The Utility of the Individual Reading Evaluation and Diagnostic Reading Inventory, a Specific Reading Skills Assessment, for Treatment Design and Implementation

Andrew J Koerner
THE UTILITY OF THE
INDIVIDUAL READING EVALUATION AND DIAGNOSTIC
(iREAD) INVENTORY,
A SPECIFIC READING SKILLS ASSESSMENT,
FOR TREATMENT DESIGN AND IMPLEMENTATION

A Dissertation Presented
by
ANDREW J. KOERNER
Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY
September 2008
School Psychology
THE UTILITY OF THE
INDIVIDUAL READING EVALUATION AND DIAGNOSTIC
(iREAD) INVENTORY,
A SPECIFIC READING SKILLS ASSESSMENT,
FOR TREATMENT DESIGN AND IMPLEMENTATION

A Dissertation Presented

by

ANDREW J. KOERNER

Approved as to style and content by:

________________________________________
John M. Hintze, Chair

________________________________________
Gary Stoner, Member

________________________________________
Elizabeth Harvey, Member

________________________________________
William J. Matthews, Member

_______________________________________
Christine B. McCormick, Dean
School of Education
ACKNOWLEDGEMENTS

I would like to thank my advisor, Gary Stoner, for his many years of encouragement and direction in supporting my development as a school psychologist. Thanks are also due to John M. Hintze, whose initial response to my question about curriculum based measures of specific reading skills led to the development of the Individualized Reading Evaluation and Diagnostic (iRead) Inventory. Professor Hintze has also been critical to the development of this study and the analysis of the results. I also extend a thank you to Elizabeth Harvey and William J. Matthews, the other members of my dissertation committee, for their support and feedback during this project.

I wish to express my appreciation to Bonnie Robinson who diligently worked with the students to support my study and their learning. Thanks go to James Cressey as well for his part in implementing the interventions and in providing feedback regarding the ongoing process of fine-tuning the iRead Inventory. Virginia Reed deserves my thank you as well for her dedication to assuring the completion of the project.

A special thank you is due to my supportive and encouraging wife, Patricia Sayre. Without her support, encouragement, and willingness to take on many family responsibilities when I was too busy, I would not have been able to complete this project.
ABSTRACT

THE UTILITY OF THE
INDIVIDUAL READING EVALUATION AND DIAGNOSTIC
(iREAD) INVENTORY,
A SPECIFIC READING SKILLS ASSESSMENT,
FOR TREATMENT DESIGN AND IMPLEMENTATION
SEPTEMBER 2008
ANDREW J. KOERNER, B.S., CORNELL UNIVERSITY
M.A., LESLEY UNIVERSITY
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST
Directed by: Professor John M. Hintze

This study was conducted to assess the effectiveness of the Individualized Reading Evaluation and Diagnosis (iRead) Inventory for accurately assessing specific decoding sub-skill weaknesses and for informing the development of targeted interventions to improve the reading abilities of students. The iRead Inventory is a curriculum-based, specific skills mastery measurement tool for assessing specific decoding weaknesses. Students read word lists targeted to specific vowel combinations to determine weaknesses with particular combinations. The study assessed whether the iRead Inventory could distinguish specific decoding sub-skill weaknesses for students and whether the iRead Inventory was effective in supporting the development of interventions to improve those decoding weaknesses.
Students were screened for dysfluency and three students were identified as having primarily decoding issues were selected for the intervention phase of the study. The intervention phase of the study involved using a multiple baseline, randomization design with the three participants receiving interventions beginning at randomly selected times. The iRead Inventory was utilized to identify specific vowel combination difficulties for intervention and the participants were provided direct, sequential instruction targeted to the identified specific decoding weaknesses. The participants’ reading progress was monitored using Reading-CBM (R-CBM) and Nonsense Word Fluency (NWF) measures. In addition, their progress with learning the specific sub-skills was monitored using the iRead Inventory.

