in certain academic activities via a computer. This is important because as pointed out early, many teachers face high student ratios and they may not have the capability to help every student all the time.

_Devolution of CAI to Address Oral Reading Fluency_

With the establishment of computer technology in educational settings it became clear that the combination of computers and effective reading instruction could help bridge the gap for struggling readers. As this technology continued to grow throughout the 1990’s developers and researchers began to customize software programs to address specific areas of reading development. These areas include phonemic awareness, reading fluency, and reading comprehension; however this section will only focus on oral reading fluency.

_Project LISTEN_. Project LISTEN began in the early 1990’s at Carnegie Mellon University with the objective of creating a computer software program to improve the reading skills of school children (Mostow, Hauptmann, Chase, & Roth, 1993). This software used voice recognition technology to assess the reading abilities of users by detecting mistakes and providing corrective feedback. More specifically, the computer recorded each student reading out loud through a noise cancelling, headset microphone. When the reader made a mistake the computer provided feedback in the following ways.
If the reader substituted a different word for the correct word, the computer highlighted the correct word and played back what the reader said. Then, it provided the correct word or phrase of the presented passage. If the reader skipped a word, the program provided corrective feedback by representing the correct word or phrase to the reader. If the reader added words or phrases, the computer ignored them and prompted the reader to continue.

The software program used the Sphinx-II speech recognizer to act as the “listener” for the reader. The basis of this speech engine was to reproduce the speech of a user, so the designers needed to adapt it to instead recognize where a reader differs from a presented text and take an appropriate action to correct mistakes. In order to accomplish this task, the system used three different types of data based text as a comparison to what the reader spoke into the microphone. Based on the information provided to the data bases, a word or phrase was either considered correct or incorrect and feedback is provided as described above. One limitation of this early prototype is that it only provided feedback after an entire page was completed by the reader. This left the potential for the reader to practice errors and develop error patterns, which in turn could affect their overall reading performance.

In order to test the efficacy of this early version of the reading tutor, the experimenters tested its accuracy to detect errors with 27 students in a large, urban elementary school system. Each student was recorded reading from a grade level text. This recoding was then transcribed and the recoding was uploaded into the voice recognition software to determine how accurately it detected any missed words. This was accomplished by calculating the actually percentage of missed words that each reader
produced and then comparing that number to the missed words identified by the program. The results indicated that the actual readers produced a missed word in 2.5% of opportunities or about 1 in 40 words. The speech software detected the same missed words in 63.6% of opportunities, which was roughly two-thirds of the time. While the authors admitted that the software did not produce 100% reliable results, they surmised that the program was of some pedagogical use and with future improvements it could be a very valuable tool to combat reading difficulties in the nation’s schools.

*Reading tutor.* With a workable prototype in hand, the researchers and developers spent next 14 years improving and testing this technology. Numerous experiments have been conducted to test the accuracy and usability of this software (Aist & Mostow, 1997; Mostow, et al. 2001; Mostow, et al. 2003; & Poulsen, Hastings, & Allbritton, 2007).

The first research study using the Reading Tutor was conducted by Aist and Mostow (1997). This was a pilot study to determine the effectiveness of the Reading Tutor. Eight third grade students, who were at-risk for reading, were selected to participate in the study. These students were reported to be the lowest readers in their classes and were 2 to 3 years below grade level in reading.

The Reading Tutor prototype was designed to interact with the student in several ways. It first allowed the student to select a story and then this story was presented on the computer screen one sentence at a time. These stories were selected from a *Weekly Reader* or other sources and it was reported that students had favorite stories which they would often reread. When the students were instructed to read the sentence, they could read aloud, click on an unknown word for help, or click a help button to have the
sentence read by the computer. The Reading Tutor responded when the end a sentence was reached, a four second pause was detected, or it detected a mouse click. If no mistakes were detected, the next sentence was presented. If the program detected a one word mistake (i.e., the reader made an error while reading the sentence), it highlighted the word in pink and spoke it. If the program detected more than one missed word, it highlighted the words and reread the entire sentence. Following both correction procedures, the student was prompted to reread the sentence.

Results from this pilot study were encouraging. Each participant’s performance was measured in two ways. First, pre and post test scores were collected using a standardized reading test. These tests were administered by the school’s reading specialist and measured accuracy, comprehension, and reading rate as well as phonemic awareness, letter recognition and letter-sound correspondence. The results for these measures indicated that all of the students increased by two grade levels on the post tests. The second measure of student performance was a reading fluency measure. Fluency was determined by comparing the number of words read by each student during their first and last encounter with each word they read. These scores were averaged across all eight participants and there was an average length of 110 days between the first and last encounter with each word. An analysis of the data revealed that there was a 16% increase on accuracy and a 35% decrease on latency between words read.

A second pilot study was conducted in 2002. Its purpose was to increase the number of participants that interacted with the Reading Tutor (RT), as well as compare it’s effectiveness with sustained silent reading (SSR) (Mosotw et al., 2002). This was
accomplished by conducted a controlled randomized experiment in a two elementary schools. A total of 178 students from grades 1-4 were selected from the two schools. It was pointed out by the experimenters that both schools were considered “Blue Ribbon Schools of Excellence” with very high standardized test scores in reading. Students were assigned to one of two groups; either the SSR group or the RT group. Students in the SSR group engaged in silent reading in their respective classes once a day for 20 minutes and students in the RT group, engaged with the RT daily for one 20 minute sessions. The study took place over the course of 7 months.

In order to determine if there were any differences between the groups, a pre and post test was administered using the Woodcock Reading Mastery Test and the Comprehensive Test of Phonological Processing (CTOPP). Both of these standardized tests contained subtest that measured reading and phonemic awareness skills. While the results varied between grade levels, which were more statically significant results were found with the younger students (i.e. first- and second-graders), the overall outcomes indicated that the RT group performed better on virtually every subtest when compared to the SSR group. The researchers concluded that the RT was a more valuable way for students to spend their time compared to independent reading. Also it was concluded that the RT was most effective for early primary grade students and those students that scored lowest on the pre-test measures.

The results of these two pilot studies were very encouraging to the researchers and to the reading community in general. They raised the possibility of using advanced computer technology (i.e., speech recognition) to increase the reading skills of students.
The implication of this line of research for classroom use is fairly obvious. First, based on the results of these studies, it is clear that the Reading Tutor was an effective tool for teaching reading skills. The fact that it was implemented by a stand alone computer software program lends support to a growing need in schools, which is the lack of enough teachers to work with struggling students. With high teacher to student ratios, it is important to find ways to supplement the instruction that struggling students are receiving. The use of computer technology as a supplemental tool could provide some of the answers for these students.

Project LISTEN is an ambitious project and it has been well documented in the academic and scientific communities. The creators of this program have published numerous studies and have presented at dozens of conferences (see http://www.cs.cmu.edu/~listen/pubs.html). Although this line of research has made substantial strides in bringing computer technology into the realm of education, it has failed to make this technology available to a wider range of school/students. One of the major drawbacks of Project LISTEN and the RT is that there is not a commercially available version for purchase or use. After over 14 years of development and research there has not been a finished product that can be widely implemented to help students who are struggling with reading.

Read Naturally. A computer program that is commercially available is Read Naturally (RN). This company has specialized in reading and building fluency for over 17 years. Although Read Naturally developed several different programs designed to improve reading fluency, only the software edition will be presented in this section. The
computer software produced by Read Naturally provides students with an opportunity to systematically advance through grade level reading material at their own pace. Students are guided through a series of stories that are designed to build vocabulary, comprehension, and oral reading fluency. The program provides guided steps that assist students through each story and allows them to actively participate in practice readings until he/she achieves a targeted level that corresponds with their grade level. Once students reach the specified goal, they are ready to be tested by a teacher and must demonstrate reading at the designated level before moving to the next story.

One of the benefits of this program is that it includes the use of guided practice to help students progress through each story/level. One of the first steps in this program is the introduction of sight words that are embedded into the story for each level. Prior to the story being presented, the RN program provides a set of vocabulary words with definitions. These words and definitions are presented using an audio file to the users, as they are prompted to read along with the program. This allows students to learn some potentially difficult words prior to seeing them in connected text.

Following the practice of the vocabulary words in the RN program, the computer prompts students to start an automated timer and begin reading. This cold reading lasts for one minute and is used to establish a baseline for that particular story. It also allows students to mark unfamiliar or difficult words. At the end of the cold reading, the student’s errors are calculated by the software and graphed. This allows students to set a fluency goal for the story.
The next step for the program incorporates a modeling phase in which the computer reads the story to the students while they are prompted to follow along. In the RN program, students are prompted to listen to the story while the computer reads it several times. Students are encouraged to sub-vocalize along with the story to ensure that they are not just listening to the audio recording. The third step is to engage in practice reading. In the RN program students are prompted to read the story several times. They are instructed to use a timer and read for one minute, just like in the pre-reading step. They are required to repeat the story until the reach a pre-determined goal.

The final step is the reading checkout. This step is used to determine the oral reading fluency rate for each student. For the RN program, this step begins after the comprehension check and requires students to get a teacher. The teacher listens while students read the same passage they practiced. They are timed for one minute and scores are determined by subtracting the number of mistakes for the total numbers of words read in the minute. If the students make more than three errors they are prompted to reengage in practicing the story on their own.

Read Naturally appears to be a viable and effective program for teaching ORF to students; however the empirical research on this program has been limited. Hasbrouck, Ihnot, and Rogers (1999) conducted a study to test the effects of the Read Naturally program. These researchers reported on the results of the program when it was used with 214 second and third grade students in a Midwest, urban elementary school. The students used the RN program for an average of 32 weeks and progress monitoring data were collected to determine their reading levels.
The results indicated that the second grade students averaged increases of 1.68 words correct per minute (wcpm)/per week and 1.60 wcpm/week for third grade students. These gains exceeded what was deemed “typical” for second graders and “ambitious” for third graders. According to study published in 1993 (Fuchs, Fuchs, Hamlett, Walz, & Germann), typical second and third grade students should improve their reading fluency by 1.5 wcpm/week. A second result of this study revealed that the average oral reading fluency for each grade increased from below the 25% percentile to between the 25% and 50% percentile by the spring. That is, in the fall second grade students were averaging 17.9 wcpm but in the spring they were averaging 71.6 wcpm. This increased average was also true for third graders who were averaging 42 wcpm in the fall but 93 wcpm in the spring. Although these results are promising, this study did not use the software edition of RN so it is difficult to determine if the results would generalize across program types.

A more recent study conducted by Denton, Fletcher, Anthony, and Frances (2006) used the RN program for students with persistent reading difficulties. The participants in this study consisted of 27 students from grades 1 through 3. All of these students participated in a previous study on reading interventions and failed to make any progress in with their reading ability (Mathes, et al. 2001). An additional thirty-five more students, who did not participate in the previous study, were included in this study. These students were nominated by their teacher as “at-risk” for reading difficulties. All of the participants first engaged in 8 weeks of structured phonics instruction. That participated in a program named Phono-Graphix for two 50 minute sessions per day. Following that
phase of the study, the students participated in the RN program for one hour a day for an additional 8 weeks.

The results of this study again demonstrated that the RN program was successful in increasing reading skills of the participants. A battery of standardized tests was used to determine the effectiveness of both programs. In order to detect any increases in fluency, the experimenters used the WJ-III sub-test of Word Attack, Letter-Word Identification, and Spelling. All of the score on these post-tests were found to be statistically significant, indicating that the RN intervention was effective in increasing the reading ability of the participants. There was also a more direct measure of reading fluency using the Gray Oral Reading Test and the results of this measure was also found to be significant. Although all the results were significant this study only evaluated the non-software edition of the Read Naturally program.

**Summary of Programs and Implication for Use**

The programs described in this section and the research to support them reveals a number of different concerns about the use of computers of teaching reading fluency. As stated in the introduction of this paper, there is a need to supplement reading instruction for struggling student without increasing the work load of teachers. Many of the students that would benefit the most from effective, supplemental reading instruction attend schools that are overcrowded and/or under funded. Both of these conditions put a serious strain on the resources that are available to teachers when it comes to providing additional and individualized attention to struggling students. This is where an effective
and efficient computer assisted reading programs could be implemented to address these needs. Although this seems like a simple solution, it is much more complex.

Both programs that were outlined in this section present with a variety of strengths and weaknesses when it comes down to practical classroom implementation. Project LISTEN’s Reading Tutor takes advantage of sophisticated technology to eliminate the need for extensive teacher monitoring. The use of a stand alone program to provide immediate and unbiased feedback is surely a way to increase the abilities of individual students that struggle in reading. There is a large amount of empirically derived research studies to support the functionality and effectiveness to this program; however, it is not without its limitation. The most immediately apparent limitation is the inaccessibility of this program on a wide scale. It is unknown at this time why the RT is not being made available commercially, especially in light of all time and effort that has been invested. The second limitation may be a reason for the first. Although not expressly stated, it was alluded to in several of the research articles that the voice recognition software that makes the RT so appealing is not at a stage where it can operate with a high degree of fidelity. Specifically, the technology is not as accurate as a human in detecting and correcting all of the mistakes that a struggling reader might make. This gap in the technology may be a reason why the program is still undergoing the developmental process and is not available for use in everyday classroom situations.

The opposite is true of the Read Naturally program. It is available for purchase and being used in schools across the country. This program is beneficial as a supplemental program and has some research to support its effectiveness. The main
problem with the research is the fact that it was not conducted using the software edition of the program. This research used a version of the program that needed to be implemented by classroom personnel and therefore requires time allocation of the teacher. As discussed earlier, a benefit of computer technology is to allow the teacher more freedom to address other situations while supplemental instruction is occurring. In order for the RN program to be deemed effective, research should use the software edition to determine if the results of the non-software edition could be replicated. Unfortunately, even the software edition requires a fair amount of monitoring to ensure program integrity. Even though the user can engage in many of the program components on his/her own, the only way of ensuring that they are doing so properly is through direct monitoring. Even though both the RT and RN program provide positive advances in the area of CAI for reading, there is still plenty of work to be done before computer technology can be reliably used to teach reading fluency.

Conclusion

The development of instructional strategies to increase oral reading fluency and in-turn overall reading skills have been taken root in the research literature over the past two plus decades. Strategies based on repeated reading activities have been demonstrated to have substantial potential to improve of reading fluency of at-risk and struggling students. What is still unknown and largely unstudied is the actual use of these strategies in the daily schedule of classrooms across the country. With the research to practice gap still in existence it is likely that teachers have not been trained or exposed to implement these strategies.
A viable solution to the above stated problem is the use of computer technology as an instructional tool. The ability of computer assisted reading programs to implement instruction is garnering more attention in the research literature. The use of computer programs may help bridge the research to practice gap in several important ways. First, effective programs can be used as a supplemental instructional tool to provide an added layer of instruction of struggling students. Students who struggle need additional, explicit instruction to make adequate gains. Teachers may not always be able to provide this additional instruction because of other classroom demands. If students can learn how to use a computer it will not require the full attention of the teacher. Finally, computers can be highly motivating to students. Programs are typically designed to appeal to the audience for which they were created. With additional instruction and increased response opportunities provided by computer programs it is likely struggling students can make gains towards decreasing their risk for reading failure.
Chapter 3: Method

This chapter describes the methodological procedures of this study. These are the procedures that are implemented throughout the study and include the following components: participants and settings, materials, definition of the dependent variables, definitions of the independent variables, general procedures, inter-observer agreement, procedural integrity, and social validity.

Participants

The participants for this study were selected because they met the following characteristics. First, the participants attended elementary schools in an urban school district. They all resided in a Midwestern urban school district and were from low socioeconomic households. The participants ranged in age from 6 to 8 years old and were enrolled in a first-grade classroom. This study included 8 students in two separate schools (i.e., four students in each school). All of the students were recommended by either their classroom teachers or the school’s special education specialist because they exhibited evidence of reading risk. Specifically, these students were included because they (a) were exhibiting reading risk in their classroom, (b) tested in the at risk/some risk range for oral reading fluency but in the emerging range for word decoding, and (c) had parental permission for participation.
Selection Criterion

Each participant was administered two sub-tests from the Dynamic Indicators of Basic Early Literacy Skills (Good & Kaminski, 2007) prior to inclusion in this study. The two sub-tests were nonsense word fluency (NWF) and oral reading fluency (ORF). These tests were designed to check for two specific reading indicators. The NWF sub-test checked if the participants exhibited the ability to decode nonsense words and therefore were likely to possess the alphabetic principle and phonological recoding or blending. For inclusion in this study each participant needed to score in either the emerging (30-49) or established (50 or higher) range on this sub-test. The ORF sub-test checked if the participants exhibited the ability to accurately and fluently read connected text. Individuals who read fluently are less likely to spend effort on figuring words out and more likely to comprehend what they are reading. For inclusion in this study each participant needed to score either at risk (0-7 cwpm) or some risk (8-19 cwpm) on the ORF sub-test.

School 1 (Century Community School)

Marvin. Marvin was a 7-year-old African American male in the first grade. Marvin scored 36 (emerging) on the DIBELS NWF sub-test and 7 (at-risk) on the ORF sub-test. He was placed in Reading Mastery level II and received small group instruction five days per week for 45 minutes per day.

Clyde. Clyde was an African American male in the first grade. He was 6 years old during this study. Clyde scored 39 (emerging) on the DIBLES NWF sub-test and 9 (some
risk) on the ORF sub-test. He was placed in Reading Mastery level II and received 45 minutes of small group reading instruction five days per week.

Malik. Malik was an African American male in the first grade. He attended a charter school in a large Midwestern city. He was 7-years old during this study. Malik scored 39 (emerging) on the DIBLES NWF sub-test and 14 (some risk) on the ORF sub-test. He was placed in Reading Mastery level II and received small group instruction five day per week for 45 minutes per day.

Tyrone. Tyrone was a 7-year old, African American first grader at the time of this study. Tyrone scored 31 (emerging) on the DIBELS NWF sub-test and 9 (some risk) on the DIBLE ORF sub-test. He was placed in Reading Mastery level II and received 45 minutes of small group reading instruction, five days per week.

School 2 (C. J. West School))

Lance. Lance was a first-grade African American male. He was 8 years old at the time of this study. Lance scored 30 (emerging) on the DIBLES of NWF sub-test and 6 (at-risk) on the ORF sub-test. He was retained in the first grade the previous year. Lance was placed in Reading Mastery level II and attended a small group for intensive reading instruction for 45 minutes per day, five days a week.

Stevie. Stevie was a 6-year-old African American male in the first grade at the time of this study. Stevie scored 32 (emerging) on the DIBLES NWF sub-test and 7 (at risk) on the DIBLES ORF sub-test. He was placed in Reading Mastery level II where he received small group reading instruction for 45 minutes per day, five days per week.
Sheba. Sheba was a 6-year-old African American female in first grade at the time of this study. She scored 32 (emerging) on the DIBLES NFW sub-test and 16 (some risk) on the ORF sub-test. She was placed in Reading Mastery level II where she received small group reading instruction for 45 minutes per day, five days per week.

Ashley. Ashley was a 9-year-old African American first-grade female at the time of this study. Ashley was retained in the previous school year so she was repeating first grade; however she did not attend the same charter school during the previous year. She scored 42 (emerging) on the DIBLES NWF and 15 (some risk) on the DIBLES ORF sub-test. She was placed in Reading Mastery level II where she received small group reading instruction for 45 minutes per day, five days per week.