The iRead Inventory was found to reliably assess specific decoding deficits. Interventions that were developed using the iRead Inventory were shown to improve the decoding abilities of all the participants. The two participants who received interventions earlier showed gains in oral reading skills and mastered a number of specific vowel combination decoding skills. The participant who began interventions last showed less gain in both abilities. In addition, there seemed to be a learning curve phenomenon whereby participants did not exhibit gains associated with the interventions until approximately two and one half weeks after interventions were initiated. Further research can include assessing the reliability of the iRead Inventory, researching its utility for designing interventions for a broader population, and assessing the implications of a potential learning curve phenomenon for making educational decisions.
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>A. Importance of Literacy Instruction</td>
<td>1</td>
</tr>
<tr>
<td>B. Findings of Research Relevant to the Acquisition of Reading Skills</td>
<td>1</td>
</tr>
<tr>
<td>C. Legislative Efforts to Address Reading Instruction</td>
<td>6</td>
</tr>
<tr>
<td>D. Response to Intervention and the Implications for Education</td>
<td>8</td>
</tr>
<tr>
<td>E. Curriculum Based Measurement and Formative Assessment</td>
<td>12</td>
</tr>
<tr>
<td>F. The Nature of Specific Sub-Skill Mastery Measurement</td>
<td>15</td>
</tr>
<tr>
<td>G. Efforts to Use CBM to Assess and Target Specific Skills</td>
<td>17</td>
</tr>
<tr>
<td>E. The Development of the iRead Inventory</td>
<td>18</td>
</tr>
<tr>
<td>F. Current Study</td>
<td>20</td>
</tr>
<tr>
<td>II. METHODS</td>
<td>23</td>
</tr>
<tr>
<td>A. Participants and Setting</td>
<td>23</td>
</tr>
<tr>
<td>B. Materials/Tools</td>
<td>24</td>
</tr>
<tr>
<td>C. Procedures</td>
<td>27</td>
</tr>
<tr>
<td>1. Design</td>
<td>27</td>
</tr>
<tr>
<td>2. Screening</td>
<td>30</td>
</tr>
<tr>
<td>3. Progress Monitoring and Data Collection</td>
<td>32</td>
</tr>
<tr>
<td>4. Intervention Procedures</td>
<td>33</td>
</tr>
<tr>
<td>5. Treatment Integrity and Inter-observer Agreement</td>
<td>35</td>
</tr>
<tr>
<td>III. RESULTS</td>
<td>38</td>
</tr>
<tr>
<td>A. Data Screening</td>
<td>38</td>
</tr>
<tr>
<td>B. Summative Analysis</td>
<td>39</td>
</tr>
<tr>
<td>C. Visual Analysis</td>
<td>41</td>
</tr>
<tr>
<td>D. Statistical Analysis</td>
<td>47</td>
</tr>
</tbody>
</table>
1. Randomization Analysis ................................................................. 47
2. Effect Size Calculation ................................................................. 48
3. Percentage of All Non-Overlapping Data Results ...................... 50
4. Allison Regression Approach Results ........................................ 51

IV. DISCUSSION .................................................................................... 53

A. The iRead Inventory as a Tool for Identifying
   Specific Decoding Weaknesses ..................................................... 54
B. The iRead Inventory as a Tool for Supporting Reading Improvement .... 55

1. Summative Findings .................................................................. 56
2. Visual Inspection Findings .......................................................... 59
3. Randomization Analysis Findings ................................................. 62
4. Percentage of All Non-Overlapping Data Findings ..................... 62
5. Allison Effect Size Calculation Findings ....................................... 63

C. Implications .................................................................................. 65
D. Study Limitations and Directions for Future Research .................. 68

APPENDICES

A. CONSENT FORM ........................................................................... 71
B. IREAD INVENTORY
   GENERAL ADMINISTRATION INSTRUCTIONS ....................... 72
C. ASSESSMENT SAMPLE PAGES .................................................... 73
D. LESSON PLAN SAMPLES .......................................................... 75
E. INTERVENTION FIDELITY FORM ................................................ 78
F. DETAILED IREAD INVENTORY SUB-SKILL RESULTS
   FOR ALL PARTICIPANTS ............................................................. 79

REFERENCES ...................................................................................... 82
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. R-CBM, iRead, NWF, and PSF Screening Results</td>
<td>32</td>
</tr>
<tr>
<td>2. Pre-Intervention and Post-Intervention NWF Results</td>
<td>39</td>
</tr>
<tr>
<td>3. Pre-Intervention and Post-Intervention R-CBM Results</td>
<td>40</td>
</tr>
<tr>
<td>4. Pre-Intervention and Post-Intervention iRead Results</td>
<td>40</td>
</tr>
<tr>
<td>5. Mean Difference Results</td>
<td>42</td>
</tr>
<tr>
<td>6. Effect Size Results for Multiple-Baseline Data for all Three Participants</td>
<td>52</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>1.</td>
<td>Participant 1: NWF mean difference results</td>
</tr>
<tr>
<td>2.</td>
<td>Participant 2: NWF mean difference results</td>
</tr>
<tr>
<td>3.</td>
<td>Participant 3: NWF mean difference results</td>
</tr>
<tr>
<td>4.</td>
<td>Participant 1: R-CBM mean difference results</td>
</tr>
<tr>
<td>5.</td>
<td>Participant 2: R-CBM mean difference results</td>
</tr>
<tr>
<td>6.</td>
<td>Participant 3: R-CBM mean difference results</td>
</tr>
<tr>
<td>7.</td>
<td>Participant 1: NWF slope difference results</td>
</tr>
<tr>
<td>8.</td>
<td>Participant 2: NWF slope difference results</td>
</tr>
<tr>
<td>9.</td>
<td>Participant 3: NWF slope difference results</td>
</tr>
<tr>
<td>10.</td>
<td>Participant 1: R-CBM slope difference results</td>
</tr>
<tr>
<td>11.</td>
<td>Participant 2: R-CBM slope difference results</td>
</tr>
<tr>
<td>12.</td>
<td>Participant 3: R-CBM slope difference results</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Importance of Literacy Instruction

Reading is a skill necessary for anyone to be a productive member of modern American society. It is a cultural imperative that, by the time they are adults, children will have learned to read. Yet, there is ongoing concern about the ability of the American public school system to educate all students to read. Statistics indicate that one quarter of adults are functionally illiterate (Riley, 1996; United States Department of Labor, 2006) and this leads to problems such as high drop out rates (Juel, 1995), incarceration, poor health maintenance, and poverty (United States Department of Labor [USDL], 2006). Furthermore statistics continue to indicate that American students are falling behind their peers in other developed countries and even in some developing countries (National Research Council [NRC], 1998; USDL, 2006). These statistics have significant implications for the economy and well being of all of the members of our American society. Over the years many efforts have been made to address how best to support children in learning to read. This study will focus on the utility of a curriculum based assessment approach for designing effective educational interventions for beginning readers who are struggling with basic decoding skills.

Findings of Research Relevant to the Acquisition of Reading Skills

Difficulty with reading skills emerges very early in a given child’s development and, if not addressed early, becomes increasingly difficult to remediate. Stanovich (1986) described a phenomenon that he called the “Matthew Effect” in reading achievement whereby the gap between poor and good readers continued to
widen as they progressed through the grades. The slope of reading skill improvement for poor readers was generally less than that of good readers as they progressed through school. He argued that interventions not only need to improve the reading skills acquisition slope for poor readers to parallel those of the good reader, the slope needed to be greater for poor readers so they could catch up. Juel (1988) found a similar result in a longitudinal study of reading and writing acquisition. Children who were poor readers in first grade tended to be poor readers in fourth grade. Juel found that 88% of the good readers at the end of first grade were proficient at the end of fourth grade and 87% of the poor first grader readers continued to struggle in fourth grade. As the students got older, the level of discrepancy between poor and good readers increased as well, a result similar to that of Stanovich (1986). Additionally, remediation of the reading difficulties of the poor readers took increasingly greater resources as they grew older. Another significant finding was that poor readers read significantly less on their own and this contributed to the increasing gap as well as limiting their participation in the one simple activity that might help them improve in reading.

Research also has found that differences in reading abilities begin prior to schooling (Hart & Risley, 1995). Hart and Risley conducted a longitudinal study that focused on the acquisition of vocabulary skills, a necessary preliteracy skill. They found that significant differences existed by age three among children of different socio-economic classes, regardless of race. Children from low socio-economic status (SES) homes had learned significantly fewer words than children from middle or high SES homes. By the time the children had reached four years old the differences were
stark. In poverty stricken families the child had learned an average of 13 million words, the working class child 26 million and the child in a professional family 45 million words (Hart & Risley, 1995). The authors concluded that, given these discrepancies, it was critical to intervene early in the student’s school history in order to remediate these skills that have not been acquired.