Teachers

Given that both schools used Reading Mastery as their reading curriculum, participants were split into small groups for reading instruction. These groups were formed based on the reading level and progress of each individual participant. For these reasons, two of the teachers were actually teaching assistants; however these assistants ran their own reading group. These assistants were the reading instructors for 3 of the 8 participants. Three of the group instructors were the actual classroom teachers. These teachers conducted Reading Mastery reading groups as part of their typical classroom curriculum. The teachers had three of the eight participants in their reading groups, although one of the participants (i.e., Ashley) was moved to a lower group about one quarter of the way through the study. The final individual to complete the survey was a reading
specialist for school 1 (Century). He conducted a group in which two of the participants were a part of.
<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Race</th>
<th>School</th>
<th>Grade</th>
<th>DIBELS Scores</th>
<th>NWF</th>
<th>ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marvin</td>
<td>7-5</td>
<td>AA</td>
<td>Century</td>
<td>1st</td>
<td>36</td>
<td>Emerging</td>
<td>7</td>
</tr>
<tr>
<td>Clyde</td>
<td>6-</td>
<td>AA</td>
<td>Century</td>
<td>1st</td>
<td>39</td>
<td>Emerging</td>
<td>9</td>
</tr>
<tr>
<td>Malik</td>
<td>7-1</td>
<td>AA</td>
<td>Century</td>
<td>1st</td>
<td>39</td>
<td>Emerging</td>
<td>14</td>
</tr>
<tr>
<td>Tyrone</td>
<td>6-10</td>
<td>AA</td>
<td>Century</td>
<td>1st</td>
<td>31</td>
<td>Emerging</td>
<td>9</td>
</tr>
<tr>
<td>Lance</td>
<td>8-0</td>
<td>AA</td>
<td>C. J. West</td>
<td>1st</td>
<td>30</td>
<td>Emerging</td>
<td>6</td>
</tr>
<tr>
<td>Stevie</td>
<td>6-4</td>
<td>AA</td>
<td>C. J. West</td>
<td>1st</td>
<td>32</td>
<td>Emerging</td>
<td>7</td>
</tr>
<tr>
<td>Sheba</td>
<td>6-3</td>
<td>AA</td>
<td>C. J. West</td>
<td>1st</td>
<td>32</td>
<td>Emerging</td>
<td>16</td>
</tr>
<tr>
<td>Ashley</td>
<td>8-0</td>
<td>AA</td>
<td>C. J. West</td>
<td>1st</td>
<td>42</td>
<td>Emerging</td>
<td>15</td>
</tr>
</tbody>
</table>

- Age: refers to the age of each participant in years and months at the time DIBLES sub-tests were administered (12/01/2008)
- Race: refers to the racial background of each participant; AA= African American
- School: refers to the school that each participant attended at the time of this study
- NWF: Nonsense word fluency
- ORF: Oral reading fluency

Table 3.1 Participant demographic information and DIBELS selection criteria score
Experimenter and Secondary Observer

The primary experimenter for this study was a doctoral candidate in Special Education in the School of Physical Activity and Educational Services at The Ohio State University. He earned his Bachelor’s of Arts in Psychology (Children and Youth) from Eastern Connecticut State University (Willimantic, CT) in 1999 and a Master’s of Science in Applied Behavior Analysis from Northeastern University (Boston, MA) in 2002. Following completion of the MS, the experimenter worked as a behavior analyst in a school for children with special needed in Windsor, CT. His primary responsibilities included writing behavior objectives for students’ IEPs, conducting functional behavior assessments, developing behavior intervention plans, participating in interdisciplinary team meetings, training school personnel, and implementing behavior management strategies. In 2006 the experimenter began his doctoral studies in special education at The Ohio State University.

The secondary observer was used to assess inter-observer agreement and procedural integrity. She was a second year doctoral student in the Special Education program at The Ohio State University. She completed her Bachelor’s of Art in Psychology in 2008 from OSU. She completed her Master’s of Education in Special Education in 2005 from University of Toledo. Following the completion of her M.Ed, she became a special education teacher in a charter school in the city of Toledo, OH for five years. During her years as a teacher she became proficient in implementing evidence based academic interventions for students with or at risk for disabilities. She was trained
as the second observer by the primary experimenter. This training took place in the graduate associate office on the OSU campus.

Settings

All teaching and testing sessions were conducted in a quiet area of a classroom or in a separate teaching area in the charter schools that the students attended. Two schools were identified as settings for this study.

The Century Community School

The Century Community School is a charter school located in a predominately African American section of the city. The residents are considered low to low-middle income with nearly 70% of the students eligible for free or reduced lunch. The school lists the ethnic make up of its student body as 96% African American and 4% other racial backgrounds (i.e. White, Latino, and Multi-racial). Their enrollment of 600 students (50% males) was spread over grades K – 8. With 28 teachers, their student-teacher ratio is 23.6:1. Thirteen percent of its students are considered to have a disability compared to 14% average for the state. Century used a Direct Instruction program Reading Mastery, as the core instructional program for reading.

All instructional and probe sessions were conducted in the office of the special education specialist. This provided a relatively secluded and quiet area for the students to complete each training sequence. The room contained a teacher’s desk, a desk-top computer, a lap-top computer, several chairs, and other educational material. There were relatively few distractions, although occasionally one of the school’s personnel would come into the office to retrieve a personal item.
**C. J. West Community School**

C. J. West Community School was an elementary charter school within the city of Columbus, OH. It is located within a predominately African American community and enrolled approximately 220 students. African Americans make up 97% of these students and 94% of the student body were considered economically disadvantaged. Approximately 7% of the students were labeled with a disability. Class sizes averaged 15 students per class, which made a low teacher to student ratio. C. J. West Community School also used Reading Mastery as the core instructional program for reading.

All instructional and probe sessions took place in an unoccupied classroom. This room was used by the special education teacher in the morning but was available in the afternoon. The room contained a teacher and several pupil desks, several chairs, two laptop computers, and various educational materials (e.g., books, dry-erase board, etc.). There were virtually no outside distractions during interventions sessions.

**Definition and Measurement of Dependent Variables**

There were two sets of dependent variables for this research study. The first set was a direct measure of the functional relation between the dependent and the independent variable. This set of dependent variables was taken from the Read Naturally stories (as treatment probes) and from the Aimsweb progress monitoring stories (generalization probes). The second set of dependent variables was an indirect measure of the relationship between these dependent variables and independent variable. This set of dependent variables was taken from the pre/post test scores on the WJ-III sub-tests (i.e.
passage comprehension, letter-word ID, and word attack) and the pre/post scores of the DIBELS sub-tests (i.e. ORF and NWF).

**Primary Dependent Variables**

The first dependent variable for this study was the number of correct words read during a one-minute reading. Each reading consisted of a connected text passage that participant practiced during the teaching procedures. These readings were presented on the computer screen at the end of each teaching sequence. Participants were required to meet a pre-set goal during the practice reading phase and to answer all of the questions during the comprehension quiz phase before they qualified to pass each story. The number of correct words read at the end of the one-minute timing was collected. In order for a word to be considered correct the participants needed to pronounce the word accurately within 3 seconds of the previous word read. A word was considered incorrect if the participant mispronounced the word or failed to read the word within 3 seconds of the previous word being read.

The second dependent variable for this study was the number of correct comprehension questions answered by the participants. Following the completion of the practice phase (once the participant reached their goal criterion for the specific story), a series of comprehension questions were presented on the computer screen. These questions were directly related to the specific story and be delivered by the computer. Comprehension questions consisted of four multiple choice questions with three choices and one short answer question.
The third dependent variable was the number of correct words read per minute on a generalization passage. This passage was taken from the first grade Aimsweb progress monitoring assessment for oral reading fluency. It was administered directly following each ORF and comprehension treatment probe. The data collected was the number of correct words read at the end of the one-minute timing. A word was counted as correct when participants pronounced the word accurately within 3 seconds of the previous word being read. A word was counted incorrect if the participant mispronounced the word or failed to read the word within 3 seconds of the previous word being read.

The fourth dependent variable was the number of words re-told on a one minute timing following the ORF treatment probe. After each participant completed the ORF treatment probe, they were asked to tell the experimenter everything they remembered about what they just read. A timer was set for one minute and started as soon as the first word was retold. A word was considered retold correctly if it met the following criteria: (a) it was in or directly related to the passage read, (b) if it part of a comprehensive statement, and (c) was not repeated. Words that were repeated outside of the context of a new statement were not counted. Non-words such as “um” were not counted.

The fifth dependent variable was the number of words re-told on a one minute timing following the generalization probes. After each participant completed the ORF generalization probe, they were asked to tell the experimenter everything they remembered about what they just read. A timer was set for one minute and started as soon as the first word was retold. A word was considered retold correctly if it met the following criteria: (a) it was in or directly related to the passage read, (b) if it part of a
comprehensive statement, and (c) was not repetitive (e.g., “the, the, the”). Words that were repeated outside of the context of a new statement were not counted. Non-words such as “um” were not counted.

Secondary Dependent Variables

The sixth dependent variable was the age and grade equivalence scores for each participant on the Woodcock Johnson standardized reading assessment sub-tests. The WJ-III sub-tests was used for this study and administered prior to the intervention beginning (pre-test) and immediately after all of the intervention sessions were complete (post-test). Pre-test and posttest were administered using two different forms (i.e., form A for pre-test and form B for posttest) in order to avoid practice effects.

The sub-tests used were letter/word identification, word attack, and passage comprehension. The experimenter administered each sub-test to the participants. The ceiling for all three sub-tests was six incorrect responses in a row. Once a student reached the ceiling a final score was calculated as total number of correct responses up to the ceiling for the sub-test. Standard and grade equivalent scores were derived from the raw scores. Additionally, classifications based on standard scores were reported for each participant on all the subtests.

The seventh dependent variable for this study was the pretest and posttest scores from the DIBLES sub-tests. The sub-test used was oral reading fluency. During the ORF sub-test the participants were given a first grade connected text passage and instructed to read it for one minute. The number of word read correctly was calculated at the end of the one minute timing. The total number of words per minute served as the dependent
variable. A word was considered incorrect if the participant pronounced the word wrong or did not read the word within 3 seconds of the previous word being read.

Definition and Measurement of Independent Variables

In this study students participated in a computerized reading program named *Read Naturally*. Each participant was trained on how to use this program and engaged with the computer approximately three to four times per week. The program consisted of the following sequence: key words, one minute cold reading, read along, practice reading, pass timing.

*Key words.* Prior to reading the practice passage, the program introduced the participant to the new key words for that passage. These words were selected from the story that the student was currently on. At the beginning of the activity the program presented each key word to the participant on the computer screen. The program read each word to the participant along with a definition. The program then prompted the participant to read the words silently to him/herself. They were instructed to click on each word to hear it again if they need additional help. This procedure continued until the participant was able to read all of the key words correctly without any prompting.

*One-minute cold reading.* This activity consisted of the computer displaying the current story on the screen. The program instructed the participant to read the story for one minute and click on any words that they did not know. At the end of one minute the participant was instructed to click on the last word that was read. From this information the computer displayed the number of words the participant read and displayed the “goal” for the story. The goal for phase one of the intervention was set at the benchmark level of
end of first grade ORF, which was 40 words correct per minute. The goal was reset for phase two and was based on individualized criteria for each participant. This criterion was determined by averaging the two highest reading rates during phase one. For example, if a student’s two highest reading rates were 45 and 52, the new goal for phase two would be set at 49 words per minute.

*Read Along.* Following the completion of the cold reading, the computer introduced the current story. The computer read the story to the participants as they were instructed to follow along. The program prompted the participant to read quietly along with the story. The sentences were highlighted in blue as it was being read by the computer. During this phase participants were required to listen/read along three consecutive times for each story.

*Practice Reading.* Following the read along step, the participants engaged in a practice reading activity. The computer instructed the participants to engage in reading the selected story. During this step the participants had an opportunity to reach their goal of reading to the set criterion within a one minute time period. The computer instructed the participants to click a clock icon at the top of the page (directly above the story) and begin reading. If a student did not know or had difficulty with any of the words in the story they could click on the word. The computer read the word to the participant and that word was counted as incorrect. At the end of one minute the bell rang and the computer instructed the participant to click on the last word that was read. The total number of words read correctly was automatically calculated by the computer and displayed in a text box to the right of the story. If the participant did not reach his/her
goal, he/she was instructed to practice the story again. Once the student reached his/her goal, the student was instructed to click the next button to proceed to the next activity.

*Comprehension Tests.* Once the participants reached their goal during the practice reading, they were introduced to the comprehension activity. The computer introduced the comprehension quiz on the screen. This quiz consisted of a series of multiple choice and short answer questions directly related to the story. The participants were instructed to answer all of the questions by either clicking the response they thought was correct or by dictating their answer to the experimenter while it was typed into the space provided. If the participants did not answer all of the questions correctly the computer represented the question until it was correctly answered. The final comprehension score was determined by using only the initial answer to each question.

*Pass Timing.* The final component of the software sequence was the pass timing. After the participants completed the comprehension quiz they were instructed to find a teacher to “pass them.” The experimenter served as the teacher for the pass timing and the person to enter the password for the student to access the pass timing. This timing was identical to the cold timing and the scores were used for treatment probe data.

**General Procedures**

**Experimental Design and Conditions**

A multiple probe across students design was used for this study. This design consisted of three tiers with two students in the first tier and one student in both the second and third tier. There were four students for each school and a total of eight students for the entire study. All three tiers began with baseline and entered into
treatment phases as the study progressed. Visual inspection of the data was used to determine when a tier will begin the intervention. Participants in tier one were first to receive the intervention once their intervention data were stable tier two was introduced to the intervention and likewise for tier three. For approximately every three treatment probes the tiers not in intervention received a baseline probe. This continued until all three tiers were receiving intervention. Both schools served as their own separate study. That is, treatment decisions were made for each school independently and did not depend on what was occurring with the other school’s participants.

*Baseline*

Following the pre-testing sessions, baseline was implemented. Baseline consisted of one minute timed readings of a selected connected text passage. Each participant was presented the story in a word document on the computer screen. They were instructed to read as many words as they could within a one minute time period. These reading passages were selected from the Read Naturally SE 1.0 level. The experimenters presented the participant with the passage and instructed him/her to begin reading when ready. The experimenter stated, “Here is a story for today, I want you to read as many words as you can and as fast and the best as you can. If you do not know a word I will tell it to you so you can keep reading. Try to remember what you read about because I am going to ask you to tell me about what you read at the end.” The timer was set for one minute and started once the participant began to read the first word of the passage or once 3 seconds had elapsed. If the participant failed to read the first word or any of the subsequent words within 3 seconds, the experimenter provided the correct pronunciation.
The participant was instructed to go to the next word. The word was scored as incorrect. At the end of the one-minute timing, the experimenter counted the total number of correct words and recorded this number on the data collection sheet.

Following the completion of the one-minute timing the participant was presented a series of comprehension questions. These questions were typed on a word document and presented on the computer screen. The experimenter stated to the participant, “I am going show you some questions about the story; I want you to read each questions and all of the choices. Once you think you know the best answer tell it to me and I will write it down.” The experimenter wrote the answers on the data sheet.

Both schools used Reading Mastery as their core reading curriculum. Reading Mastery is a scripted curriculum and the lessons emphasize phonemic awareness skills beginning with level I and advancing through level III. Students are placed on one of the levels as a result of a placement test. They receive approximately one 45 minute lesson per day in a small group format. Students can be moved to different groups depending on their progress or lack there of. Groupings do not depend upon the grade or age of the student but rather their performance level.

**Training**

Following the baseline phase, each participant was introduced to the intervention. Prior to implementation of the intervention, the participants were exposed to a training sequence that familiarized them with the computer program. This training consisted of instructing the participant through each component of one story. During this instruction, the experimenter supervised the participants as they navigated through one lesson on the
program. Prior to and following each activity, the experimenter provided explicit instructions and a review of the activity. The student was given the opportunity to ask questions. The participant needed to exhibit a thorough understanding of how to use the program before the training was complete. The experimenter used a training checklist to ensure that the participants were able to adequately engage with the software (see appendix A). A secondary observer independently scored the training sequence for over 30% of the participants.

Implementation of the Independent Variable: Phase I

Once the participants completed baseline and were trained on how to use the Read Naturally program, phase one of the intervention was implemented. During this phase participants were placed in front of the computer and instructed to click on the story they were currently working on. In phase one, all of the participants had a set goal level they needed to reach during the practice reading activity. This goal was to read 40 words correctly by the end of the one minute timing. This goal was selected because it is the benchmark for the end of first grade; meaning, by the end of first grade students are considered to be reading as benchmark if they can independently read 40 words of a novel connected text passage within one minute. During treatment sessions, participants were instructed to engage in all of the activities in the software program for the current instructional story. Once the participants meet their goal on the practice activity and completed the comprehension quiz, they were presented with the treatment probe, pass timing.
Implementation of the Independent Variable: Phase II

Phase two was identical to phase one with the exception of the goal for each participant. During this phase, the participants had their goals increased, based on their individual performance in phase one. The two highest treatment probes in phase one were averaged and used to set the new goal for phase two. For example, if a student’s two highest reading rates were 45 and 52, the new goal for phase two would be set at 49 words per minute. The experimenter and his advisor had weekly research meetings to evaluate the treatment data. Following the completion of approximately 3 generalization probes, it was determined that there was not an increase in ORF on these probes. This was when the decision was made to implement phase two of treatment.

Treatment Probes

Once the participants reached their goal in either of the phases for each story and they completed the reading comprehension test, they were given a treatment probe. These probes consisted of one-minute readings of the practiced passage and corresponded with the “pass timing” activity of the software program sequence. The participant was instructed to click start and begin reading the story displayed on the computer screen. The timer was started at the click. If the participant did not correctly pronounce a word or if 3 seconds elapsed without the participant reading the word, the experimenter provided the correct pronunciation to the participant and instructed him/her to go to the next word. The missed word was scored as incorrect. After the timer went off, the experimenter instructed the participant to click on the last word that was read and the program automatically calculated the number of correct words that were read by the participant.
Following the pass timing treatment probe, the participants were instructed to engage in a retell activity. The participants were instructed to tell the experimenter everything that was remembered about the story. A timer was set for one minute and the experimenter said, “go.” The participants were allowed to verbally state everything that was remembered about the story that was read. The experimenter counted every word that was directly related to the story as a word retold. Words such as “um” and repeats were not counted as a word retold.

**Generalization Probes**

Following the completion of three treatment probes, each participant was given a generalization probe. These probes consisted of a one-minute timed reading selected from the Aimsweb ORF progress monitoring passages and the word retell probe of that passage. These probes were identical to treatment probes with the exception of the use of the Aimsweb progress monitoring passages.

**Materials**

The materials used in this study consisted of the following: assessment materials (i.e. WJ-III and DIBLES), the Read Naturally software program, three computers (two lap-tops for Cupe and one lap-top and one desktop for Millennium), a timer, tangible reinforcers and teaching procedures materials. The assessment materials consisted of both standardized and informal instruments to determine the students’ reading levels, oral reading fluency, oral retell, and comprehension. The assessment instruments as described in the following section were employed prior to, during, and following the intervention.
**Read Naturally Software**

The Read Naturally (RN) software consisted of a computerized reading program that was uploaded on to each computer hard drive. The software corresponded with the reading level of the participants. All of the participants used the first grade (level 1.0) software for this study. The first grade level software program consists of 24 stories. Students that progress through all of the stories of level 1.0 are promoted to level 1.5, which also contained 24 stories.

Each story was approximately 60-90 words and contained two to five key words. All of the stories were on a CD-ROM disk. This disk needed to be inserted into the computer’s disk drive in order for the participants to be able to access the story. Each story’s difficulty level was evaluated using a readability scale. Appendix B provides an example of several stories from the RN 1.0 level software.

**Standardized Testing Materials**

*DIBELS* (Good & Kaminski, 2007). The DIBELS Oral Reading Fluency 1st-grade winter benchmark assessment was used to evaluate each student’s oral reading fluency (ORF). The ORF sub-test was a standardized, individually administered assessment of reading fluency using connected text and comprehension of what the students read. Students were assessed to determine if they qualified for inclusion in this study. In order to qualify they needed to score as either “at-risk” or “some risk.” The DIBELS Nonsense Word Fluency (NWF) 1st grade fall benchmark assessment was used to evaluate students’ ability to decode words adequately. This assessment was used as a screening tool. In order for students to qualify for inclusion they needed to score in either the emerging or
the established range. The DIBLES winter assessments were administered as screening measures prior to the implementation of the intervention. The DIBELS spring assessments were used following the completion of all intervention sessions.

*AIMSweb.* AIMSweb is an academic assessment based on fluency measure. It is used in three different ways to provide a complete picture of the performance level of students across grade level academic skills. These skills include reading, writing, and math.