In 1997, Congress created the National Reading Panel (NRP), in conjunction with the Secretary of Education and the Director of the National Institute of Child Health and Human Development (NICHD), in an effort to study the research accumulated to date regarding how best to teach reading. The NRP was charged with the task of assessing “…the status of research-based knowledge, including the effectiveness of various approaches to teaching children to read” (National Reading Panel [NRP], 2000, p. 1-1). After considering the foundational work of the National Research Council (NRC) Committee on Preventing Reading Difficulties in Young Children (Snow, 1998) the NRP conducted hearings and adopted the following five key component of literacy instruction to study: alphabetics (phonemic awareness and phonics instruction), fluency, comprehension (vocabulary, text comprehension, and strategy instruction), teacher education, and computer technology. The NRP members then developed a rigorous set of methodological standards to apply to the process of analyzing the broad field of research in reading instruction. These standards were then applied to a meta-analysis of the research available to date. There were some areas of research for which there were not enough studies available and so for these areas more subjective qualitative analyses were used to arrive at conclusions and recommendations.
After analyzing the literature, the NRP identified five basic areas of reading development: phonemic awareness, phonics, fluency, vocabulary and comprehension. These were identified as key components for any child’s process of learning to read and for planning instruction. Phonemic awareness, phonics and fluency were viewed as foundational skills that needed to be addressed to learn to read and vocabulary and comprehension skills were secondary level skills that were necessary to reading to learn. The emphasis of instruction in the earlier grades should focus on the foundational skills and shift to the secondary level skills as students’ progress through the grades. These skills are, however, not acquired solely in a linear fashion. Therefore, instruction needs to occur in all five areas throughout all of the elementary grades, though the need for instruction in foundational skills diminishes over time unless if a student is struggling.

The three foundational skills involve understanding the alphabetic principle and developing automaticity with the reading process. Phonemic awareness is “…the ability to notice, think about, and work with the individual sounds (phonemes) in spoken words…” (Schumm, 2006, p. 91). The NRP found that this skill was a significant indicator of learning to read and spell. Furthermore, it was found that there are effective ways to teach Phonemic Awareness and that the effects of that instruction lasted well beyond the initial lessons. Phonics is “…the teaching of sound-symbol relationships and patterns.” (Schumm, 2006, p. 119). The NRP meta-analysis found that systematic and explicit instruction of letter-sound correspondences was the most effective way to teach students how to decode words and how to spell. These two aspects of the alphabetic principle were key building blocks for the next step of
reading development, reading fluently. Fluency refers to the ability of readers to “…read text with speed, accuracy, and proper expression.” (NRP, 2000, p. 3-1)

Fluency involves automatically recognizing words one knows and rapidly decoding unfamiliar words and is important for comprehension, particularly as passages become longer. The NRP found that reading fluency is a skill that must receive greater emphasis in classroom instruction and that it can be improved through guided oral reading and repeated readings. Fluency was found to be a good indicator of progress with reading but the NRP cautioned that “…word recognition accuracy is not the end point of reading instruction.” (NRP, 2000, p. 3-3) and the focus of reading instruction is ultimately understanding and comprehension of the material in order to be able to learn.

The secondary level skills analyzed by the NRP included vocabulary skills and comprehension. These two areas proved more difficult to analyze because of their complexity and the inter-relationships of multiple skills and the NRP was not able to conduct meta-analyses of either area. As noted by a number of researchers (Hart & Risley, 1995; Juel, 1988; Stanovich, 1986), vocabulary is a pre-reading skill that is influenced by SES and certainly has a strong influence on the process of learning to read and understand what is being read. Vocabulary can be thought of as “…having full and precise understanding of the meanings of words,” (Schumm, 2006, p. 299). The NRP did recommend a mix of direct and indirect instruction methods and emphasized the need for multiple exposures to new words within a context.

Comprehension is “…an understanding of words and how those words are used to create meaning.” (Schumm, 2006, p. 223). Comprehension is a complex process
involving word knowledge, analyzing context, evaluating words, phrases and sentences and using prior knowledge. With respect to comprehension, the NRP found that a combination of reading comprehension strategies were useful when taught explicitly though the research is still fairly inconclusive regarding which strategies were most effective at which ages. The NRP concluded that more research was necessary regarding both vocabulary and comprehension instruction.

A critical recommendation of the NRP report was that evidence based practices should be used in the process of reading instruction. As part of this recommendation, the NRP outlined specific areas for further research into assessment and instruction as well as delineating standards for the quality of that research. Additionally, the NRP recommended that assessment of each of these skill areas needed to be an ongoing part of instruction. The NRP called for the development of assessment tools that could be easily utilized in the classroom and be utilized over time in a formative manner to assist teachers in modifying their instructional approaches to meet the needs of individual students. Lastly, the NRP report functioned as a respected synthesis of research based approaches to reading instruction. The findings and recommendations of this report continue to provide a framework for educational, policy and legislative efforts to address the issue of improving the literacy of all students.

**Legislative Efforts to Address Reading Instruction**

Legislation has been enacted since the mid-1950’s to assure that all children grow up learning to read. The most recent legislation that was passed to address the issue of reading instruction was the NCLB Act of 2001. This legislation established a goal that, by the end of third grade, all students must be proficient readers. In addition,
it mandated the use of research-based instruction, statewide assessments regarding students’ progress towards proficiency and the development of a system to make schools accountable if students fail to learn. NCLB also included a provision for early assessment and intervention to address reading difficulties. These mandates were all part of a comprehensive effort to bring many years of research into teaching reading into the daily practices of teaching children in all American schools.

Another significant legislative effort was the Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004). IDEA 2004 was a reauthorization of the special education laws designed to address the needs of those students who struggle the most with learning. IDEA 2004 had a particular focus on early intervention for reading and on linking assessment and instruction. In addition, IDEA 2004 clarified procedures regarding monitoring the progress of students who were receiving services and emphasized that interventions needed to change if they were not effective. Lastly, IDEA 2004 allowed for the utilization of a Response to Intervention model (RTI) for identifying students with disabilities. The RTI model, while incorporated into a special education law, has significant implications for the process of general education. RTI involves interventions prior to placement in special education and includes the assumption that interventions within the general education setting will be evidence based, implemented with integrity and monitored for effectiveness. Therefore, IDEA 2004 has major implications for the assessment and educational planning process in general education.
Response to Intervention and the Implications for Education

The primary method for determining eligibility for special education services has been the discrepancy model whereby a child became eligible for services only if there was a significant discrepancy between the child’s ability (generally measured by IQ testing) and achievement (generally measured by a combination of classroom performance and achievement testing). The discrepancy model has been found to be problematic in several ways (Aaron, 1997; Bray, Kehle, & Hintze, 1998; Fuchs, Mock, Morgan, & Young, 2003; Gresham & Witt, 1997; Reschly, 1997).

Aaron (1997) argued that the assumptions behind the discrepancy model have been shown to be false, namely there is little or no measurable difference in the etiology of learning disabled vs. non-learning disabled readers and the instructional approaches to remediate reading difficulties do not, in fact, differ between the two groups. Research indicates that focusing on the specific cause of the reading difficulty is more effective for remediation, regardless of classification through the use of the discrepancy model. Bray et al. (1998) argue that the discrepancy model and use of profile analysis is flawed because of psychometric issues (profiles used to define LD occur in high percentages of the population), validity issues (IQ only accounts for 25% of the variance in academic achievement) and ethical issues (while IQ tests measure overall intelligence fairly accurately, the way they are used is not in line with their intended purpose). Reschly (1997) echoes these concerns as well and emphasizes the ethical issues involved in using an assessment approach that is not supported by research. Fuchs, Mock, et al. (2003) also argue that the IQ discrepancy model does not accurately differentiate LD students from non-LD students and, furthermore, excludes
a number of students who need services from those services because they do not qualify with a significant enough discrepancy. Gresham and Witt (1997) echo these criticisms and assert that IQ testing is not a cost effective use of time, time that could be better spent assessing specific skill deficits and designing interventions to address those deficits.

All of these authors (Aaron, 1997; Bray et al., 1998; Fuchs, Mock, et al., 2003; Gresham & Witt, 1997) also point out that IQ testing does not provide information that is useful for the design of instructional interventions. Lastly, they point out that the “wait to fail” approach is problematic. Quite frequently students struggle for a number of years prior to developing a discrepancy between performance and ability that is significant enough to qualify for special education services. Therefore, even those student who qualify for special education services receive the services at a point where remediation is more difficult, if not impossible (Juel, 1988). As research has shown, remediation for reading disabilities needs to occur early in a child’s career to be effective (Hart & Risley, 1995; Juel, 1988; NRP, 2000; Stanovich, 1986).