The first way AIMSweb is used is to provide benchmark assessments of the specified skills. For example, student performance on reading fluency can be determined using the benchmark assessment for reading. The benchmark assessments are administered at the beginning, the middle, and the end of the school year. Benchmarks are given at the beginning of the year to determine which students are at-risk for academic difficulty (early identification). They are given in the middle of the year as a general progress monitoring tool. That is, mid-year benchmark assessments are used to determine if students are making the expected grade level gains. The end of the year benchmark is used to assess each student’s annual yearly progress and can be used as a system of accountability for educational decisions that are made.

The second way in which AIMSweb can be used is for strategic monitoring. These assessments are administered on a monthly basis and the information gained is used to make educational decisions. The data are used to track the effectiveness of any instructional changes. Strategic monitoring is used for students who score in a risk category during benchmark assessments. These students require some supplements to the
standard academic curriculum in order to achieve benchmark performances. The final way to use AIMSweb is as a progress monitoring tool. These assessments are administered on a much more frequent basis (i.e., weekly or bi-weekly) and used for students who need more intensive instruction. These students are likely to get either small group or individual instruction according to need. Progress monitoring data from the AIMSweb assessment are used to determine if the intensive instruction is effective.

For the purpose of this study, the AIMSweb, first grade ORF progress monitoring passages were used to test for generalization effects. These passages consisted of connected text and were administered as probes following the completion of every three treatment probes. Appendix C provides an example of several stories from the Aimsweb passages.

*Woodcock-Johnson Tests of Achievement – Third Edition (WJ-III ACH)* (Woodcock, McGrew, & Mather, 2001). The WJ-III ACH evaluates educational progress in curriculum areas such as reading, mathematics, and written language. A key feature of this assessment is that the norms are based on data collected from the same sample of subjects. Thus, a direct comparison can be made between and within subjects with a high degree of accuracy. In the area of reading ability, the standard battery evaluates letter-word identification and passage comprehension. The supplemental battery consists of word-attack and reading vocabulary tests. For this study, tests from both the standard and the extended battery were selected. From the standard battery the passage comprehension and letter-word identification sub-tested were administered. From the extended battery, the word-attack sub-test was administered.
Reliability data for each sub-test was reported as an indicator of the how reliable the scores are compared to population norms. These data are reported as median reliability coefficients ($r_{11}$) and median standard errors of measurement (SEM). Reliability coefficients and standard error of measurement were reported for the following sub-tests: passage comprehension $r_{11} = 0.88$ and SEM = 5.12; letter-word identification $r_{11} = 0.94$ and SEM = 3.81; word attack $r_{11} = 0.87$ and SEM = 5.36. The WJ-III ACH tests yield several scoring and interpretation possibilities such as comparison, qualitative and level of development. For the purpose of this study, age equivalent and grade equivalent scores, as well as standard scores were used because these scores provide the most direct information regarding level of reading development.

*Timer.* A digital kitchen timer was used to time all of the 1-minute timed readings and retell sessions. This timer was set for one minute at the beginning of the timed reading and the participants were instructed to read a pre-selected passage for one minute.

*Reinforcers.* A variety of tangible reinforcers (e.g. candy) were provided to the participants for completion of each treatment story. Every time a participant selected a reinforcer a note was made by the experimenter to assess for preference. This information was used to purchase desired reinforcers once the previous ones were gone.

*Computers*

*Laptop 1.* A Sony Vaio laptop computer was used for two participants at Century Community and two participants at C. J. West This computer used the Microsoft Vista operating system.
Laptop 2. A Dell laptop computer was used for two participants at C. J. West. This computer used the Microsoft XP operating system.

Desktop. A Dell desktop computer was used for two participants at Century Community. This computer used the Microsoft XP operating system.

Interobserver Agreement

A second observer independently recorded data for at least 33% of the baseline, treatment, and generalization probes. All the probes were recorded using an audio/video recorder and these recordings along with a copy of the passage were provided to the second observer. The second observer recorded the number of correct or incorrect words that were read within one minute by the participant. Responses were considered correct if the participant correctly pronounced a word within 3 seconds of the previous word being read. A response was considered incorrect if the experimenter provided the word to the participant because of mispronunciation and/or more than a 3-second pause between words. The second observer also recorded the number of comprehension questions the participant answered for each treatment probe and the number of words retold.

Interobserver agreement was calculated using the exact agreement method. In this method, agreement was scored for each interval in which observer 1 and observer 2 scored identical frequencies of behavior. For this study each dependent variable served as one frequency of a behavior (i.e., the number of words read correctly per session). Agreement is then calculated using the following formula: Agreement

\[
\text{Agreement} = \frac{\text{Frequency}}{\text{Agreement Frequency} + \text{Disagreement Frequency}} \times 100 = \___\%.
\]
Interobserver agreement was calculated separately for each story passage and then averaged for all of the passages across all of the participants.

Given that all IOA sessions were recorded and therefore reviewable, there was 100% agreement for all dependent variables in this study. Following the completion of the second observer recording the data, a session-by-session comparison was conducted. During this comparison, any disagreements were reviewed using the video recording to come to a consensus regarding the data. A final decision was made about whether or not the data point in question was an occurrence or non-occurrence.

**Procedural Integrity**

An independent observer collected procedural integrity data in at least 33% of baseline, treatment, and generalization sessions. Several types of procedural integrity data were collected during this study. The first type of data collected was used to assess the implementation of the independent variable. This assessment used a checklist (see appendix D) to ensure the Read Naturally software functioned properly throughout treatment sessions. The percentage of steps in which the program operated correctly was tallied and calculated across all sessions and all participants. The total percentage of correct steps ranged from 82% to 100% and averaged 97.6%.

The second type of procedural integrity involved the collection of probe data. The experimenter needed to provide specific directions to the participants during probes and also needed to engage in specific procedures during treatment sessions. These sessions were recorded and used to collect procedural integrity data. A second observer used the recordings to collect data using a checklist (see appendix E). The total percentage of steps
correctly implemented was averaged across all participants. The procedural integrity data ranged from 71% to 100% with an average of 89%.

The third and final type of procedural integrity data involved assessing how the participants used the software program. A checklist (see appendix F) was used to collect data on the number of steps each participant engaged in correctly. An independent observer completed this checklist for each participant in at least 33% of all treatment sessions. The total percentage of steps completed correctly averaged across all participants. The percentage ranged from 66% to 100% and averaged 85%.

Social Validity

Social validity measures were used to assess the participants’ and teachers’ satisfaction with the treatment procedures. Following the completion of the study the experimenter administered questionnaires to the participants. All eight of the participants were asked to a series of questions following the completion of the intervention. These questions were designed to solicit the specific opinion of the participants about how much they enjoyed the program and whether they felt it was effective in helping them become better readers. The questionnaire consisted of five “yes or no” questions with a follow up component for participants to elaborate on their answers.

Questionnaires were given to the teachers after the participants completed the study. This questionnaire contained a series of questions that were generated by the experimenter, using a modified Likert type scale. There were a total of six teachers for the participants in this study. A questionnaire was given to the instructors following the completion of the intervention. This questionnaire consisted of five multiple choice
questions that were designed to solicit the opinion of these instructors about the importance of fluency instruction and the effects of this study of the fluency of their students. There were three possible answers for each question; a little, somewhat, and a lot. There was an additional space for the instructor to provide comments. These questionnaires were handed to the teachers who were instructed to complete them at their earliest convenience. All of the questionnaires were completed outside of the presence of the experimenter and were collected at later date. Appendix G and H contain copies of student and teacher social validity questionnaires.
Chapter 4: Results

This chapter presents the results of this study according to the dependent variables. The primary dependent variables were (a) correct words per minute (CWPM) and word retell on treatment and generalization probes and (b) percentage of comprehension questions answered correctly on treatment probes. The secondary dependent variables were (a) pretest and posttest scores on the WJ-III and DIBELS subtests and (b) the results of the social validity questionnaires for both teachers and students. Each section is further broken down according to the school each participant attended.

Analysis of Primary Dependent Variables

This section presents the results obtained from implementing the Read Naturally program. There were five dependent variables for each participant: (a) correct words per minute on treatment probes, (b) total number of words “retold” on treatment probes, (c) correct words per minute on generalization probes, (d) total number of words retold on generalization probes, and (e) percentage of comprehension questions answered correctly on treatment probes. The use of the multiple-baseline probe design allowed for analysis across baseline and treatment conditions for each participant. The treatment condition consisted of two phases and data analysis across these phases is provided for each participant.
Effect Sizes

Additional information regarding effect sizes using Cohen’s d (Cohen, 1992) is provided for comparison of baseline, treatment, and generalization probe averages for CWPM. Effect sizes were calculated using the following formula:

\[ d = \frac{\text{mean}_1 - \text{mean}_2}{\sqrt{(SD_1^2 + SD_2^2)/2}} \]

The \( d \) statistic was calculated by subtracting the mean of group one (i.e., baseline) from the mean of group two (i.e., treatment) and dividing that by the pooled standard deviation of the two groups. In order to control for the different number of data points between baseline and treatment conditions (and therefore different \( N \)'s) the standard deviations were pooled prior to being used to calculate the effect sizes. Small effects sizes are considered to be 0.2 or lower, medium effect sizes are considered to be 0.5, and large effect sizes are considered to be 0.8 and higher (Cohen, 1988).

School 1 (Century Community School)

Figure 4.1 presents the multiple baseline data for Century Community School students. This figure presents treatment and generalization probe data for CWPM and IWPM across baseline and treatment conditions for each participant. Table 4.1 presents the averages for CWPM on baseline, treatment phase and generalization conditions as well as effect sizes. Effect sizes are presented across the following parameters: baseline average versus total treatment probe average, generalization baseline average versus total generalization treatment average, baseline average versus phase one treatment average,
baseline versus phase two treatment average, phase one treatment versus phase two treatment average, generalization baseline average versus generalization treatment phase one average, generalization baseline average versus generalization treatment phase two average, and generalization treatment phase one average versus generalization treatment phase two average. Figure 4.2 presents the multiple baseline data for re-tell data for Century Community School students. This figure presents baseline, treatment and generalization probe data for word retell. Figure 4.3 presents the multiple baseline data for comprehension questions answered for each student. These data are presented as percentages and are across baseline and treatment probes.

Marvin

The first participant averaged 12.6 CWPM during baseline. There was an immediate increase once the intervention was introduced. Throughout the treatment condition (i.e., phases one and two combined) Marvin averaged 49.9 CWPM. Visual inspection of the data shows that Marvin increased ORF to either at or above the benchmark level of 40 CWPM. There was a very large effect size when comparing baseline averages to total treatment condition averages. This effect size was 2.9 with a pooled standard deviation 10.7.

To further evaluate these data, the treatment condition was split into two phases. During the first phase Marvin’s responding increased to an average of 41.7 CWPM. This was an increase of 29.1 correct words. In treatment phase one Marvin’s responding immediately increased to benchmark levels and remained there throughout the first
treatment phase. Comparison of the baseline average to the average of the first treatment phase reveals a very large effect size of 2.28 (SD 12.7).

During treatment phase two Marvin’s responding increased to an average of 45.5, which was the goal set for this phase. Once phase two was implemented responding immediately increased and remained stable for the remainder of the intervention. When baseline average was compared to treatment phase two average, there was an even larger effect size of 2.45 (SD 13.4). The effect size comparing treatment one versus treatment two was 1.23 (SD 3.0).

The final comparison for the first dependent variable was the average CWPM for treatment phase one and treatment phase two. When these two averages were compared there was a large effect size of 1.23 (SD 3.0). Although this is considered a large effect, when compared to the other three effect sizes for the first dependent variable it is considerably lower.

The second dependent variable was the percentage of comprehension questions answered correctly. During baseline Marvin only answered 26.8% of the comprehension questions correctly. Once the intervention was implemented he increased this average to 79% for the entire intervention condition (i.e., both phase one and two). Further analysis of intervention phases reveal differences in responding between these phases. Marvin averaged 64.4% of comprehension questions correctly in phase one and increased to an average of 90% during phase two of the intervention.

The third dependent variable was CWPM on generalization passages. The results are reported as averages across phases and the effect sizes are also reported. During
baseline Marvin read 19 words per minute correctly on the generalization passages. During the intervention his CWPM on generalization probes increased to an average of 24.5. This increase of 5 correct words was considered to have a medium effect size of 0.77 (SD 7.2).

A finer analysis of generalization results for Marvin is presented as comparisons between phases. Generalization baseline average (19 CWPM) when compared to treatment phase one average (17 CWPM) produced an effect size of -0.74 (SD 2.7), suggesting the intervention did not have any effect on responding. Once the criterion was adjusted in phase two, Marvin’s responding increased to an average of 30.25 with an effect size of 1.97 (SD 5.7), which is considered to be a large effect. The final effect size comparison of generalization averages for treatment phase one and treatment phase two was 1.78 (SD 7.4).

Results for the fourth dependent variable, word retell on treatment probes, evidenced a similar increase from a baseline average of 6.3 words retold in a minute to a treatment condition average of 29.8 words retold. During the baseline condition Marvin maintained a stable and low rate of responding. There was a slight increase in level once the intervention was implemented. Marvin’s responding increase to an average of 30 words retold per minute in the first treatment phase. The data were variable during this phase but they remained higher than any of the baseline data points. Once phase two was implemented there were not any immediate changes in level and the average words retold remained basically unchanged at 29.75. The most salient difference is the stability of the data throughout the treatment phase.
The fifth dependent variable was the word retell on generalization probe. Marvin averaged 6 words retold per minute. After intervention was introduced he increased his retell to an average of 14.8 words per minute for the total treatment condition. During the first phase, Marvin averaged 9.6 words retold per minute. This average nearly doubled to 18.8 words in treatment phase two.

*Clyde*

Clyde averaged 16.3 CWPM during the baseline condition. There was an immediate increase in level once the intervention was introduced. During the entire treatment condition (i.e. phase one and two combined) Clyde averaged 61.9 CWPM. A visual inspection of figure 4.1, shows that Clyde increased responding above the benchmark level of 40 CWPM. There was a very large effect size of 2.5 (SD 18) when comparing baseline averages to total treatment condition average.

During the first treatment phase, Clyde’s CWPM increased to an average of 51.6. These data varied but remained above benchmark levels for the duration of this phase. The comparison of baseline and treatment phase one resulted in a large effect size of 2.14 (SD 16.4).

During treatment phase two Clyde’s responding increased to an average of 69.9 CWPM. This increase exceeded the goal of 63 set for this phase. The level of responding immediately increased in this phase and remained stable. Baseline average compared to the average for this phase resulted in a large effect size of 2.44 (SD 21.8). The final comparison for the first dependent variable was the average CWPM for treatment phase one and treatment phase two, which produced a large effect size of 1.69 (SD 10.6).
Although this is considered a large effect, when compared to the other three effect sizes for the first dependent variable it is considerably lower.

The second dependent variable was the percentage of comprehension questions answered correctly. Clyde answered 13.6% of the comprehension questions correctly during baseline. During intervention, this average increased to 69.5% for both intervention conditions. For the first intervention phase, Clyde answered 48.9% of comprehension questions correctly and 85% correctly during the second phase.

The third dependent variable was CWPM on generalization passages. The results are reported as averages across phases and the effect sizes are also reported. During baseline Clyde read 22 CWPM on generalization probes but this increased to an average of 31.7 CWPM during intervention. This increase of correct words had a large effect size of 0.88 (SD 10.9).

An analysis of generalization results for Clyde is presented as comparisons between phases. A comparison of the generalization baseline average of 22 CWPM and treatment phase one average (20 CWPM) produced a negative effect size of -0.62 (SD 3.2), suggesting the intervention did not have any effect on responding. During phase two, however, Clyde’s responding increased to an average of 40.5 CWPM with a large effect size of 2.1 (SD 8.7). The final comparison was between generalization averages for treatment phase one and treatment phase two. The effect size for this comparison was 1.83 (SD 11.2). These effect sizes suggest a greater impact of treatment phase two compared to phase one.
Results for the fourth dependent variable, word retell on treatment probes, demonstrated a similar increase from a baseline average of 1.6 words retold in a minute to a treatment condition average of 32 words. During the baseline condition Clyde barely cited anything that was read. Immediately upon intervention increased to an average of 27.4 words retold per minute. During phase two more gradual increases were noted for an average word retell per minute increased to 36 per minute. There was also much greater variability in responding during phase two.

The fifth dependent variable was the word retell on generalization probe. Clyde averaged 3 words retold during baseline. During intervention he increased his retell to an average of 16.4 words per minute for the combined treatment condition. During the first phase, Clyde averaged 8.6 words retold per minute. This average nearly tripled to 22 words in treatment phase two.

Malik

Malik averaged 15.3 CWPM during baseline. Data were stable throughout the baseline condition. There was an immediate increase in level during the first intervention treatment probe produced 54 CWPM. During the entire treatment condition (i.e. phase one and two combined) Malik averaged 69 CWPM, well above the 40 correct words needed to reach benchmark. There was a very large effect size of 2.8 (SD 19.4) for the compared baseline and treatment averages.

During the first treatment phase Malik’s responding increased to an average of 63.7 CWPM, which was above benchmark level and produced a very large effect size of 2.28 (SD 21.1).
Malik’s responding increased to an average of 74.3 CWPM during the second treatment phase, which was the goal for this phase. Once phase two was implemented there was an immediate increase in the level of responding. There was a slight increasing trend throughout phase two and baseline average compared to treatment phase two average yielded a large effect size of 2.34 (SD 25.1). The comparison of the average CWPM for treatment phases one and two produced a large effect size of 1.39 (SD 7.6), albeit lower than the other three effect sizes for this variable.

The second dependent variable was the percentage of comprehension questions answered correctly. During baseline, Malik answered one third of the comprehension questions correctly (33.3%). Once the intervention was implemented he increased this average to 86% for the entire intervention condition (i.e. both phase one and two). When the intervention phases were reviewed independently and compared, slight differences were noted in that, Malik averaged 80% of comprehension questions correctly in the first intervention and increased to 92% in the second treatment phase.

The third dependent variable was CWPM on generalization passages. The results are reported as averages across phases and the effect sizes are also reported. In the baseline condition Malik averaged 24 CWPM. Immediately upon intervention his responding on generalization probes increased to an average of 35 CWPM. This increase of 11 correct words was found to have a large effect size of 1.4 (SD 8.3).

A further analysis of generalization results for Malik is presented as comparisons between phases. The baseline average for generalization probes (24 CWPM) when compared to treatment phase one generalization average (28 CWPM) produced a large
effect size of 1.1 (SD 3.6). During phase two, Malik’s responding increased to an average of 42 CWPM, with a large effect size of 2.20 (SD 9.2) compared to baseline. The final comparison was between generalization averages for treatment phase one and treatment phase two had an effect size of 1.85 (SD 7.9).

The fourth dependent variable was the word retell on treatment probes. Malik averaged 9 words retold per minute during baseline but increased to an average of 32.3 words per minute for the total treatment condition. During the first phase, Malik averaged 32 words retold per minute, which continued at 32.6 words in treatment phase two.

Results for the fifth dependent variable, word retell on generalization probes, are presented next. During the baseline condition Malik’s responding demonstrated a slight increasing trend but averaged 12.5 words per minute. During intervention Malik’s responding increased to an average of 23.8 words retold per minute in the combined treatment condition. In the first treatment phase Malik averaged 21.3 words retold and during the second treatment phase be averaged 26.3 words retold. Visual inspection of the data revealed an increasing trend throughout the entire treatment condition.

**Tyrone**

The final participant in the first school (Century) was Tyrone. He averaged 19.8 CWPM during baseline of the first dependent variable. These data were variable throughout baseline but showed a decreasing trend prior to the intervention. There was an immediate and dramatic increase in level once the intervention was introduced. This increase was of substantial magnitude from the last baseline data point of 14 CWPM to the first intervention data point of 65 CWPM. During the entire treatment condition (i.e.,
phase one and two combined) Tyrone averaged 66.2 CWPM. Visual inspection of the data indicated that Tyrone increased CWPM well above the benchmark level of 40. These data were variable with no distinguishable trend throughout phase one. There was a very large effect size of 2.3 (19.7) when comparing baseline averages to total treatment condition averages.