In response to the growing body of research and experience that does not support the use of the discrepancy model for determining qualification for special education a new model, Response to Intervention (RTI) has evolved. RTI involves the following essential components: all students being provided with an evidence based curriculum using effective instructional practices, monitoring of all students’ progress (periodic benchmark assessments), implementing supplementary instruction in the general education setting for those who are not responding to the curriculum, monitoring the progress of those students more closely over time, and providing
special education services to those students who fail to respond to intervention (Batsche et al., 2005; Fuchs & Fuchs, 2001; Fuchs, Mock, et al., 2003). This model of evaluating and educating children with difficulties has several advantages over the discrepancy model (Batsche et al., 2005; Fuchs, Mock, et al., 2003). RTI involves monitoring progress early in all children’s’ educational careers, includes interventions that focus on skill deficits, provides ongoing, formative assessment of progress to evaluate the effectiveness of interventions and provides targeted interventions that can assist in the planning of special education services. There is also a strong emphasis on using research based practices both in the general education process and in the intervention and assessment process. The benefits of RTI as an assessment and intervention approach seem to be well aligned with the findings of research in effective reading instruction.

There are two models for the implementation of RTI that are currently being developed and assessed. One of these is the Standard-Protocol model which involves providing a specific, research supported small group or individual intervention to students who are performing poorly. Those that do not respond to the intervention are deemed non-responsive to intervention and eligible for more intensive services. A number of studies have been conducted regarding this approach with some promising results (D. Fuchs, L. S. Fuchs, & Compton, 2004; Torgesen, Alexander, Wagner, Rashotte, Voeller, & Conway, 2001; Vaughn, Linan-Thompson, & Hickman, 2003; Vellutino, Scanlon, Small, & Fanuele, 2006). Generally, the research shows that students with more intensive needs are identified and those students who only need some additional support to progress adequately are given that support.
The second model for the implementation of RTI is the Problem-Solving model. This model is more individually focused whereby a given child’s personal skill deficit is identified, an intervention is developed within the general education classroom to address that need and the student’s progress is monitored formatively. Generally, exhibiting a significant discrepancy from their peers within the classroom or school identifies students as needing targeted services. Response to intervention is determined by the use of a dual discrepancy approach (Batsche et al., 2005; Fuchs, 2003; Speece, Case, & Molloy, 2003) whereby the student’s progress is measured by both a target level of skill and a target rate of improvement. The goal is for the student to achieve the same level of skill as his or her peers and to do so at a rate greater than the rate of improvement of the peers. If the student does not meet the level improvement and rate of improvement targets, he or she is deemed eligible for more intensive services. The research in the use of this model is less developed than for the Standard-Protocol method though some promising work has been done in the state of Iowa in implementing this model state wide (Grimes, Kurns, & Tilly, 2006; Ideka, Grimes, Tilly, Allison, Kurns, & Stumme, 2002).

Both of the models of RTI are in their initial phases of development. They have been tried in various settings (Fuchs, Mock, et al., 2003) with some success. However, Fuchs, Mock, et al. (2003) suggest that more research needs to be conducted in order to assess which components are critical and how this promising approach to intervention and assessment can be utilized. It is clear, however, that RTI has significant implications for both how special education eligibility is determined and for the process of providing instruction in the general education setting (Batsche et al.,
RTI is a systemic change, which will require that empirically supported curricula, instructional methods and intervention approaches be used to address the needs of students. In addition, formative assessment tools need to be used to assess students regarding skill deficits and to monitor their progress once interventions have been implemented. Given the high stakes nature of the decisions being made it is also important that these tools are standardized and have empirical support for their use.

Curriculum Based Measurement and Formative Assessment

Curriculum Based Measurement (CBM) tools are a form of assessment that meet the criteria outlined by research, legislation and the parameters of RTI. The concept of CBM was initially developed by Stan Deno in the 1970’s and early 1980’s (Deno, 1985). CBM tools can be used for both formative and summative assessment and have been researched extensively to assure their standardization. In developing CBM tools, researchers have focused on creating what are referred to as *general outcome measures* (GOM). That is, the research focused on finding curriculum-based activities that provide an indication of how well a student is doing with global skills that accurately reflect overall ability and progress in one general area (Fuchs & Deno, 1991). For example, Reading-Curriculum Based Measurement (R-CBM) has been demonstrated to be a global indicator of overall reading progress. While previous efforts at skill measurement had focused on mastery measures that focused on student progress with a specific sub-skill in a hierarchy of skills being taught (e.g. learning the short “a” sound, then the short “e” sound, etc.), CBM tools have been purposely developed to have the following qualities: 1. They represent activities associated with the curriculum that students are learning in school, 2. They are of short duration so
educators could use them frequently, 3. They are available in multiple forms for repeated and regular administration, 4. They are inexpensive to produce both in terms of cost and time and, 5. They had to be sensitive to change over time (Shinn, 1989; Shinn, Nolet, Knutson, Thomas, & Grimes, 1990). Lastly, CBM measures were designed to be administered in a standardized fashion so that comparisons could be made between children and over time with reliability and validity comparable to summative achievement tests (Martson, 1989).