Treatment phases were evaluated separately using averages and effect sizes. During the first phase Tyrone’s responding increased to an average of 61.1 CWPM. This was an increase of 46.4 correct words. In treatment phase one Tyrone’s responding immediately increased to above benchmark levels and remained there throughout treatment phase one, producing an effect size of 2.02 (SD 20.4) compared to baseline.

During treatment phase two Tyrone’s responding increased to an average of 70.8 CWPM, which nearly reached his phase criterion of 71 CWPM. Once phase two was implemented responding immediately increased and remained stable for the remainder of the intervention. When baseline average was compared to treatment phase two average, there was a large effect size of 2.08 (SD 24.4) Comparison of the averages for the two treatment phases resulted in a large effect size of 1.02 (SD 9.4), which is lower than the other effect sizes for this variable.

The second dependent variable was the percentage of comprehension questions answered correctly. Tyrone answered 50% of the comprehension questions correctly during baseline but rose to an average to 81% for the combined intervention condition (i.e. both phase one and two). There was a distinct difference between treatment phases.
During the first intervention phase, Tyrone averaged 71.1% of comprehension questions correctly, which increased to 90% during phase two of the intervention.

The third dependent variable was CWPM on generalization passages. The results are reported as averages across phases and the effect sizes are also reported. During baseline Tyrone read an average of 27.3 correct words per minute, which increased to an average of 36.6 in intervention. This increase of 9 correct words was found to have a large effect size of 1.6 (SD 5.6).

A more in-depth analysis of generalization results for Tyrone is presented as comparisons between phases. When generalization baseline average (27.3 CWPM) was compared to treatment phase one average (34.3 CWPM) an effect size of 1.37 (SD 5.1) was produced. Once the criterion was adjusted in phase two, Tyrone’s responding increased to an average of 38 CWPM with a large effect size of 1.74 (SD 6.1). The final comparison was between generalization averages for both treatment phases. The effect size for this comparison was 0.90 (SD 4.1), a large effect although much less than the effects for the other comparisons.

The fourth dependent variable was word retell on treatment probes. During baseline Tyrone’s level of responding varied and averaged 12 words per minute. During intervention responding increased to an average of 26.5 words per minute; for the first treatment phase his responding averaged 26.6 words retold per minute. Comparable levels of responding were found in treatment phase two. Phases differed mainly in that data were more stable in treatment two.
The fifth dependent variable was the word retell on generalization probe. Tyrone averaged 9.3 words retold per minute but nearly doubled his level of responding (18.1) during intervention. During the first phase, Tyrone averaged 17.6 words retold per minute and during phase two his average was slightly higher at 18.6 words per minute.
Figure 4.1 Treatment and generalization probe session data (CWPM) for school one participants
<table>
<thead>
<tr>
<th>Student</th>
<th>BL</th>
<th>I</th>
<th>ES  (SD)</th>
<th>BL</th>
<th>I</th>
<th>ES  (SD)</th>
<th>BL</th>
<th>BL</th>
<th>Tx1</th>
<th>BL</th>
<th>BL</th>
<th>Tx1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marvin</td>
<td>12.6</td>
<td>43.9</td>
<td>2.9 (10.7)</td>
<td>19</td>
<td>24.5</td>
<td>0.77 (7.2)</td>
<td>2.28 (12.7)</td>
<td>2.45 (13.4)</td>
<td>1.23 (3.0)</td>
<td>-0.74 (2.7)</td>
<td>1.97 (5.7)</td>
<td>1.78 (7.4)</td>
</tr>
<tr>
<td>Clyde</td>
<td>16.3</td>
<td>61.9</td>
<td>2.5 (18.0)</td>
<td>22</td>
<td>31.7</td>
<td>0.88 (10.9)</td>
<td>2.14 (16.4)</td>
<td>2.44 (21.8)</td>
<td>1.69 (10.6)</td>
<td>-0.62 (3.2)</td>
<td>2.1 (8.7)</td>
<td>1.83 (11.2)</td>
</tr>
<tr>
<td>Malik</td>
<td>15.3</td>
<td>69</td>
<td>2.8 (19.4)</td>
<td>24</td>
<td>35.3</td>
<td>1.4 (8.3)</td>
<td>2.28 (21.1)</td>
<td>2.34 (25.1)</td>
<td>1.39 (7.6)</td>
<td>1.10 (3.6)</td>
<td>2.02 (9.2)</td>
<td>1.85 (7.9)</td>
</tr>
<tr>
<td>Tyrone</td>
<td>19.8</td>
<td>66.2</td>
<td>2.3 (19.7)</td>
<td>27.3</td>
<td>36.3</td>
<td>1.6 (5.6)</td>
<td>2.02 (20.4)</td>
<td>2.08 (24.4)</td>
<td>1.02 (9.4)</td>
<td>1.37 (5.1)</td>
<td>1.74 (6.1)</td>
<td>0.90 (4.1)</td>
</tr>
</tbody>
</table>

Table 4.1 Baseline and intervention averages of CWPM on treatment and generalization probes for school one participants; effect sizes and standard deviations included.
Figure 4.2 Treatment and generalization probe session data (retell) for school one participants

```
Figure 4.3 Treatment and generalization probe session data (percentage correct)
School 2 (C. J. West Community School)

Figure 4.4 presents the multiple baseline data for C. J. West Community School students. This figure presents treatment and generalization probe data for CWPM and IWPM across baseline and treatment conditions for each participant. Table 4.2 presents the averages for CWPM during baseline, treatment phases, and generalization conditions as well as effect sizes. Effect sizes are presented according to the following parameters: baseline average versus total treatment probe average, generalization baseline average versus total generalization treatment average, baseline average versus phase one treatment average, baseline versus phase two treatment average, phase one treatment versus phase two treatment average, generalization baseline average versus generalization treatment phase one average, generalization baseline average versus generalization treatment phase two average, and generalization treatment phase one average versus generalization treatment phase two average. Figure 4.5 presents the multiple baseline, treatment, and generalization probe data for word retell. Figure 4.6 presents the multiple baseline data for comprehension questions, which are presented as percentages across baseline and treatment probes.

Lance

Lance averaged 12.7 CWPM during baseline of the first dependent variable. There was an immediate increase in CWPM once the intervention was introduced; however, this increase was not of a large magnitude, increasing only by seven words from the final baseline probe to the initial treatment probe. Throughout the combined treatment conditions Lance averaged 53.4 CWPM. Visual inspection of the data revealed
that Lance mostly responded at or above benchmark. There was a very large effect size of 2.5 (SD 15.9) when comparing baseline averages to total treatment condition averages.

A comparison of the two treatment conditions showed that during the first treatment phase Lance’s responding increased to an average of 47.3 CWPM. An increase of 34.6 words over baseline, Lance’s responding immediately reached benchmark levels and remained at this level, producing a large effect size of 2.10 (SD 16.4).

During treatment phase two Lance’s responding gradually increased to an average of 58 CWPM, which exceeded the goal of 55 CWPM set for this phase. Comparison to baseline resulted in a large effect size of 2.45 (SD 19.2). Differences between the two treatment phases also yielded a large effect size of 1.16 (SD 9.2), with higher levels of responding under the second treatment conditions.

The second dependent variable was the percentage of comprehension questions answered correctly. During baseline Lance only answered 26.6% of the comprehension questions correctly. Once the intervention was implemented he increased this average to 87.6% for the combined intervention condition. Lance answered an average of 84% and 90% of the comprehension questions correctly during the first and second treatment phases, respectively.

The third dependent variable was CWPM on generalization passages. The results are reported as averages across phases and effect sizes. During baseline Lance read 11 words per minute correctly. After the intervention was implemented his responding on generalization probes increased to an average of 21.9. This increase of 10.9 correct words
was considered to have a large effect size of 1.5 (SD 7.0). Compared to the effect size for the intervention phase of the main treatment data, this effect size was substantially less.

A finer analysis of generalization results for Lance is presented as comparisons between phases. The generalization baseline average (11 CWPM) compared to treatment phase one average (17.25 CWPM) produced a large effect size of 1.57 (SD 3.9). Once the criterion was adjusted in phase two, Lance’s responding increased to an average of 26.75 CWPM with a large effect size of 2.37 (SD 6.6). The comparison between generalization averages for treatment phases one and two produced a large effect size of 1.46 (SD 6.5).

Results for the fourth dependent variable, word retell on treatment probes, demonstrated an increase from a baseline average of 7.3 words retold in a minute to a treatment condition average of 29.9 words retold. During the baseline condition, Lance maintained a stable and low rate of responding. There was a slight increase in level once the intervention was implemented. Lance’s responding increased to an average of 30 words retold per minute in the first treatment phase. These data were variable in phase one but there was an increasing trend. Once phase two was implemented there was not a change in level and the average words retold remained basically unchanged at 29.9, but the level of responding was more stable.

The fifth dependent variable was the word retell on generalization probes. Lance averaged 11 words retold per minute during baseline. After intervention was introduced he increased his retell to an average of 15.8 words per minute for the combined treatment conditions. Lance averaged 12 and 18.75 words retold per minute, respectively, for treatment phases one and two.
Stevie averaged 10.6 CWPM during the baseline condition for the first dependent variable. There was an immediate increase in responding once the intervention was introduced. Stevie averaged 62 CWPM during the combined treatment conditions. A visual inspection of Figure 4.1 shows that Stevie’s responding increased to or above the benchmark level of 40 CWPM. There was a large effect size of 2.4 (SD 21.2) when comparing baseline averages to total treatment condition average.

When treatment phase one average is compared to baseline levels of responding it reveals that Stevie’s CWPM increased to benchmark levels. These data were somewhat variable but did not fall below baseline levels of responding at any point. A large effect size of 2.03 (SD 20) was yielded for this comparison.

During treatment phase two Stevie’s responding increased to an average of 70.1 CWPM, exceeding the specified goal of 66 CWPM. After a slight decrease, responding during this phase steadily increased. A large effect size of 2.36 (SD 25.2) resulted from baseline comparisons. A comparison of the CWPM for the two treatment phases yields a large effect size of 1.37 (SD 13.6).

The second dependent variable was the percentage of comprehension questions answered correctly. Stevie answered a third of the questions correctly (33.3%) during baseline, but this average increased to 77.1% for the combined intervention condition. Correct responding increased further across treatment phases with averages of 66% and 83.3% for phases one and two, respectively.
The third dependent variable was CWPM on generalization passages. The results are reported as averages across phases and the effect sizes are also reported. During baseline Stevie read an average of 8 words per minute correctly on generalization probes. During intervention his responding doubled to an average of 19.8 CWPM, resulting in a large effect size of 1.8 (SD 6.5).

A further analysis of generalization results for Stevie is presented as comparisons between phases. The generalization baseline average of 8 CWPM when compared to treatment phase one average of 13.6 CWPM produced a large effect size of 2.19 (SD 2.6). Stevie’s responding increased in phase two to an average of 24.5 CWPM with a large effect size of 2.42 (SD 6.8). A comparison between generalization averages for treatment phase one and treatment phase two resulted in an effect size of 1.94 (SD 5.6).

Results for the fourth dependent variable, word retell on treatment probes, offer a similar increase from a baseline average of 3.3 words retold in a minute to a treatment condition average of 26.8 words retold. There was an increasing but variable trend in the level of the responding throughout phase one of the intervention. Stevie’s responding increased to an average of 20.7 words retold per minute in the first treatment phase. After phase two was implemented responding further increased to 30 per minute. There was also much more stability in responding during phase two.

The fifth dependent variable was the word retell on generalization probe. Stevie averaged 3 words retold during baseline. After intervention was introduced he increased his retell to an average of 12.1 words per minute for the combined treatment condition.
During phase one Stevie averaged 9 words retold per minute. This average increased to 14.5 words in treatment phase two.

Sheba

The third participant was Sheba who averaged 20.3 CWPM during baseline and immediately increased to 54 CWPM when treatment was introduced. During the combined treatment conditions Sheba averaged 89.7 CWPM, which exceeded the 40 correct words needed to reach benchmark and produced a very large effect size of 2.5 (SD 27.7).

During the first treatment phase Sheba’s responding increased to an average of 78.4 CWPM. Although the data were variable, there was an increasing trend throughout the phase. Comparing baseline and treatment averages resulted in a large effect size of 2.09 (SD 27.7).

Sheba’s responding increased to an average of 100 CWPM during the second treatment phase, which was greater than the set goal of 97 CWPM. During phase two there was not an immediate change in the level of responding, but there was an increasing trend throughout phase two. A large effect size of 2.39 (SD 33.2) resulted from the comparison to baseline. A large effect size of 1.38 (SD 15.6) emerged when comparing the two treatment averages.

The second dependent variable was the percentage of comprehension questions answered correctly. During baseline Sheba answered just over one half of the comprehension questions correctly (53.3%). Once the intervention was implemented she increased this average to 88.5% for the entire intervention condition. During the first
intervention phase, Sheba answered an average of 84% of the comprehension questions correctly, and increased to 92.7% during phase two of the intervention.

The third dependent variable was CWPM on generalization passages. The results are reported as averages across phases and the effect sizes are also reported. In the baseline condition Sheba averaged 24.5 CWPM and increased to an average of 34.4 CWPM during intervention with a large effect size of 0.9 (SD 11.1).

A further analysis of generalization results for Sheba is presented as comparisons between phases. The baseline average for generalization probes (24.5 CWPM) when compared to treatment phase one generalization average (22.3 CWPM) produced a negative effect size of -0.49 (SD 4.4), which indicated that intervention did not have an effect on Sheba’s responding. After phase two was implemented, Sheba’s responding increased to an average of 43.4 CWPM. The effect size for this phase compared to baseline was 1.92 (SD 9.8), considered a large effect. The final comparison between generalization averages for treatment phase one and treatment phase two produced a large effect size of 1.66 (SD 12.7).

The fourth dependent variable was the word retell on treatment probes. Sheba averaged 18.6 words retold per minute during baseline. After intervention was introduced, she increased her retell to an average of 35.9 words per minute for the total treatment condition. During the first phase, Sheba averaged 35 words retold per minute, remained basically the same at 36.7 words in treatment phase two.

Results for the fifth dependent variable, word retell on generalization probes, are presented next. During the baseline condition Sheba averaged 12.5 words per minute.
There was a slight increase once the intervention was implemented to an average of 20.8 words retold per minute. In the first treatment phase Sheba averaged 15.3 words retold and during the second treatment phase be averaged 25 words retold. Visual inspection of the data revealed an increasing trend throughout phase two of the treatment.

*Ashley*

Ashley averaged 27.7 CWPM during baseline of the first dependent variable. These data showed a slightly increasing trend throughout baseline. There was an immediate and substantial increase in level once the intervention was introduced. During the entire treatment condition Ashley averaged 75.6 CWPM. Visual inspection of the data demonstrated that Ashley increased responding to well above the benchmark level of 40 CWPM. These data were fairly stable with the exception of two higher data points during phase one. There was a large effect size of 2.4 (SD 19.9) when comparing baseline averages to total treatment condition averages.

During the first treatment phase Ashley’s responding increased to an average of 70.4 CWPM. This was an increase of 42.7 correct words. Ashley’s responding immediately increased to above benchmark levels and remained there throughout treatment phase one. A large effect size of 2.06 (SD 20.6) was found when compared to baseline.

During treatment phase two Ashley’s responding increased to an average of 80.8 CWPM, which was slightly below the set criterion of 86 CWPM. After phase two was implemented responding immediately increased and continued on an increasing for the
remainder of the intervention. When baseline average was compared to treatment phase two average, there was a large effect size of 2.13 (SD 24.8).

The final comparison for the first dependent variable was the average CWPM for treatment phase one and treatment phase two. When these two averages were compared there was a large effect size of 1.09 (SD 9.4).

The second dependent variable was the percentage of comprehension questions answered correctly. Ashley answered 60% of the comprehension questions correctly during baseline. Once the intervention was implemented she increased this average to 93% for the combined intervention condition. There was practically no difference between treatment phases, as indicated in averages of 94% and 92% correct responses, respectively, for the first and second treatment phases.

The third dependent variable was CWPM on generalization passages. The results are reported as averages across phases and the effect sizes are also reported. During baseline Ashley read an average of 21.6 correct words per minute, which increased to an average of 42.6 during treatment, and produced a large effect size of 1.4 (SD 14.6).

When the generalization baseline average (21.6 CWPM) was compared to treatment phase one average (35.6 CWPM) an effect size of 1.79 (SD 7.7) was produced. Once the criterion was adjusted in phase two, Ashley’s responding increased to an average of 49.6 CWPM with a large effect size of 1.58 (SD 17). Comparing the generalization averages for both treatment phases yielded an effect size of 1.07 (SD 13).

The fourth dependent variable was word retell on treatment probes. During the baseline Ashley maintained a stable rate of responding and averaged 16 words per minute.
There was a small increase in level once the intervention was implemented and Ashley averaged 41.1 words per minute on treatment probes for the entire condition. During the first treatment phase her responding increased to an average of 38.8 words retold per minute. Responding increased during this phase and remained higher than baseline. A small increase to 43.7 words per minute with stable responding was observed in the second treatment phase.

The fifth dependent variable was the word retell on generalization probe. Ashley averaged 13 words retold per minute. After intervention was introduced she increased the retell average to 23.5 words per minute for the all of the generalization treatment probes. During the first phase, Ashley averaged 19 words retold per minute and increased in phase two to 28 words per minute.
Figure 4.4 Treatment and generalization probe session data (CWPM) for school two
<table>
<thead>
<tr>
<th>Student</th>
<th>BL</th>
<th>I</th>
<th>ES (SD)</th>
<th>BL</th>
<th>I</th>
<th>ES (SD)</th>
<th>Treatment Probes</th>
<th>Generalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lance</td>
<td>12.7</td>
<td>53.4</td>
<td>2.5</td>
<td>11</td>
<td>21.9</td>
<td>1.5</td>
<td>2.1 (16.4)</td>
<td>1.57 (3.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(15.9)</td>
<td></td>
<td></td>
<td></td>
<td>2.45 (19.2)</td>
<td>2.37 (6.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.16 (9.2)</td>
<td>1.46 (6.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.57 (3.9)</td>
<td>2.37 (6.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.46 (6.5)</td>
<td>2.37 (6.6)</td>
</tr>
<tr>
<td>Stevie</td>
<td>10.6</td>
<td>62.1</td>
<td>2.4</td>
<td>8</td>
<td>19.8</td>
<td>1.8</td>
<td>2.03 (20)</td>
<td>2.19 (2.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(21.2)</td>
<td></td>
<td></td>
<td></td>
<td>2.36 (25.2)</td>
<td>2.42 (6.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.37 (13.6)</td>
<td>1.94 (5.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.19 (2.6)</td>
<td>2.42 (6.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.94 (5.6)</td>
<td>2.42 (6.8)</td>
</tr>
<tr>
<td>Sheba</td>
<td>20.3</td>
<td>89.7</td>
<td>2.5</td>
<td>24.5</td>
<td>34.4</td>
<td>0.9</td>
<td>2.09 (27.7)</td>
<td>-0.49 (4.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(27.2)</td>
<td></td>
<td></td>
<td></td>
<td>2.39 (33.2)</td>
<td>1.92 (9.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.38 (15.6)</td>
<td>1.66 (12.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.49 (4.4)</td>
<td>1.92 (9.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.66 (12.7)</td>
<td>2.42 (6.8)</td>
</tr>
<tr>
<td>Ashley</td>
<td>27.8</td>
<td>75.6</td>
<td>2.4</td>
<td>21.6</td>
<td>42.6</td>
<td>1.4</td>
<td>2.06 (20.6)</td>
<td>1.79 (7.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(19.9)</td>
<td></td>
<td></td>
<td></td>
<td>2.13 (24.8)</td>
<td>1.58 (17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.09 (9.4)</td>
<td>1.07 (13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.79 (7.7)</td>
<td>1.58 (17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.07 (13)</td>
<td>2.42 (6.8)</td>
</tr>
</tbody>
</table>

Table 4.2 Baseline and intervention averages for school two participants; effect sizes and standard deviations included
Figure 4.5 Treatment and generalization probe session data (retell) for school two
Figure 4.6 Treatment and generalization probe session data (percentage correct) for school two participants.
Analysis of Secondary Dependent Variables

This section presents the results of the secondary dependent variables in the form of pretest and posttest scores and associated information. There were two different assessment tools that were used to obtain pretest and posttest scores. These assessments were the Woodcock Johnson Test of Achievement III (ref) and the Dynamic Indicator of Basic Early Literacy Skills (DIBLES). Both of these assessment tools have several sub-tests that were administered individually.

The WJ-III sub-tests included the following: Letter/word Identification (LWID), Word Attack (WA), and Passage Comprehension (PC). Raw scores for these sub-tests were used to calculate grade equivalent rankings, standard scores and associated classifications based on the standard scores. For the DIBLES the winter and spring benchmark assessments were used as pretest and posttest measures. Although the nonsense word fluency benchmark was used prior to intervention, this subtest was used only as a tool to assess if participants met selection criteria to be included in the study. The pretest and posttest measure used for analysis in this section was the ORF subtest. The scores on this test are reported in the form of correct words per minute and the risk category associated with each score.

School 1 (Century Community School)

Table 4.3 presents the scores for the WJ-III subtest for participants at Century School. These results are separated by pre-test and posttest scores for grade equivalences, standard scores, and classification for each sub-test. Table 4.4 presents winter and spring benchmark scores for the ORF subtest as well as risk status for each participant.
Marvin

Marvin was assessed on all three subtests of the WJ-III prior to and following the Read Naturally intervention. As seen in Table 4.3, pre and posttest scores for LWID resulted in a grade equivalence of 1.5 and 1.9, both yielding an average classification according to standard scores. For WA Marvin scored grade equivalents of 1.9 and 2.0 pre and post, respectively. The standard score classifications for these equivalents declined from high average to average. The final WJ-III subtest was passage comprehension where Marvin attained grade equivalent scores of 1.2 and 1.8, pre and post, resulting in average classifications for both standard scores.

Winter benchmark scores on the DIBLES were used to determine the pre-intervention reading rate and risk category for each participant. Marvin read 7 words correct in one minute on the winter benchmark assessment. He was classified as “at-risk” for reading failure. The spring post intervention yielded an ORF of 21 CWPM, reducing his risk status to “some risk.”

Clyde

Clyde was assessed on all three subtests of the WJ-III prior to and following the Read Naturally intervention. Pretest and posttest grade equivalent scores for LWID were 1.4 and 1.9 years, respectively. Both scores are associated with the classification of average. The second subtest, WA yielded grade equivalent score that were considered to be in the average category as well. Clyde posted scores of 1.8 (pretest) and 2.3 (posttest). The pretest grade equivalence score for Clyde was 1.8 and the standard range score was 105-111. The final WJ-III subtest was passage comprehension and Clyde’s pretest, grade
equivalent score was 1.5 years. There was no change in grade equivalent score; it was 1.5 years and both classifications were considered average.

Clyde read 9 words correct in one-minute on the DIBLES winter benchmark assessment. He was considered to be exhibiting “some-risk” for reading failure. Although, Clyde increased his ORF to 31 CWPM on the spring benchmark assessment; this 14 word gain did not affect his risk status; he remained at “some risk.”

Malik

Malik’s pretest and posttest scores for LWID were 1.8 years and 2.0 years, respectively. Both of these scores were considered to be within the average classification range. Malik’s WA pretest score was 2.1 year and was classified as high average. His posttest score decreased to 2.0 years and a classification of average. The finally on the passage comprehension subtest, Malik scored at 1.4 and 1.8 years (pretest and posttest) and a classification of average for both.

On the DIBLES winter benchmark Malik read 14 words correct in one-minute on the winter benchmark assessment. He was considered to be exhibiting “some-risk” for reading failure. Following the intervention he read 27 CWPM on the spring benchmark and remained at “some risk” for reading failure.

Tyrone

Tyrone’s pretest score for LWID resulted in a grade equivalence of 1.5 years. This score increased to 1.9 years on the posttest. Both scores were considered to be in the average range. On the WA subtest, Tyrone scored a grade equivalent of 1.9 and 2.0, respectively. These scores were associated with the classification of average. On passage
comprehension, Tyrone was able to score 1.1 years pretest and increase his score to 1.5 years on posttest. These classifications were average.

Tyrone was reading 9 words correct in one-minute on the DIBELS winter benchmark assessment and was considered to be at “some-risk” for reading failure. On the spring benchmark assessment, Tyrone increased his ORF to 25 CWPM. He gained 16 words but his risk status remained at “some risk.”
### Table 4.3 Woodcock Johnson III pretest and posttest scores for LWID, WA, and PC sub-tests for participants in school one

<table>
<thead>
<tr>
<th>Students</th>
<th>1GE</th>
<th>2SS</th>
<th>3Class</th>
<th>GE</th>
<th>SS</th>
<th>Class</th>
<th>1GE</th>
<th>2SS</th>
<th>Class</th>
<th>GE</th>
<th>SS</th>
<th>Class</th>
<th>GE</th>
<th>SS</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marvin</td>
<td>1.5</td>
<td>99-</td>
<td>Ave</td>
<td>1.9</td>
<td>98-</td>
<td>Ave</td>
<td>1.9</td>
<td>106-</td>
<td>Ave</td>
<td>2.0</td>
<td>101-</td>
<td>Ave</td>
<td>1.2</td>
<td>90-</td>
<td>Ave</td>
</tr>
<tr>
<td>Clyde</td>
<td>1.4</td>
<td>94-</td>
<td>Ave</td>
<td>1.9</td>
<td>100-</td>
<td>Ave</td>
<td>1.8</td>
<td>105-</td>
<td>Ave</td>
<td>2.3</td>
<td>106-</td>
<td>Ave</td>
<td>1.5</td>
<td>98-</td>
<td>Ave</td>
</tr>
<tr>
<td>Malik</td>
<td>1.8</td>
<td>104-</td>
<td>Ave</td>
<td>2.0</td>
<td>101-</td>
<td>Ave</td>
<td>2.1</td>
<td>109-</td>
<td>H. Ave</td>
<td>2.0</td>
<td>101-</td>
<td>Ave</td>
<td>1.4</td>
<td>96-</td>
<td>Ave</td>
</tr>
<tr>
<td>Tyrone</td>
<td>1.5</td>
<td>99-</td>
<td>Ave</td>
<td>1.9</td>
<td>100-</td>
<td>Ave</td>
<td>1.9</td>
<td>106-</td>
<td>H. Ave</td>
<td>2.0</td>
<td>101-</td>
<td>Ave</td>
<td>1.1</td>
<td>88-</td>
<td>Ave</td>
</tr>
</tbody>
</table>

1. GE = Grade Equivalent Scores  
2. SS = Standard Scores  
3. Class = Classification based on SS  
4. Ave = Average  
5. H. Ave = High Average  

Table: 4.3 Woodcock Johnson III pretest and posttest scores for LWID, WA, and PC sub-tests for participants in school one
## Table 4.4 Winter and spring DIBELS ORF benchmark scores and risk statuses for school one students

<table>
<thead>
<tr>
<th>Student</th>
<th>DIBELS Winter</th>
<th>DIBELS Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1&lt;sup&gt;st&lt;/sup&gt; Grade-Benchmarks</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; Grade-Benchmarks</td>
</tr>
<tr>
<td></td>
<td>ORF&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Risk Status</td>
</tr>
<tr>
<td>Marvin</td>
<td>7</td>
<td>At Risk</td>
</tr>
<tr>
<td>Clyde</td>
<td>9</td>
<td>Some Risk</td>
</tr>
<tr>
<td>Malik</td>
<td>14</td>
<td>Some Risk</td>
</tr>
<tr>
<td>Tyrone</td>
<td>9</td>
<td>Some Risk</td>
</tr>
</tbody>
</table>

<sup>1</sup> ORF = oral reading fluency
School 2 (C. J. West Community School)

Table 4.4 presents the scores for the WJ-III subtest for participants at C. J. West Community School. These results are separated by pre-test and posttest scores for grade equivalences, standard scores, and classification for each sub-test. Table 4.5 presents winter and spring benchmark scores for the ORF subtest as well as risk status for each participant.

Lance

Lance’s pretest and posttest grade equivalent scores for LWID were 1.5 and 1.9 years, respectively. Both were classified as average. The word attack subtest, the pretest score was 1.8 years and the posttest score was 2.0 years. This range was associated with the classification of average. The final WJ-III, passage comprehension Lance’s grade equivalent score on the pretest was 1.2 years on the posttest it was 1.8 years. Again, both of these scores were classified as average.

Lance read 6 words correct in one-minute on the DIBLES winter benchmark assessment. He was considered to be “at-risk” for reading failure. He increased his ORF to 21 CWPM on the spring benchmark and decreased his risk status to “some risk.”

Stevie

Pretest grade equivalent scores for Stevie’s LWID assessment was 1.6 years. Posttest, grade equivalence score for increased to 2.0 years. Both scores were associated with the average classification. For the word attack subtest, Stevie scored a grade equivalent of 2.1 and 2.0 year, respectively. This range was associated with the classification of average. On passage comprehension, Stevie’s pretest assessment yielded
a grade equivalent score of K.9 years, which was in the low average range. After the completion of the intervention he increased to a grade equivalence of 1.5 years a classification of average. On the DIBLES winter benchmark, Stevie was reading 7 WCPM. He was considered to be “at-risk” for reading failure. Stevie increased his ORF to 21 CWPM on the spring benchmark and decreased his risk status to “some risk.”

Sheba

Pretest and posttest scores for LWID for Sheba resulted in a grade equivalence of 1.3 years and 2.1, respectively. These scores were both in the average classification range. For word attack, Sheba’s pretest grade equivalence score was 2.1 and her posttest score was 1.8 years. These ranges were associated with the classification of average. Passage comprehension subtests yielded a grade equivalent score of 1.3 years and 1.8, pretest and posttest. Both scored had the classification of average. On the DIBLES winter benchmark Sheba read 16 CWPM on the winter benchmark assessment. She was considered to be exhibiting “some-risk” for reading failure. Her ORF increased to 40 CWPM on the spring benchmark and her risk status decreased to “low risk.”

Ashley

Ashley’s pretest and posttest scores for letter/word identification resulted in a grade equivalence of 1.7 and 1.9 year, respectively. These scores were considered to be average. Her scores on the word attack subtest were 1.9 (pre) and 2.0 years (post). The pretest score was classified as of high average but the posttest score was average. On the passage comprehension subtest, Ashley’s pretest and posttest were 1.5 and 2.2 years. Again the pretest score was considered to be high average and the posttest score was
average. Ashley’s winter benchmark score on the DIBLES ORF subtest was 15 CWPM. She was considered to be at “some-risk” for reading failure. Ashley increased her ORF to 49 CWPM on the spring benchmark and decreased her risk status to “low-risk.”
<table>
<thead>
<tr>
<th>Students</th>
<th>Letter-Word Identification</th>
<th></th>
<th>Word Attack</th>
<th></th>
<th>Passage Comprehension</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
<td>Pretest</td>
<td>Posttest</td>
<td>Pretest</td>
<td>Posttest</td>
</tr>
<tr>
<td></td>
<td>GE SS Class</td>
<td>GE SS Class</td>
<td>GE SS Class</td>
<td>GE SS Class</td>
<td>GE SS Class</td>
<td>GE SS Class</td>
</tr>
<tr>
<td>Lance</td>
<td>1.5 96-102</td>
<td>1.9 97-103</td>
<td>1.8 105-111</td>
<td>Ave</td>
<td>2.0 101-106</td>
<td>Ave</td>
</tr>
<tr>
<td>Stevie</td>
<td>1.6 101-105 Ave</td>
<td>2.0 101-106</td>
<td>2.1 N/A N/A</td>
<td>106 Ave</td>
<td>2.0 101-106</td>
<td>Ave</td>
</tr>
<tr>
<td>Sheba</td>
<td>1.3 92-99 Ave</td>
<td>2.1 102-107</td>
<td>2.1 105-111</td>
<td>Ave</td>
<td>1.8 102-107</td>
<td>Ave</td>
</tr>
<tr>
<td>Ashley</td>
<td>1.7 102-107 Ave</td>
<td>1.9 100-104</td>
<td>1.9 106-113</td>
<td>5H. Ave</td>
<td>1.9 98-105</td>
<td>Ave</td>
</tr>
</tbody>
</table>

1 GE = Grade Equivalent Scores  
2 SS = Standard Scores  
3 Class = Classification based on SS  
4 Ave = Average  
5 H. Ave = High Average  
6 L. Ave = Low Average

Table: 4.5 Woodcock Johnson III pretest and posttest scores for LWID, WA, and PC sub-tests for participants in school two
<table>
<thead>
<tr>
<th></th>
<th>DIBELS Winter</th>
<th></th>
<th>DIBELS Spring</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st Grade-Benchmarks</td>
<td></td>
<td>1st Grade-Benchmarks</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>ORF(^1) Risk Status</td>
<td>ORF Risk Status Words Gained</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lance</td>
<td>6 At Risk</td>
<td>21 Some Risk</td>
<td>+15</td>
<td></td>
</tr>
<tr>
<td>Stevie</td>
<td>7 At Risk</td>
<td>21 Some Risk</td>
<td>+14</td>
<td></td>
</tr>
<tr>
<td>Sheba</td>
<td>16 Some Risk</td>
<td>40 Low Risk</td>
<td>+24</td>
<td></td>
</tr>