CBM tools have been developed for early literacy skills, reading, written expression, spelling, and mathematical computation, concepts, and application. All of the tools involve the student engaging in activities derived from and related to the curriculum they are learning in school. The largest body of research has been with the reading tools as reading is a foundational skill for all students. Reading-Curriculum Based Measurement (R-CBM) is used as the GOM for reading assessment from grade one on. Students are required to read a grade level passage for one minute and scored on the number of words correct per minute. As noted above, R-CBM has been found to be a very valid measure of reading fluency and skill and of comprehension as well as being correlated with other reading measures (Martson, 1989; Tindal, Germann, & Deno, 1983). R-CBM is reliable as a repeated measure and sensitively indicates change in reading skills (Hintze & Shapiro, 1997; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006).

While R-CBM is a useful GOM for the second and higher grades, tools have also been developed to assess early literacy skills in an effort to provide for earlier intervention. The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is one
version of early literacy assessment that has research support and is widely used (Good, Gruba, Kaminski, Thomas, & Grimes, 2002; Kaminski & Good, 1998). The DIBELS measurements consist of Letter Naming Fluency (LNF), Initial Sound Fluency (ISF), Phoneme Segmentation Fluency (PSF), and Nonsense Word Fluency (NWF). Both ISF and PSF are measures of phonological awareness. The ISF task involves selecting a picture from four that are presented that begins with a particular initial sound and producing the initial sound for one of those pictures. The PSF task involves segmenting a series of words into individual sounds. The NWF task is a measure of phonetic skills that includes knowing letter sound associations and blending those sounds into words. It involves looking at a series of Consonant-Vowel-Consonant (CVC) nonsense words and either identifying the sounds for each of the letters or reading the whole word. The LNF task is a task of letter identification whereby students are asked to label a series of letter. It is a task that has been shown to be highly predictive of later reading skill. These tasks are all considered to be GOMs for pre-reading skills. As such, they provide information regarding a student’s early reading abilities and whether those abilities are developing as they should.

These reading CBM measures are all very effective for evaluating students’ general progress in reading. They allow for a summative evaluation at any given point in time and for a normative comparison to peers. CBM measures can also be used formatively to assess students’ overall progress in developing necessary reading skills and to modify interventions if progress is not being made. These measures, however, do not provide specific information regarding particular skills a teacher should focus on for instruction (Fuchs & Deno, 1991). As Hintze, Christ and Methe (2006) indicate,
this next level of assessment for developing instruction involves a mastery measurement model. The weakness of the GOM approach adopted by the developers of CBM measures is that GOM measures do not provide specific information regarding the focus for intervention (Fuchs & Deno, 1991; Fuchs, Fuchs, Hosp, & Hamlett, 2003; Hintze et al., 2006). As explained by Hintze et al. (2006) and Fuchs, Hosp, et al. (2003), specific skill mastery measurement tools could provide this level of specific information to be utilized in developing targeted interventions. Furthermore, specific skill mastery measurement tools could be designed to be used for progress monitoring of the intervention and the acquisition of the specific skill. When utilized in conjunction with CBM general outcome tools, mastery measurement tools could provide useful information regarding both the design of interventions and the progress that students are making in the particular skill. One problem with mastery measurement tools is that in the past, they have not been standardized either in the gathering of information or in the analysis of that information (Fuchs & Deno, 1991). Currently there are efforts under way to develop assessment tools that allow for a standardized approach to sub-skill assessment and that can be used to monitor progress in a fashion similar to that provided by CBM.

*The Nature of Specific Sub-skill Mastery Measurement*

The nature of specific sub-skill mastery measurement (SSMM) is that it focuses on assessing a specific sub-skill once a weakness has been assessed using a GOM. Generally, a survey level assessment of a student’s skills is first conducted using a CBM tool. Then, once a weakness has been identified, SSMM tools are used to assess the specific sub-skill that contributes to the weakness and that must be
addressed in order for the student to develop the general skill, i.e. decoding of particular diphthongs or blending of discreet sounds in words. The assumption is that this sub-skill is necessary in order for the student to develop the general skill (Hintze et al., 2006). Interventions can then be developed to target the specific issue and help the student master that sub-skill.