<tr>
<td>Ashley</td>
<td>15 Some Risk</td>
<td>49 Low Risk</td>
<td>+34</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)ORF = oral reading fluency

Table 4.6 Winter and spring DIBELS ORF benchmark scores and risk statuses for school two students
Results of Social Validity Questionnaire

This section presents the results of the social validity questionnaires for both the teachers and the participants. These results are presented together for each subgroup (i.e. teachers and students).

Teachers

Table 4.7 displays the results of the questionnaire for each instructor. The first question was about the participants’ overall reading performance following the completion of the intervention. Five of the six instructors answered “a lot” to this question and one instructor answered “somewhat.” The second question asked if the instructor thought the participants became more fluent readers after the intervention was complete. Three instructors answered “a lot,” two answered “somewhat,” and one answered “a little.” The third question asked if the instructor would allow their student to participate in similar programs in the future. All but one of the instructors answered “a lot.” The final instructor answered “somewhat.” The fourth item questioned if the instructors believed if the participants enjoyed the intervention. All but one of the instructors answered “a lot” to this question and the last instructor answered “somewhat.” The final question asked the instructors how important reading interventions were for struggling readers. Again, five of the six instructors answered “a lot” and the sixth instructor answered “somewhat.”
### Social Validity Questionnaire Results – Instructors

<table>
<thead>
<tr>
<th>Question Type</th>
<th>A Little</th>
<th>Somewhat</th>
<th>A Lot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Performance</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2. Fluency</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3. Participant</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>4. Enjoyment</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>5. Importance</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 4.7 Social validity questionnaire results for the teachers of the participants (see appendix G for full questions)

Only two of the instructors completed the comment section of the questionnaire. The first instructor indicated that his student performed exceptionally well in his reading group once the intervention was started. His statement included a ringing endorsement of the intervention and use of the computer software. The other instructor indicated that her student’s performance began to decline towards the end of the year, but this was due to an increase in tardiness that resulted in him missing his reading group.

**Students**

Table 4.8 display the results of the student social validity questionnaire. The first question asked if the participants enjoyed using the computer for reading. All of the participants answered yes to this question. When asked what their favorite part of using
the computer was; answers ranged from actual hardware (e.g., the mouse) to the stories. In fact, five of the six participants said the stories were their favorite part. The second question asked if the participants thought they became better readers and all of the students indicated that they did. Answers on the open component of the second question were variable, but most of the participants made references to being able to read faster. The third and fourth question was about motivation. Participants were asked if they enjoyed working for prizes (i.e., candy) and they were also asked if they would have worked even if they were not able to earn prizes. Again all of the participants answered that they enjoyed working for prizes but would have worked using the program even without receiving candy for their performance. The open ended part of the fifth question was used to determine why the participants would want to use the program even if they were not working for prizes. Nearly all of the participants gave answers that referred to the fact that the computer helped them become faster readers. The final question asked the participants if they would like to continue to use the computer program for reading. They all responded by saying that they would like to continue. When asked why, many said things such as “it helps me read better” (Marvin), “it helps me get A’s” (Clyde), and “I do better in reading group” (Sheba).
### Social Validity Questionnaire Results - Participants

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Enjoyment</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>2. Effectiveness</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>3. Motivation/Prize</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>4. Work w/o prizes</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>5. Continue with the program</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.8 Social validity questionnaire results for the participants (see appendix H for full questions)
Chapter 5: Discussion

This study sought to determine the effectiveness of a computer software program designed to increase the reading fluency of users. The current study investigated effects of the Read Naturally program on the oral reading fluency and comprehension of the participants. Eight students, all of whom were in first grade, participated in this study. These students were taught how to use the computer program and then interacted with it three to four times per week, for a total of between 14 to 16 weeks. Students were instructed to sit at a laptop computer and engage with a selected story during each session. A session consisted of a sequence of activities, using the selected story, designed to improve reading fluency and comprehension. This program was to be used as a supplemental reading activity and designed to be used by students with minimal supervision.

This study used a multiple baseline probe design. Students were separated into three different tiers (i.e., groups) for the purpose of demonstrating experimental control. This was accomplished by introducing each intervention phase one tier at a time, while keeping the other tiers in the baseline condition. Following an analysis of the tier one responding, the intervention was then implemented in tier two while tier three remained in the baseline condition. After treatment effects were established for tier two, the intervention was implemented in the final tier. In order to ensure there were enough
stories for the entire intervention, a probe design was used. This design used separate probe sessions during the baseline condition. That meant, once tier one students were introduced to the treatment condition, students in tiers two and three received only one baseline probe for every three treatment probes given to tier one students.

In this study there were 4 students in two different schools, for a total of 8 students. Each school had 2 students in tier one, one student in tier two, and one student in tier three. Prior to and after baseline and intervention sessions all of the students were administered pretest and posttest measures. These measures consisted of the Woodcock-Johnson III subtest and the DIBELS oral reading fluency benchmark assessments. The WJ-III subtests included letter/word identification (LWID), word attack (WA), and passage comprehension (PC). During baseline and treatment sessions, data were collected on the following dependent variables: correct words per minute (CWPM) on treatment stories, incorrect words per minute (IWPM) on treatment stories, CWPM on generalization stories, IWPM on generalization stories, words retold per minute on treatment and generalization stories, and the percentage of comprehension questions answered correctly on treatment stories. Treatment probes were administered following the completion of every story sequence and generalization probes were given after every three treatment probes. There were a total of 24 treatment probes (across baseline and intervention sessions) and approximately 8 to 10 generalization probes for each student in this study.

This chapter discusses the results of this study in the context of the research questions that were presented in chapter one. Each question is answered using the
information obtained through data collect and analysis. Limitations, implications for classroom implementation, and future research directions are also discussed.

Research Question One

*What effect will the Read Naturally computer software program have on the oral reading fluency of urban, first-grade students at risk for reading failure?*

*Correct Words per Minute on Treatment Probes*

This first research question required a direct comparison of oral reading fluency scores in baseline and treatment conditions. Oral reading fluency was directly measured by recording each participant’s reading of a connected text passage for one minute and calculating the number of words read correctly. The number of total words read was subtracted by the number of incorrect words to get the number of correct words per minute (CWPM).

The overall results for all the participants indicated that the intervention was very successful in increasing the ORF over baseline. As noted in the results chapter, the average increase for participants in school one (Century) was 36 CWPM and a corresponding increase of 45 CWPM for participants at school two. The results for individual students in both schools evidenced similar gains.

Given the substantial amount of research on the success of repeated reading strategies to increase reading fluency, these results are not surprising (Dowhower, 1987; Rashotte & Torgesen, 1985; Therrien, 2004; Therrien & Kubina, 2006; Weinstein & Cooke, 1992). These results also support the use of a multiple component strategy to improve the ORF of struggling students. Martens and colleagues (2006) used a multiple
component fluency activity to improve the ORF of second and third graders. Two key components of this activity, were the use of listen passage preview and the error correction procedure. By providing participants a model of fluency (the passage preview) and error correction, the experimenters were able to increase the ORF of the participants.

Participants in the current study were also exposed to similar procedures. For example, participants were required to listen to each treatment story (read along activity) three times prior to practicing it on their own. Error correction came in the form of self evaluation. Participants were instructed to “click” on the words they did not know during the cold reading and practice activities. The program provided the correct pronunciation of the unknown word immediately and highlighted the word for further correction following the completion of the story. There were some differences between the correction procedure of the current study and the Martens et al. (2006) study. The current study depended on the participants’ ability to determine if a word was correct or incorrect. It is possible that the participants did not realize they were pronouncing words incorrectly and failed to have those words corrected. Procedural integrity data were collected if the participants clicked on unknown words during practice readings. The results of this check indicated that the participants clicked on the unknown words 68.4% of the time. In the Martens et al. study an experimenter was used to determine mistakes on pre-session readings. These words were used for a phrase drill correction procedure during the intervention sessions. Although the participants in the current study did not accurately identify all unknown words, they were still able to achieve significant gains in ORF following the implementation of the RN software program.
The current study extends previous research in the following ways. The first major difference between this study and the vast majority of the existing literature is the presentation format. Most fluency strategies need to be implemented by trained personnel (i.e., teachers, classroom assistants, etc.) in small group or individual formats. These factors make implementation less likely to occur due to classroom conditions such as lack of personnel and time constraints. By using existing technology, students can be fairly independent when engaging in this repeated reading strategy. Additionally, the participants in this study indicated on several occasions that they enjoyed using the computer to help them with their reading skills. There is a paucity of empirical research on the use of computer-based, repeated reading strategies and the results of this study begin to examine the potential benefits of its use.

The second difference between this study and the existing literature was the age/grade level of the participants. There are only a few studies (Hapstak & Tracey, 2007) that have investigated the effects of repeated reading strategies for first-graders. It may be assumed that first grade students have not had enough time to exhibit difficulties in reading fluency; however, the DIBELS begins to measure ORF on the first grade winter benchmark. Students who are considered to be “at-risk” should certainly receive interventions immediately. With early identification and intervention, it may be possible to reduce the risk status of very young readers. As evidenced in the current study, 5 of the participants were able to decrease their risk status by the end of the school year. These results, build support for the use of reading fluency strategies for early primary students.
Differences Between Phases

Another finding demonstrated by the results of this study was an increase in ORF following an increase in goal criteria for each participant. The treatment condition was split into two phases in an attempt to generate higher rates of responding on generalization passages. During the first phase all of the participants were required to read 40 CWPM on the practice passage activity before they were allowed to take the comprehension test and final timed checkout. In phase two, the goal criterion to qualify for the comprehension test and final reading checkout was individualized for each participant. The new goal was determined by the performance of the participant during the first phase. The two highest rates of responding (i.e., number of words read correctly) in phase one were averaged to set the new goal for phase two. For example, Lance’s two highest CWPM levels in the first phase were 55 and 55. The average of 55 CWPM was the new goal criterion for Lance in phase two of the intervention. For Sheba, the two highest session rates in the first phase were 99 and 95 CWPM. Her reset goal in the second phase was 97 CWPM. The new goal criterion was calculated and reset for each participant individually.

All participants made substantial progress when comparing baseline averages to the overall treatment averages; however, by increasing the goal criteria all the participants further increased their ORF when comparing phases. The combined phase one average for participants in school one, for example, was 55 CWPM. This average increased by ten correct words to 65 CWPM during the second phase.
Averages for individual participants were variable, but the increases closely matched or exceeded the increased goal criteria. For example, Marvin’s goal increased from 40 to 44 CWPM and he averaged 45.5 CWPM during the second phase of the intervention. On the higher end, Clyde’s goal criterion increased from 40 to 62 CWPM for phase two. He averaged 52 CWPM in phase one and 70 CWPM in phase two. This increase of 18 CWPM exceeded the fluency goal by six words.

Similar results were demonstrated for participants in school two. The combined averages in phase one for these participants was 62 CWPM. This average increased by 15 correct words to 77 CWPM in phase two. Also similar to the results in school one, there was a variable range of increases between participants that either matched or exceeded the goal criterion increase. Two participants (Lance and Ashley) both gained an average of 11 CWPM in the second phase of the intervention, whereas the two other participants gained averages of 19 and 22 CWPM.

Much of the research on the use of repeated reading strategies typically does not focus on manipulating the number of words read during training stories. The seminal article by Samuels (1979) required participants to read the same story a specific number of times prior to being timed for ORF. Many other researchers used this approach when developing repeated reading strategies. That is, they required participants to read the same story for a set number of times before reassessing their ORF (Ardoin, McCall & Klubnik, 2007; Begeny, Daly III, & Valleley, 2006; Rashotte & Torgesen, 1985). Other researchers set specific goals (i.e., words per minute) before the participants were timed for ORF or introduced to more difficult material (Dowhower, 1987). Another strategy
used in previous research was to increase the difficulty level of the material if participants met ORF goals (Martens et al., 2007; Yurick, Robinson, Cartledge, Lo, & Evans, 2006).

These methods differed from the current study in several ways. Similar to some of the previous research, the current study set a specific goal criterion for all the participants in the first phase (i.e., 40 CWPM). This goal was based on benchmark requirements for low-risk students at the end of first grade. The results demonstrated that all of the participants in this study met this benchmark on treatment probes; however, many of them failed to display the same progress on generalization passages. Therefore, the first difference was implemented by increasing the goal criterion for each participant based on his/her individual phase one performance. In turn, all of the participants were able to increase their ORF accordingly in phase two, with corresponding increases in ORF for generalization passages (see discussion for research question four).

Another difference in the procedures of the current study was the requirement of the participants to continue with the practice reading activity until they reach their goal criterion. As pointed out above, many of the research procedures in previous studies only required participants to reread stories a set number of times. The current study required participants to reread the practice story until their specific goal was reached. This requirement caused some of the participants to become frustrated following the initial increase in criterion; however, they were able to adjust quickly. For example, Sheba’s criterion increased from 40 CWPM in phase one to 97 CWPM in phase two. During the first phase she averaged 1.5 times on the practice reading before she reached her goal. Once the criterion was increased on the first practice session, it took her six timings to
reach the new goal. She voiced frustration by saying the new goal was “too hard” and “impossible” to reach. By the end of the second phase, she was reading the practice stories on average only twice before reaching this goal.

This observation was similar for all of the participants. When the increased goal criterion was implemented in phase two, there was an increase in the number of times a story needed to be practiced; however, the number of practice readings decreased as the phase progressed. In fact, the average number of practice readings was only slightly higher by the end of phase two than phase one.

The final major difference in the procedures of the current study and some previous research was the use of grade level material throughout the entire study. This study used stories that were reported to be on a first grade level. Other studies started off using stories that were on or below the grade level of the participants and typically ended using stories that were above the participants’ grade level. The difficulty level was increased based on the performance of the participants. The current study only manipulated the goal criterion while keeping the grade level material consistent. It is unclear which is more beneficial or which manipulation would produce higher rates of responding. One possible risk of increasing the difficulty of the material is a decrease in overall ORF. There are conflicting findings surrounding the increase in the difficulty of the material. In the Begney et al. (2006) study, the participants’ ORF showed a decreasing trend on all three treatment conditions. It was speculated that this was due to the increasing difficulty of the material throughout the study. Other studies reported mixed results when training material became increasingly more difficult (Martens et al.,
In the Martens study, several participants were classified as non-responders because they failed to make fluency gains on more difficult material, whereas other participants did make gains. Future research should focus on a direct assessment of both manipulations to determine if one is more effective.

Research Question Two

*What effect will the Read Naturally computer software program have on the percentage of comprehension questions answered correctly by the participants?*

**Percentage of Correct Comprehension Questions**

To assess comprehension, participants were asked a series of comprehension questions for each treatment story. There were five questions for each story with four multiple-choice questions and one short answer question. The comprehension quiz was the fifth activity in the program sequence and it came immediately following the practice reading activity. In order for participants to be eligible to take the comprehension quiz, they needed to first reach their CWPM goal on the practice reading.

The overall results for all students demonstrated that the Read Naturally software was very effective in increasing the percentage of questions answered correctly. The combined baseline average for all participants was 37% correct. This average increased to 83% for intervention sessions, which is an increase of 46%. Participants in both schools demonstrated similar patterns of responding across baseline and combined treatment conditions.

These results are similar to and support past research on the correlation between ORF and comprehension (Buchanan, 2006; Cates, Thomason, & Havey, 2007; Fuchs et
al., 2001; Huang, Nelson, & Nelson, 2008; Klauda & Guthrie, 2008; Therrien & Hughes, 2008). Dowhower (1987) used a repeated reading strategy with a group of second grade students to determine its effects on fluency and comprehension. Following the intervention it was found that participants increased both their fluency and comprehension on practiced and unpracticed passages. Although the current study only checked for comprehension of treatment session passages (i.e., stories read during the practice activity), the results indicate that when the participants’ ORF increased so did their comprehension scores.

More recently, Therrien and Hughes (2008) compared two different strategies to promote comprehension for struggling students. The first strategy was the use of tutor lead, repeated reading and the second strategy was called “question generation”. In this strategy, participants were prompted through a series of generic questions about the story. The results of this comparison demonstrated that the repeated reading strategy generated a more significant increase in comprehension scores than did the question-generation strategy. The authors speculated that participants’ comprehension scores improved more with the repeated reading because of the repeated exposure to the same words. This is in line with LaBerge and Samuel’s automaticity theory (1974) which states that the more an individual is exposed to the same words; the more likely it will be that he/she begins to automatically recognize them when reading and therefore devote more attention to understanding their meaning. This may also help to explain the results of the current research question. By the end of the combined treatment phases, participants were reading an increasingly high number of words. Most of the participants were completing
the treatment stories and some of them were finishing the stories well before the one minute timing was over. By increasing their ORF, the participants were expending much less effort on figuring out the words and therefore could put more effort on the meaning of the entire story.

Differences between Phases

Intervention was divided into two phases and the administration of the comprehension quiz was based on the number of words read correctly during the practice phase. Participants were required to read more words correctly in phase two before they took the comprehension quiz. This phase change had a direct effect on the percentage of comprehension questions answered correctly for the majority of the participants.

Participants in school one averaged 66% correct on the comprehension quiz. During phase two, this average increased to 89% correct. In school two, participants averaged 82% correct in phase one and 90% correct on phase two comprehension quiz questions. All of the participants increased their average percentage of correct questions from phase one to phase two, with the exception of Ashley. She was already answering correctly at a high percentage during phase one (94%) and answered slightly lower during phase two.

A possible reason for the increasing percentages in phase two is that participants were required to read more of the story during the practice activity. Most of the treatment stories contained between 60 and 90 words. The criterion to move to the comprehension quiz during the first phase was 40 CWPM. All of the participants could have met this requirement before completing the story. Analyzing the averages of the participants for
the first phase reveals many of them were not reading enough to complete the story. This is particularly true for students in tier one, who averaged between 42 and 52 words per minute. By increasing the number of words a participant was required to read, they were exposed to more of the story. There was also a slight increase in the number of times participants needed to practice stories before they reached their goal. It is likely that being exposed to more of the story (i.e., more words) participants were able to answer correctly more of the comprehension questions.

Similar results on the link between increases in ORF and comprehension were found by Yurick et al. (2006), where participants were exposed to both a sustained silent reading (SSR) and a repeated reading procedure. The results of these two procedures demonstrated that participants answered more comprehension questions correctly when they engaged in the repeated reading procedure over SSR. A key feature in this study was the use of a set goal criterion that participants needed to reach before they were administered the comprehension quiz. In the SSR condition there was no such requirement and no way to determine how much of the passage was actually being read by the participants. In the repeated reading activity, participants could not access the comprehension questions until they reached their goal. This requirement was in line with the requirement in the current study. Participants needed to keep reading the practice passage until their goal was reached. When the goal was increased during the second phase of the intervention, 7 out of 8 participants increased the percentage of comprehension questions answered correctly. The eighth participant’s responding only decreased by one percentage point. During the second phase all of the participants
increased their correct responding to 80% or over, with six increasing to over 90% correct. These results support the use of increased goal criteria to promote higher ORF and better comprehension.

Research Question Three

*What effect will the Read Naturally computer software program have on the oral re-tell of the participants?*

The third research question also involved a measurement of the participants’ comprehension. This was measured by calculating the number of words each participant was able to recall after finishing the one-minute timed reading checkout. Following each treatment and generalization oral reading fluency probe, the experimenter asked participants to verbally state everything they can remember about the passage they just read. This word retell assessment was identical to the DIBELS retell fluency (Good & Kaminski, 2002), which consists of a one-minute timed retell activity administered immediately following the ORF timed reading. These researchers indicate that students who are able to read at least 40 CWPM should be able to retell 25% of the number of words read to be considered on track for comprehension.

*Word Retell on Treatment Probes*

Participants in both schools were assessed on word retell fluency (WRF) for all treatment stories in baseline and intervention conditions. The results for these assessments indicated that all of the participants were able to increase the WRF over baseline levels. The average WRF during baseline sessions for all eight participants was 9 words per minute. This average increased to 32 words retold during treatment phases.
This increase of 23 words was representative of gains demonstrated by participants in both schools. Considering that individuals need to be able to retell at least 25% of what they read (according to DIBELS criteria), these participants exceeded this expectation. The combined average CWPM for all the participants was 58 correct words. In order to meet the 25% criterion, participants needed to retell at least 15 words per minute. They doubled this with an average of 32 words retold; approximately half of all words read.

Participants in school one averaged 7 words retold per minute during baseline sessions. This average increased to 30 words retold per minute during treatment condition probes. This increase of 23 words per minute exceeded the DIBLES criterion of 25% of ORF as comprehension norms. Participants in school one averaged 52 CWPM for ORF during treatment probes. These participants only needed to retell 13 words per minute in order to meet the 25% criterion. By obtaining an average of 30 words per minute participants retold nearly 60% of what was read during treatment stories. There were only slight differences between the two separate treatment phases. Participants averaged 29 words retold per minute in phase one and 31 words retold in phase two.

Participants in the second school exhibited the same pattern of responding. The combined average for participants in school two was 11 words per minute during baseline sessions. This average increased to 31 words per minute during treatment sessions. Comparing the average number of words retold to the average number of words read demonstrates that participants were retelling an average of nearly 50% of words read. When comparing differences between phase one and phase two, there was only a small
increase. Participants averaged 31 words retold in the first phase and 35 words in the second phase.

The results of this analysis add more support to the link between ORF and comprehension. The participants were able to retell much more of the story passages following the implementation of the RN program. With an average increase of 23 words once treatment phases were implemented and in comparison to recommendations set further by the DIBELS WRF sub-test, it can be determined that the intervention was effective for increasing word retell fluency of these participants. These results are similar to those found by Welsch (2007) in which participants increased their ORF and their word recall across both treatment and generalization passages. Not as clear are the variables that caused participants to retell double the number of words required to be considered on track for comprehension.

One possibility is the fact that specific activities during the intervention are favorable for the retell assessment. Participants were exposed to the story multiple times throughout intervention sessions. First, they are required to follow along as the computer reads the treatment story three times during the “read along” activity. Second, participants were required to practice the treatment story as many times as needed to reach their goal. Although participants were generally consistent in the number of times they needed to practice, the repeated reading aspect of the intervention differed from DIBELS ORF probes. DIBELS probes consisted of a single, one-minute timed reading, followed by the WRF assessment versus being repeatedly exposed to the story that needs to be retold. Research on the use of retell fluency in repeated reading interventions is
nonexistent and therefore cannot be used to support or refute these findings. Part of the next research question was used to address WRF on unpracticed reading passages.

Research Question Four

What effect will the Read Naturally computer software program have on the ORF and word retell of generalization passages (i.e., Aimsweb ORF progress monitoring measures)?

Participants performed very well on treatment story probes during both phases of the intervention. The results are in line with the accumulation of other empirical research on repeated reading strategies. Although these results are promising, it still needed to be determined if they would generalize to untrained passages. It is intuitive to think that fluency will improve on passages that are repeatedly read. Some research focused on testing repeated reading procedures for ORF on the same passages used for training (Hapstak & Tracey, 2007; Rashotte & Torgesen, 1985; Samuels, 1979). These studies revealed some valuable information about repeated reading procedures but were missing a critical analysis: how well the acquired skills transferred to unpracticed material. Researchers quickly turned their attention to this analysis and began to assess ORF on generalization material (Ardoin, McCall & Klubnik, 2007; Begeny, Daly III & Valleley, 2006; Dowhower, 1987; Therrin & Kubina, 2007; Martens, et al., 2007; Vandenberg, Boon, Fore, & Bender, 2008; Welsch, 2007). The current study also focused on the generalization of ORF on untrained passages.
Correct Words per Minute on Generalization Probes

Following the completion of probe sessions on every three training stories, participants were given an untrained, grade level passage to assess for ORF. The results of these generalization probes demonstrated that the intervention was effective for increasing the ORF on these passages; however, the majority of the participants needed an additional manipulation before treatment results were demonstrated. For this reason the treatment condition was split into two phases. In the second phase, participants’ goal criterion was increased for the treatment story, practice readings. This manipulation was used because it was easy to implement and it did not interfere with the integrity of the independent variable (i.e., the program sequence). The experimenter simply needed to change the goal criteria in the software program’s administration options.

During baseline conditions, participants in both schools averaged 16.5 CWPM on generalization probes (16 CWPM for Century and 17 CWPM for West). The average for both schools increased following the implementation of the intervention. The combined result for both schools was 31 CWPM and the separate results were 32 and 30 CWPM, respectively. Although these results indicated that the intervention was effective for increasing ORF of untrained passages, they are less impressive when considering the difference between baseline and treatment phase one. The combined average for phase one was 24 CWPM, which is an increase of 7 words and not close to the benchmark average of 40 CWPM needed to be considered at “low risk” for reading failure.
A closer examination of the data reveals that none of the participants were able to obtain the 40 CWPM benchmark during phase one of the treatment. Only two participants averaged over 30 CWPM and two others actually deceased in average CWPM during phase one when compared to baseline. Of the remaining 4 participants, 3 only made minimal gains of between 3 and 6 words over their baseline averages. Although the results of the effect size calculations suggest a large effect (i.e. over 0.8) for 5 of the 8 participants, these effect sizes are by far the lowest of all the other comparisons. In fact, this baseline versus treatment phase one comparison for generalization probes yielded the negative effect sizes for three participants. These results suggest that the 40 CWPM fluency goal was only minimally effective for promoting generalization on untrained passages.

Once the goal criteria were increased based on the individual participants’ performances in phase one, CWPM on generalization probes also increased significantly. Phase two averages for all the participants combined increased to 38 CWPM; 40 CWPM for school one and 37 CWPM for school two. Moreover, 4 of 8 participants reached an average that was over benchmark and another two increased their average to 30 or more correct words per minute. All of the participants had effect sizes for the baseline versus phase two that were considered very large; including the three who had negative effect sizes in phase one.

Few published studies have focused specifically on how to promote increased fluency on generalization material. Ardoin, Eckert, and Cole (2008) recognized this problem and conducted a study to investigate their theory that participants may exhibit
more generalization if they are exposed to multiple reading passages. These researchers compared a typical repeated reading strategy and the multiple exemplar strategy. The repeated reading strategy consisted of participants reading the same passage three times and the multiple exemplar strategy consisted of participants reading three different stories once each. Following the completion of each procedure, participants were given generalization passages to read. These generalization passages either contained high or medium word overlap. The results of this comparison demonstrated that participants’ ORF was significantly higher following the repeated reading condition. Generalization results were mixed with significantly higher scores on ORF passages with medium word overlap and no difference on passages with high word overlap.

Although the results from the Ardoin et al. (2008) study were not definitive on what promoted generalization, the researchers suggested that several variables of the study may have contributed. They identified the use of multiple reading examples and also cited the possibility of setting the fluency criterion of 100 CWPM on practice passages. The results of the current study also suggested that the participants’ generalization improved when they were required to read more words during the second phase of the intervention. It is possible that increasing the number of words read on practice passages had an effect on generalization. Participants in the current study were only in first grade and therefore may not have needed to read 100 CWPM, but by increasing the number of words based on their highest reading rate, generalization can be influenced. For the two participants whose increased fluency goal approached 100 words in phase two (i.e. Sheba’s goal was 97 CWPM and Ashley’s goal was 86 CWPM), the
highest ORF scores were obtained on generalization passages. Both of these girls increased their ORF average to above benchmark level of 40 CWPM following the increase in fluency goals.

The results of phase two are promising because they suggest at least two important considerations. First, they suggest that procedures to increase the ORF of struggling readers can use individualized criteria to promote generalization. Much of the past research on repeated reading and generalization fail to specifically identify the manipulation that resulted in increased responding on untrained passages. Furthermore, the most prevalent manipulation used in repeated reading strategies is to increase the difficulty of the material as opposed to the actually fluency rate. These results indicated that increasing the fluency rate may have a direct effect on the ORF of untrained passages.

Second, by increasing the ORF criteria on the practice passages but keeping the grade level difficulty static, participants may be able to still make gains in fluency without experiencing the frustration associated with more difficult material. There has not been much research comparing grade level material with increasingly difficult reading material on ORF. Much of the existing research either increased the difficulty level of the material (Begney & Silber, 2006; Staubitz, Cartledge, Yurick, & Lo, 2005; Valleley & Shriver, 2003; Yurick, Robinson, Cartledge, Lo, & Evans, 2006) or kept it at consistent difficulty level (Ardoin, Eckert, & Cole, 2008; Cates, Thomason, Havey, & McCormick, 2006; McComas, Wagner, Chaffin, Holton, McDonnell, & Monn, 2009; O’Connor, White, & Swanson, 2007; Therrien & Hughes, 2008). O’Connor and colleagues (2002) conducted a direct comparison of difficulty level and found that some participants
performed significantly better on ORF measures when they engaged in repeated reading strategies using material that was on their instructional reading level as opposed to their classroom reading level. Although the current study did not conduct an assessment to determine the reading level of each participant, pretest assessments and classroom reading groupings suggested that even though they exhibited some risk, they were able to read first-grade material. The results suggest that the manipulation of goal criterion alone was enough to increase ORF on both treatment and generalization probes. Hence, there was not a need to increase the difficulty level of the material. By the end of the study, 5 of the 8 participants reached benchmark levels on generalization passages and the data for the remaining students were accelerating.

**Words Retold on Generalization Probes**

One way to assess the participants’ retell comprehension that is similar to the DIBELS WRF sub-test was to use generalization passages. These passages were taken from the Aimsweb assessment, which is a similar curriculum based measure as the DIBELS. Following each generalization timed reading for ORF, participants were asked to recall “as many things” as they could about their passage. Results for these generalization probes were consistent with the finding on the treatment probes.

During baseline probes, the combined retell average for all the participants was nine words per minute. After the intervention was introduced this average increased to 18 words per minute for both intervention phases. When compared to the average number of correct words per minute for both schools during ORF generalization probes (31 CWPM), participants were retelling over 58% of the total words read. A school by school
comparison demonstrates nearly identical results. Participants in school one read 32 CWPM on generalization probes and retold 18 words during the WRF assessment. This is over 56% of all words read. School two participants read 30 CWPM on generalization probes and retold 18 words (60%) during retell probes. Both schools retold well over the 25% standard set by the DIBELS.

When conducting a comparison across treatment phases, it can be determined that were was an increase in word retell fluency after phase two goal criteria was increased for each participant. Total retell average for treatment phase one sessions was 14 words and this average increased to 22 words for phase two sessions. In terms of percentage of words retold, participants read an average of 24 CWPM on generalization probes during phase one and retold and average of over 58%, which was nearly identical to total treatment percentages. During the second phase of the treatment, participants’ average of CWPM on generalization probes increased to 38 words. The average of 22 words retold was a bit under 58% of the words read.

These combined results lend strong support to the positions that the participants in this study increased their reading comprehension as a result of the RN intervention and treatment manipulation. Although these results are convincing when held up to the criteria set by the DIBELS retell fluency activity, there is the question of total number of words read. The DIBELS criteria are based on students reading at least 40 CWPM to be assessed on WRF. Combined treatment averages were slightly lower than this criteria at 31 CWPM (total), 24 CWPM (phase 1), and 38 (phases 2). It may be useful to compare treatment averages of CWPM and WRF for participants who did meet the 40 word
criterion. Four participants (two in each school) met the benchmark criterion during the second treatment phase. These participants averaged 41, 43, 44, and 50 CWPM on generalization probes. The retell fluency scores for these participants were 26, 22, 25, and 28 words. These numbers represent percentages of 63%, 51%, 57%, and 56%, respectively and provide convincing evidence of strong WRF on the generalization passages, which support the link between increased ORF and increased comprehension (Welsch, 2007).

Research Question Five

What effect will the intervention have on the reading scores of the three Woodcock-Johnson reading sub-tests as administered pre and post the intervention?

The use of standardized tests helps to draw comparisons between the participants of this study and their grade level peers. Three sub-tests from the WJ-III were used in this study: letter/word identification, word attack, and passage comprehension. Each participant was administered these sub-tests prior to and following the intervention. Each participant’s raw scores were used to generate grade equivalent and standard scores, as well as classifications.

Letter/word Identification (LWID)

The first subtest was used to determine how many letters and words participants were able to read independently. The results indicated that all of the participants performed within the typical range for their grade level. Most of the participants had grade equivalent and standard scores that indicated normal performance at the pretest. Grade level scores were between 1.3 and 1.7. Following the intervention participants
remained either at grade level (e.g. 1.9) or slightly above grade level at 2.0 or 2.1. Their standard scores and classifications were consistent with the grade level peer norms and were considered to be in the average range of performance.

**Word Attack**

The second sub-test was used to determine the decoding skills of each participant. The results indicated that 4 of 7 participants performed above their grade level on pretest administration. Three other participants performed on grade level and the final participant’s performance could not be determined due to his age at the time of the pretest. Participants performed between grade level 1.8 and 2.1; these performances fell within the average to high average category when the standard scores were compared to peer norms. During the posttest administration it was determined that the participants’ performances regressed somewhat on word attack. Even though all of the participants maintained their “on grade level” performance, the participants who scored in the “high average” range on the pretest only performed in the “average” range on posttests.

**Passage Comprehension**

The third subtest was designed to determine how well participants would perform on comprehension measures. The results of the pretest indicated that all of the participants were performing in the average range with the exception of two. One participant performed in the low average range (K.9) and the other in the high average range (1.5) at the time of the pretest. Following the intervention, all of the participants scored in the average range for passage comprehension.
Summary of Findings

The results of the WJ-III subtests do not provide much additional evidence about the effectiveness of the intervention. Overall, the participants were at or around the average range of performance on all three sub-tests during both pretest and posttest. It is unclear as to why these participants did not make more significant gains on the standardized test; however, the vast majority of the participants’ scores did improve on all three posttest measures. It just happens to be that the increases were consistent with grade level norms. That is, participants ended the school year in the range of normal performance for these subtest measures. It should be noted that none of the sub-tests were measures of reading fluency. It is possible that the participants would have made more substantial gains on this type of subtest considering that the intervention specifically targeted oral reading fluency.

Research Question Six

What effect will the intervention have on the pretest and posttest scores on the DIBLES ORF benchmark assessment?

Unlike the WJ-III sub-tests, the DIBELS ORF benchmark assessment was used as a direct assessment of pretest and posttest oral reading fluency. Participants were administered the DIBELS subtest prior to the intervention as part of the selection criteria. To be considered for inclusion in this study, participants needed to fall either in the “at risk” or “some risk” category based on winter benchmark scores. Five of the participants selected scored in the “some risk” category; however, two of these participants were on
the lower end of the “some risk” criteria. Clyde and Tyrone both obtained an ORF score of 9 CWPM and the “at risk” category range is between 0 and 8 CWPM, meaning that they missed being considered “at risk” by one word. The three remaining participants scored in the “at risk” range on the winter benchmark.

The spring benchmark was the final assessment administered to all of the participants at the end of this study. The results of this assessment demonstrated that all of the participants made improvements in their ORF. Five out of eight participants were able to decrease their risk category and none of the participants were considered to be “at-risk” on the spring benchmark. Sheba and Ashley made the most improvement in terms of lowering their risk; both started in the “some risk” category and were able to improve their ORF to be considered at “low risk” in the spring. Lance, Stevie, and Marvin were considered to be “at risk” when the intervention started but were able to improve their performance and decrease their risk status to “some risk” after the intervention was completed. Finally, Clyde, Malik, and Tyrone were all considered to be exhibiting “some risk” prior to and following the intervention. Although they did not decrease their risk status, they all made substantial gains over the course of the treatment. This is particularly true for Clyde and Tyrone because as noted earlier, these two participants were only one word over what was considered to be “at risk” on the winter benchmark.
Research Question Seven

How will teachers rate the effects of the intervention on the participants ORF and comprehension?

There were a total of six teachers who completed the social validity questionnaire following the completion of the study. This questionnaire consisted of five multiple choice questions designed to solicit the opinion of the teachers about the effects of the intervention. There was also a place on the bottom of the questionnaire for each teacher to make comments regarding the intervention’s effect on the ORF of their students.

Each question was designed to assess the opinion of a specific aspect of the intervention. The most important questions were the ones pertaining to students’ overall reading performance and increases in reading fluency. For the first question about overall reading performance, 5 of the 6 teachers indicated that their students improved their reading following the completion of the intervention. Question two asked if the students became more fluent readers as a result of the intervention. These results were a bit more mixed, with 3 teachers indicating that their students improved “a lot,” 2 “somewhat,” and 1 “a little.” For the remaining three questions, all of the teachers except one agreed that their students enjoyed the intervention, it was important, and they would allow future students to participate in similar interventions.

The teachers who believed that the intervention was only somewhat effective may have thought so for different reasons. At least one of the teachers, who gave the outcome a lower mark, made few inquiries throughout the program and expressed little interest in learning about the intervention. When asked by the experimenter how she thought her
student was progressing throughout the year, she typically indicated that she was not sure. In contrast, the other teachers were very willing to offer information about their students’ progress in the class reading groups and regularly asked questions about the intervention. A possible reason for the difference of the teachers in interest level and overall attitude towards the study may have been related to professional training. All of the personnel in school two (C. J. West) were certified teachers, whereas the personnel in school one (Century) consisted of one reading specialist and one paraprofessional. The individual who seemed the most indifferent about the study happened to be the paraprofessional and the individuals who seemed the most interested were the certified teachers.

Research Question Eight

*Will the participants respond favorably to the Read Naturally program, as measured by a questionnaire administered at end of the study?*

Each participant was asked a series of questions after all intervention and posttest assessments were complete. These questions were designed to solicit information about what each participant thought about the computer program. Questions were asked about the effectiveness, enjoyment, motivation, and future participation in similar interventions. All eight participants gave the highest rank for all five social validity questions. Of particular interest was the question regarding motivation; participants were asked if they would still be willing to work on the computer even if they could not earn prizes. All of them indicated they would still like to work because they enjoyed using the computer. When asked if they wanted to continue to use the program for reading, they all said they
would like that “a lot”. When asked why, many indicated because it helped them to be faster readers.

Limitations

Although the results of this study are very encouraging in terms of increases of both ORF and comprehension on treatment and generalization passages, there are several limitations that should be discussed further.

*Failure to Implement in Natural Classroom Environment*

The combined results of this study support the viability of the Read Naturally software program for struggling students. All of the participants made substantial gains in ORF and comprehension on treatment and generalization stories. Although these results are promising and the intervention appears to be effective, they were obtained under clinical conditions. That is, participants were pulled out of their typical classroom and placed in a very quite environment. All sessions in the respective schools were conducted in empty classrooms, free from most distractions. Furthermore, the experimenter engaged in continuous monitoring of the teaching sessions.

This artificial learning environment is not typical of the daily conditions in many classrooms across the country. Moreover, many classrooms have very high student to teacher ratios that are not conducive to individual instruction or the close supervision required to ensure program integrity. These facts alone place a limitation on how far these results might generalize to more typical classroom conditions. It is unclear if these same results could be obtained under conditions in which a teacher instructs students to
use the Read Naturally software with minimal or no supervision. It is likely that less robust results would be obtained and should be a research interest for future study.

*Lack of Comprehension Questions for Generalization Passages*

The conclusions of the generalization measures were impressive for both ORF and word retell scores. After the treatment and phase two manipulations were implemented, generalization scores increased for all of the participants across both dependent variables. These results suggested that the intervention was effective for increasing comprehension of unpracticed reading material; however, these results only included one measure of comprehension.

The Read Naturally software used a comprehension quiz activity to measure the user’s ability to provide information about the main idea, to recall facts, to determine the meaning of vocabulary words, and to draw conclusions about each story. This analysis of comprehension is more in-depth than asking students to recall words that have been read. By combining the two assessments (i.e. comprehension questions and word recall), it is possible to gain a more well rounded picture of a student’s comprehension. This type of analysis was completed for training stories but was not conducted for generalization passages. The experimenter opted to use only the word-retell activity to check for comprehension on generalization passages. Although the results of this assessment were strong, they only provided one type of comprehension data to analyze. The link between ORF and comprehension would have been much stronger if a comprehension quiz similar to the one used in the RN software, was used with the generalization passages.
Lack of Standardized Reading Fluency Measure

One way to assess the effects of this intervention is through the use of standardized measures. This study used Woodcock-Johnson III subtests that focused on identifying letters and words, decoding skills, and comprehension. The results of these assessments revealed that participants’ skills were comparable to their same grade peers who scored in the average range. Although it appears that the intervention helped the participants to maintain similar scores as their peers, this study failed to conduct a standardized assessment that directly measured reading fluency.

There was some compensation for the lack of a fluency assessment with the use of the DIBELS winter and spring benchmark assessment; however, this is a curriculum based measure (Marston, Mirkin, & Deno, 1984) and not a norm referenced assessment. So, although 5 of 8 participants were able to decrease their risk status on the spring benchmark it still would have been beneficial to include another measure of post-intervention fluency. By failing to do so, it is difficult to draw any definitive conclusions in regards to comparing the ORF of the participants to their grade level peers.

Failure to Assess Generalization in Actual Classroom Reading Groups

As noted previously, this intervention was conducted under clinical conditions. In addition to controlling for environmental variables, the experimenter used reading material that was not an actual part of the participants’ classroom curriculum. Although, all of the materials used in this study were considered to be on grade level (i.e., first grade difficulty level) it is unclear if the results generalized to the participants’ classroom reading groups. There is mixed evidence to support the assumption that the skills may
have generalized; 3 out of 6 teachers responded “a lot” when asked how much participants’ fluency improved as a result of the study.

Even though half of the teachers reported that their students’ reading fluency had improved, no empirical data was collected on classroom performance to support or refute that report. Additionally, the other three teachers reported that their students’ fluency improved only “somewhat” or “a little”. Without a direct assessment of classroom ORF performance, it is difficult to reliably determine if the participants made gains outside of the research setting. This study did not include assess classroom performance and therefore cannot assume classroom generalization.

Implications for classroom use

This software program produced positive results for the oral reading fluency and comprehension of the participants. In light of the fact that there is a limited amount of empirical research on the use of the Read Naturally software edition, this study lends credibility to its effectiveness for students who are struggling with reading fluently. There are several benefits of this program and the way it is implemented but also there are some concerns.

First, this program is designed to be used as a stand alone, supplemental tool to improve reading fluency. Students who can use this program range from first to sixth grade. Once instructed on how to use the program correctly, students can be fairly independent in navigating through lessons. Students in the current study were introduced to the program in a 20-minute training lesson. Although, all of the students indicated that they had previous experience with using a computer; the majority of them also indicated
that they did not have a computer at home. In spite of this fact, students were able to engage with the program independently and with procedural integrity.

Procedural integrity checks were conducted for each participant to determine how well they completed each component of the program accurately. The results indicated that the participants accurately completed 95.5% of the steps correctly. The steps missed most frequently involved correctly identifying unknown words and reading all questions and answers before selecting an answer on the comprehension quiz. Some of the participants had the tendency to guess at words that they did not recognize. The program instructs users to click on any words that are unknown and the correct pronunciation is provided; however, if the user does not recognize that words are unknown, they will continue to read the words incorrectly. Treatment integrity data and anecdotal observation indicated that participants needed additional prompting to ensure they were getting feedback on missed words.

A key element of many fluency building activities is the error correction procedure (Begney, Daly III, & Valleley, 2006; Begney & Silber, 2006; Eckert, Dunn, Ardoin, 2006; Martens et al., 2007; Nelson, Alber, & Gordy, 2004; Pany & McCoy, 1988). Considering that young readers may guess at unknown words and therefore continually read them incorrectly, some level of teacher supervision is required. It is unclear of exactly how much supervision is needed but in this study two participants worked on separate laptop computers simultaneously. The experimenter was able to easily monitor both of them effectively. Although this is the case, the environment was
clinical and not subjected to typical classroom conditions. Teachers in a full and busy classroom may have a tougher time monitoring students at computer centers.

Another implication for classroom use is the program’s ability to be individualized for each student. The software contains administrative features that allow the classroom teacher to create and manipulate files for individual students. This is beneficial for students that may work at different paces. For example, several students in the current study were able to read at a faster rate than the other participants. In traditional larger or small group instruction, the groups need to be homogenous and progress at relatively the same rate. For students that begin to fall behind, it may be hard to help them to catch up with the rest of the group. With this computer program, participants were able to progress at their own pace and therefore avoid the aversiveness and stigma of being the “slow” student.

A similar benefit involves the ability to readjust goal criterion on an individual basis. All of the participants in this study began the first intervention phase with the same goal criteria. After roughly half of the treatment stories were completed, it was evident that all of the participants were reading at different ORF rates. In order to allow for the participants to progress at their optimal pace, goal criteria were adjusted to fit each of their individual rates of responding. As a result, there were higher rates of ORF on both treatment and generalization probes. Classroom teachers face the same dilemma, in which certain students progress much faster or slower than specified. With the ability to readjust the goal criterion individually, it is more likely that students can progress at a
rate that is suited to them, and permits gradual goal adjustments to make continuous progress.

The ability for students to work independently and at their own pace may be particularly beneficial to older students, who are likely to be embarrassed in group situations for not being strong readers. With a program like Read Naturally, students engage with the computer in relative isolation, reducing the likelihood of maladaptive escape behavior.

The final classroom implication involves increased motivation to participant in reading activities. As indicated in the social validity results, all of the participants gave high marks to the RN computer program. When asked what they enjoyed most about the program, many of them said the use of the computer. Most of the participants did not have computers at home and looked forward to the chance to use one, even for academic purposes. When asked how much they would want to use the program without the ability to earn prizes all of the participants answer “a lot”. It is likely that the use of computers served as a motivator. Research supports this assumption (Liu, 2005; Zardoya & Fico, 2001) and suggests that students are more likely to be motivated by using computers for academic work over traditional presentation formats.

Additional Considerations

Although there appears to be many benefits to using the RN software to increase reading fluency, there are a few drawbacks that teachers should be aware of. First, as with the use of any technology, the software did not always operate correctly 100% of the time. Procedural integrity data collected on how well the program operated revealed that it
crashed and needed to be rebooted in 3% of total sessions. This percentage is quite small but does suggest that there were times that the experimenter needed to devote time to reestablishing the participants’ ability to use the program properly. Teachers should be willing and able to troubleshoot technology problems to allow students to continue to make progress.

A more serious concern involves supervision of students and close monitoring of their progress. As noted earlier, participants did not always accurately identify all unknown words. If the experimenter was unable to monitor this, it is possible that these individuals would not have made sufficient progress. Another supervision issue involves student deception. It may not take a great deal of effort for students to figure out ways to circumvent the program and therefore progress through stories that they do not actually read. On the practice reading activity students are supposed to click on the last word they read when the bell rings to determine if they reached their goal. While the majority of participants performed this step with integrity, there was one occurrence in which a participant clicked ahead to advance to the next activity without actually reaching his goal. It is unclear if the participant would have continued to engage in this manipulation if he was not caught and corrected, but if he figured out how to “trick” the software, older students could certainly do the same. This is why periodic teacher supervision is strongly recommended for all students using this program.

Directions for future research

The results of the current study demonstrated promise in using a computer assisted instructional program to increase ORF and comprehension of struggling first
graders. In addition to increasing ORF and comprehension on treatment stories, by
manipulating individual fluency criteria for each participant both measures were
increased on generalization passages as well. Despite these results, there is still a need to
conduct additional research in this area for several reasons. The following section
discusses the reasons and possible ways to proceed.

First, the current study manipulated the fluency rates in an attempt to promote
generalization. As discussed earlier, previous research has not considered this type of
manipulation to promote increased ORF and/or generalization. In light of the results of
this study, future research should be conducted to verify the effectiveness this type of
manipulation. More specifically, future research can perform a similar manipulation to
determine if the results of this study can be replicated. In addition, future studies may
attempt to systematically increase fluency goal requirements of repeated reading material
to determine if an optimal level can be obtained to promote generalization to unpracticed
stories.

Second, although the results of this study were very positive, they were obtained
under clinical (non-classroom) conditions. These conditions included an empty,
distraction free classroom and consistent monitoring from the experimenter. The Read
Naturally software was designed to be used under typical classroom conditions. Although,
the experimenter interacted with the participants as little as possibly, he was still
available to answer questions and provide any additional support as needed. This
contrived environment is far from the typical first grade classroom. It is unclear if the
same robust results would have been achieved under different conditions. Future research
should assess the effectiveness of the Read Naturally software under more typical classroom conditions. The information obtained in the proposed study may be useful when making recommendations to school personnel about the classroom implementation.

Third, the participants in the current study were first-graders. Many studies that have targeted oral reading fluency tended to focus on older students. The results of this study demonstrated the effectiveness of the intervention for younger students. All of the participants made substantial gains in generalization and the majority of them were able to decrease their reading risk status by the end of the intervention. Future research should continue to target this population to identify additional strategies that can be used to remediate and prevent greater reading difficulties. Also, future studies should target a wider range of participants (i.e. higher grades) to determine if similar results are produced for older students.

Summary

This study investigated the Read Naturally software program to determine if it would be effective for increasing the oral reading fluency and comprehension of first grade students. Eight first graders were selected as participants because they were identified by their teachers as exhibiting low fluency in their classroom reading groups. Pre-intervention assessments confirmed that all the participants were either within the “at-risk” or “some-risk” category on the DIBELS winter benchmark ORF assessment. The study sought to answer research questions about the effects of the program in ORF and comprehension on both practiced and unpracticed, grade-level reading material. A
multiple probe research design was used to analyze the dependent variables. In addition, pretest and posttest scores were obtained using standardized and curriculum based reading assessments.

The results of this study were that the Read Naturally software program effectively helped all eight participants increase their ORF and comprehension. Not only did the intervention produce rapid gains in reading fluency, participants were able to make gains on the standardized and curriculum based assessments. One of the most relevant findings involved participants’ ability to decrease their risk status on the DIBELS ORF spring benchmark assessment.

Despite these encouraging results, it should be noted that generalization gains were not consistent for all the participants until an additional manipulation was made to the research procedures. During the second phase of the intervention, fluency requirements for practice stories were increased and individualized for each participant. Following this increase all of the participants were able to generalize their reading skills to unpracticed stories. The outcomes of this study provide empirical evidence of the effectiveness of computer technology used as a tool for secondary interventions. The results are encouraging because they suggest that students who fall behind and need individualized instruction can make progress through the use of computer technology. In light of typical classroom conditions, where human resources are limited, the computer appears to be a potentially powerful educational tool that could be used more extensively to prevent and remedy academic risk and school failure.
REFERENCES


Glasser, W. Address delivered in Annapolis, Maryland November 5, 1981.


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(4), 1-15.

Hapstak, J., & Tracey, D. (2007). Effects of assisted-repeated reading on students of
varying reading ability: A single-subject experimental research study. *Reading
Horizons, 47*(4), 315-334.

Harlaar, N., Dale, P. S., & Plomin, R. (2007). From learning to read to reading to learn:

principle on the reading development of first graders. *Journal of Learning
Disabilities, 41*(2), 143-157.


Hasbrouck, J. E., Ihnot, C., & Rogers, G. H. (1999). "Read naturally": A strategy to
increase oral reading fluency. *Reading Research and Instruction, 39,* 27-38.

Behavioral Education, 8,* 63-79.

Hindson, B., Byrne, B., Fielding-Barnsley, R., Newman, C., Hine, D. W., & Shankweiler,


Koskinen, P. S., & Blum, I. H. (1986). *Paired repeated reading: A classroom strategy for developing fluent reading*


Rasinski, T. V. (1989). *The effects of repeated reading and repeated listening while reading on reading fluency*


Rasinski, T. V. (1989). *The effects of repeated reading and repeated listening while reading on reading fluency*


Thornton, L. H., & Vinzant, R. S. (2000). *The effects of phonemic awareness instruction in first grade on the reading scores of rural primary students*


APPENDIX A

Computer Training Procedural Integrity Checklist
Training PI Checklist

Student: ____________ Date: _________________

<table>
<thead>
<tr>
<th>Trainers Actions</th>
<th>Yes</th>
<th>No</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduced program</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asked/checked if participant can use a mouse</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructed to click on story</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduced key words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduced cold timing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduced unknown words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prompted to click last word read</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prompted to click unknown words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explained goal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explained read along</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explained practice reading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asked if there were any questions about the program</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX B

Example of Read Naturally Story
Story 1- Bears

Bears are big animals. They have big heads. They have big feet. Bears are very strong animals. They have short legs. They can kill a man with one hit. Bears have little ears. They have little eyes. They do not see well. Bears have long fur. A grizzly bear has brown fur. A polar bear has white fur. A black bear has black fur. Cubs are baby bears.
Comprehension questions for story 1 - bears

1) What color is a grizzly bear?
   a. Brown
   b. Black
   c. White

2) What is a baby bear called?
   a. cub
   b. bug
   c. doe

3) What parts of a bear are big?
   a. Feet and head
   b. Ears and eyes
   c. Nose and fur

4) What part of a bear is very strong?
   a. Ears
   b. Eyes
   c. Legs

5) Name one thing a bear can not do well.
Story 5 – Hats

People wear hats on their heads. Some people wear hats to keep them warm. Warm hats can be made of fur. Some people wear hats in the sun. Sun hats can be made of straw. Some people wear hats at work. Work hats keep them safe. Some people wear hats to be funny. Clowns wear hats to make us laugh. Some people wear hats to look good. Do you look good in a hat?
Comprehension questions for story 5 - hats

1) What do people wear on their heads?
   a. Shoes
   b. Hats
   c. Gloves

2) What hats do clowns wear?
   a. Funny hats
   b. Fur hats
   c. Work hats

3) What hats keep us safe?
   a. Clown hats
   b. Straw hats
   c. Work hats

4) What hat do you wear when it snows?
   a. A funny hat
   b. A warm hat
   c. A snow hat

5) What are three reasons people wear hats?
APPENDIX C

Examples of Aimsweb Stories
Story 1 - Tom

A boy named Tom was at the bus stop. He was waiting for the school bus. There was no one there but him. The bus was late.

Tom began to talk to himself. "Maybe the bus forgot me," he said.

Then Tom heard a dog barking. He looked up and saw his dog, Spot, running down the road. Spot ran to Tom. He was so happy to see Tom that he jumped into Tom's arms.

Just then, Tom heard the bus coming. He didn't have time to take Spot home. There was no time to think. Tom grabbed Spot and hid him under his coat.

The bus pulled up to Tom's bus stop. Tom got on the bus and went to the back. His friend Jack had saved a seat for him. Just as Tom sat down, a little yelp came from under his coat.

"What do you have under there, Tom?" asked Jack.

"If I tell you, do you promise not to tell?" replied Tom.

"You bet. I'm your best friend, aren't I?" asked Jack.

Tom told Jack what had happened. He asked his friend what he should do. Jack had an idea.
"You can tell the teacher you have something very cool for show and tell. Then you could call your mom and have her come and pick up Spot."

Tom decided that's what he would do. His teacher was surprised. His mom was mad, but Spot was very happy.
APPENDIX D

Procedural Integrity Checklist 1 - Computer Presentation
**Computer PI Checklist-2**

Student: ____________ Date: ________________

<table>
<thead>
<tr>
<th>Read Naturally Actions</th>
<th>Yes</th>
<th>No</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presents selected story</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presents all key words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timer starts when clicked</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Displays correct goal and cold timing scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presents story three times during read along</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presents “next” button after third read along</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presents all comprehension questions during quiz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Represents incorrect comprehension questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prompts participant to keep practicing until a teacher comes to pass</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presents password protection for pass timing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correctly scores all measures</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E

Procedural Integrity Checklist 2 - Data Collection PI
## Data Collection PI Checklist

Student: ___________ Date: _______________ Story: _______________

<table>
<thead>
<tr>
<th>Item</th>
<th>Yes</th>
<th>No</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presented story</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructed participant to read as fast as he/she can</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructed participant to read as best as he/she can</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Told student would provide correct word if it was unknown</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provided all unknown or mispronounced words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructed participant to remember everything about the story</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Started/stopped timer correctly</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX F

Procedural Integrity Checklist 3 - Software Uses by Participants
## Computer PI Checklist

Student: __________  Date: __________  Story: ______________

<table>
<thead>
<tr>
<th>Activity</th>
<th>Yes</th>
<th>No</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clicked key words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read along three times</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attended during read along</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clicked unknown words during practice</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clicked correct last word after bells sounds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read all questions and answers before clicking answer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asked for help when needed</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Student: __________  Date: __________  Story: ______________

<table>
<thead>
<tr>
<th>Activity</th>
<th>Yes</th>
<th>No</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clicked key words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read along three times</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attended during read along</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clicked unknown words during practice</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clicked correct last word after bells sounds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read all questions and answers before clicking answer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asked for help when needed</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX G

Social Validity Questionnaire – Teachers
Social Validity Questionnaire – Teachers

1) Do you feel like the student’s overall reading had improved after the intervention was complete?
   a. A little
   b. Somewhat
   c. A lot

2) Would you say that the student became a more fluent reader after the intervention was complete?
   a. A little
   b. Somewhat
   c. A lot

3) Would you allow any of your students to participate in similar programs in the future?
   a. A little
   b. Somewhat
   c. A lot

4) Do you think the students enjoyed this reading intervention?
   a. A little
   b. Somewhat
   c. A lot

5) How important do you think reading intervention programs are for struggling reading?
   a. A little
   b. Somewhat
   c. A lot

Comments:
APPENDIX H

Social Validity Questionnaire – Participants
Social Validity Questionnaire – Participants

1) Did you like reading on the computer?
   a. Yes
   b. No

   Why?

2) Do you think you became a better reader?
   a. Yes
   b. No

   Why?

3) Did you like working for prizes?
   a. Yes
   b. No

   Why?

4) Would you have worked on the computer even if you did not get prizes?
   a. Yes
   b. No

   Why?

5) Would you like to continue to read on the computer?
   a. Yes
   b. No

   Why?
APPENDIX I

Sample Read Naturally Lesson
Each Read Naturally level includes 24 stories in two sets of 12. The set of available stories is based on the reading level the teacher chooses for the student. Students work on the stories in a level until they pass all the stories, or they are ready for a new level. Allowing students to choose their own story provides them with a sense of ownership over their work.

1. Stories that the student has completed are marked "Passed." Students must complete a story before selecting a new one.
2. Students click a picture to select a story from the ones available at their reading level.
Key Words

Abraham Lincoln

Key Words

Civil War
lawyer
politics
self-made

1

lawyer

A lawyer is someone who works with the law.

John needed a lawyer because he broke the law.

2

3

Quit

Next
The Key Words step provides students with vocabulary instruction, teaching them key words that are used in the story. They learn how to pronounce the words and what they mean.

Learning and saying these key words helps students read them with greater confidence when they encounter the words in the story. The Key Words step also encourages students to think about what they will be learning from the story.

Screen Highlights

1. Students listen and read along as the key words are read for them.
2. Students click each of the words to see and hear their definitions.
3. Students click the Next button when they are ready to move to the next step.

Cold Timing

Abraham Lincoln was born in Kentucky. His mother died when he was 6 years old. Abraham was tall and strong. He worked hard and helped his neighbors. He loved to wrestle and tell stories. He only went to school for a year. Abe learned to read by himself.

When Abe grew up and left home, he had many jobs. He took a flatboat down the Mississippi River to New Orleans. He worked in a lumber mill. He was a stenographer. Then he was a postmaster. At the time he kept learning. He read books about the law. Abe wanted to be a lawyer. In 1836, Abe passed a test and became a lawyer.

Then he got into politics. In 1860, Abraham Lincoln was elected the President of the United States. President Lincoln became one of our greatest presidents. He led the United States in the Civil War. He is remembered as the president who freed the slaves. And Abraham Lincoln was truly a self-made man. When he was killed in 1865, it was a great loss for America.
In the Cold Timing step, the program times students for one minute as they read the story for the first time. This establishes a baseline for measuring students' improvement as they work on the story.

As students read, they should click words they don't know or have problems reading. Clicking on these difficult words increases students' awareness of unknown words and alerts teachers to words or word patterns that need to be taught.

1. Students can turn a reading guide on or off before they start reading. The guide highlights the line of text under the cursor, helping the students' eyes track across the line.
2. Students click on the stop watch button to begin the one minute timing.
3. Students read the story aloud and click words they don't know or have problems with. The words they click on change color and are subtracted from the final cold timing score. After the timing is complete, students can click the highlighted words to hear them spoken and practice reading them.
4. Students can click **Finished** if they finish the story before one minute has passed. If students click **Finished** too early (before they could have completed reading the story), the program may ask them to read again.
5. The Cold Timing graph shows the number of words the student read correctly, subtracting the number of difficult words the student clicked on from the total number of words read. The screen also displays the student's score and goal.
Abraham Lincoln was born in Kentucky in 1809. His mother died when he was 9 years old. Abraham was tall and strong. He worked hard and helped his neighbors. He loved to wrestle and tell stories. He only went to school for a year. Abe learned to read by himself.

When Abe grew up and left home, he had many jobs. He took a flatboat down the Mississippi River to New Orleans. He worked in a lumber mill. He was a storekeeper. Then he was a postmaster. In the time he kept learning. He read books about the law. Abe wanted to be a lawyer. In 1836, Abe passed a test and became a lawyer.

Then he got into politics. In 1860, Abraham Lincoln was elected the President of the United States. President Lincoln became one of our greatest presidents. He led the United States in the Civil War. He is remembered as the president who freed the slaves. And Abraham Lincoln was truly a self-made man. When he was killed in 1865, it was a great loss for America.
In the Read Along step, students read along while listening to a recording of the story, usually three times. (When setting up individual students' story options, teachers can specify the number of read alongs, as well as the speed of each read along.)

This step helps students learn new words and encourages proper pronunciation, expression, and phrasing. Students should quietly subvocalize as they read to ensure that they actually read along.

1. When the read along begins, each sentence in the story is highlighted as it is being read. Students quietly read along with the recording.
2. After each read along, students can click the blue vocabulary words to see and hear their definitions.
3. After completing the required number of read alongs, students can either click Start to read along again for more practice, or click Next to continue to the next step.
Practice

In the Practice step, students are timed as they practice reading the story without the recording. They are required to practice until they can read the story at or above the goal rate that the teacher has set for them. It may take students from three to ten practice readings to reach their goals.

1. A box displays the number of words per minute the student read. Students who did not meet their goals must read again.

2. The teacher can review details about the student's progress through the story (for example, a list of difficult words the student clicked on in the Cold Timing and Practice steps).

3. The student can return to the Read Along step, continue practicing or, if the goal has been met, move to the next step.
Quiz

Abraham Lincoln was born in Kentucky in 1809. His mother died when he was 9 years old. Abraham was tall and strong. He worked hard and helped his neighbors. He loved to wrestle and tell stories. He only went to school for a year. Abe learned to read by himself.

When Abe grew up and left home, he had many jobs. He took a flatboat down the Mississippi River to New Orleans. He worked in a lumber mill. He was a storekeeper. Then he was a postmaster. At the time he kept learning. He read books about the law. Abe wanted to be a lawyer. In 1830, Abe passed a test and became a lawyer.

Then he got into politics. In 1860, Abraham Lincoln was elected the President of the United States. President Lincoln became one of our greatest presidents. He led the United States in the Civil War. He is remembered as the president who freed the slaves. And Abraham Lincoln was truly a self-made man. When he was killed in 1865, it was a great loss for America.

1. What is the main idea of this story?
   A. Lincoln was a self-made man who became a great president.
   B. Lincoln was the President during the Civil War.
   C. Lincoln loved to work hard, wrestle, and tell stories.

2. Why do we say Abraham Lincoln was a self-made man?
   He became successful with not much help from other people.

3. What did Abraham Lincoln do when he was a storekeeper?
   He kept learning and read books about the law.
In the Quiz step, students answer up to nine comprehension questions about the story. The program checks the answers to the multiple-choice questions and repeats questions that the student answers incorrectly. During the Pass step, the teacher reviews and corrects the open-ended questions.

1. Students can review the story text and click on vocabulary words prior to answering the questions.
2. Students answer each of the questions in order. Each question focuses on a specific comprehension skill. For example, Question 1 tests whether the student understood the main story idea. The Comprehension Report enables teachers to track a pattern of errors and provide additional instruction in those areas.
3. Students type answers to open-ended questions using their own words. Teachers review and correct these questions during the Pass step, using the SE Answer Guide.
Pass Timing

Abraham Lincoln was born in Kentucky in 1809. His mother died when he was 9 years old. Abraham was tall and strong. He worked hard and helped his neighbors. He loved to wrestle and tell stories. He only went to school for a year. Abe learned to read by himself. When Abe grew up and left home, he had many jobs. He took a flatboat down the Mississippi River to New Orleans. He worked in a lumber mill. He was a storekeeper. Then he was a postmaster. At the time he kept learning. He read books about the law. Abe wanted to be a lawyer. In 1836, Abe passed a test and became a lawyer.

Then he got into politics. In 1860, Abraham Lincoln was elected the President of the United States. President Lincoln became one of our greatest presidents. He led the United States in the Civil War. He is remembered as the president who freed the slaves. And Abraham Lincoln was truly a self-made man. When he was killed in 1865, it was a great loss for America.

Words read: 105
Enter the number of words the student missed: 2
Pass timing score: 103
Enter a score from 1 to 4 to rate the student's reading expression (4 is excellent):
At the Pass step, teachers listen to students read and determine whether they are ready to move on to the next story. Students pass if they meet the following criteria:

- They reach their reading rate goal (words correct per minute).
- They make no more than three reading errors.
- They read with good expression.
- They answer the questions correctly.

If they do not pass, teachers assign some remedial activity and test the students again.

1. After students complete all previous steps, they are held at the Ready to Pass screen until the teacher is available. Students can continue practicing.
2. The teacher clicks Pass to enter a password and begin the pass timing.
3. After the timing is complete, the student clicks on the last word read (or the Finished button), and the program calculates the score.
4. The teacher enters the number of words missed.
5. The teacher enters an expression score on a scale of 1 to 4. Students must score 2 or higher to pass.
Results

When a student passes, the program provides visual and auditory feedback. It also enables the teacher to review the results of the student's progress through the story.

Screen Highlights

1. The difference between the cold timing bar and the pass timing bar shows the student's progress.
2. The teacher can display a results graph of all the stories the student has read in the current level.
3. The teacher can display the student's detailed results for the current story.
4. The student can select a new story to read.