The advantage of SSMM tools is that they provide teachers with information regarding the specific deficit that needs to be addressed. They also allow for progress monitoring of the learning of that sub-skill. When used in conjunction with ongoing CBM progress monitoring, one can assess whether the student is gaining proficiency in the sub-skill and whether that increased proficiency is impacting the general skill. As noted previously, fluency is a critical foundational component of learning to read (NRP, 2000) and it is the component measured by the R-CBM reading measure. Once R-CBM has identified a weakness in fluency the next question is one of the nature of that weakness and how to intervene to improve the student’s reading skills. R-CBM primarily provides information regarding fluency and that may well be the issue for the particular child. While NWF provides general information regarding a potential phonetic weakness, it does not provide specific information regarding the particulars of a decoding weakness. NWF also focuses largely on basic letter-sound skills and CVC word decoding. It does not assess decoding difficulties with more complex words that contain vowel and consonant combinations. Efforts to assess decoding issues began in the late 1990’s and continue today.
Efforts to use CBM to Assess and Target Specific Skills

During a CBM retreat in 1999 “Stan Deno, Chris Espin, Joe Jenkins, and Lynn Fuchs have agreed that developing and assessing the value of a CBM diagnostic analysis represented a pressing research agenda for CBM.” (Fuchs, Fuchs, et al., 2003, p. 15). In addition, two recent reviews of the literature (Stecker, Fuchs, & Fuchs, 2005; Wayman, Wallace, Wiley, Tichaj, & Espin, 2007) on CBM arrived at a similar conclusion regarding the field of CBM assessment. Both reviews called for the development of standardized, reliable, valid CBM tools to assess specific skill weaknesses. Stecker et al. (2005) noted that more research has been completed in the area of specific skills assessment in the area of math than in the area of reading. The authors also noted that teachers who received detailed information regarding student weaknesses were better able to modify instruction and support student learning than teachers who did not have the specific information (Stecker et al., 2005, p. 802). Wayman et al. (2007) also point out that the purpose of CBM has changed with the introduction of RTI and CBM is being used to make high stakes decisions regarding the provision of special education services. They state that this creates a greater need for assuring the validity of these instruments and for making sure they assess needs in a way that allows for effective intervention (Wayman et al., 2007). These concerns have led to the development of at least one specific skills mastery measurement tool based on CBM principles.

That effort to extend CBM approaches to specific sub-skills measures involves the work of Fuchs, Fuchs, et al. (2003). In this study the authors attempted to develop CBM cut points that could differentiate between students who needed decoding,
fluency, or comprehension instruction and to develop follow up assessments that would provide a diagnostic analysis for decoding and comprehension strengths and weaknesses (Fuchs, Fuchs, et al., 2003). The authors developed a computerized assessment that sought to meet these goals. They used word reading fluency in combination with oral reading fluency and comprehension questions in this assessment. Unfortunately, the reliability and validity data for the comprehension component of the assessment was weak and that was not included in the diagnostic system. The utility of the word-reading component of the assessment was supported by the research findings. The authors were, however, unable to develop reliable cut points for using R-CBM to differentiate student needs for types of instruction (Fuchs, Fuchs, et al., 2003). The effort to combine the use of oral reading fluency with specific word reading measures seemed to be promising as an initial step towards developing a specific skills mastery measurement approach to reading assessment (Fuchs, Fuchs, et al., 2003).

The Development of the iRead Inventory

At the University of Massachusetts-Amherst there is currently a project under way to develop a comprehensive, psychometrically sound SSMM assessment tool. Given that decoding is so central to fluency, the researchers at UMass decided to begin the process of developing a specific sub-skill mastery measurement tool by focusing on assessing decoding skill, with a focus on more complex decoding skills. The Individual Reading Evaluation and Diagnostic (iRead) Inventory (Koerner, McGurl, Farrell-Meier, & Hintze, 2006) is being developed as a comprehensive, standardized, research supported SSMM tool. The researchers began with a focus on the assessment
of the sub-skills central to effective decoding. This process has taken several steps and
the development of the tool continues. The initial step for the UMass researchers was
to develop a data base of specific decoding sub-skills that are taught in grades
kindergarten to fourth grade. This began with an extensive content analysis of four
empirically supported reading curricula (i.e., Scott Foresman, Houghton Mifflin,
SRA/McGraw-Hill, and Harcourt) to identify the specific sub-skills that were taught at
each grade. This provided a measurement net of decoding sub-skills that can be used
in developing the tools and assessing students. The next step was to develop lists of
words that correspond to the specific skill objectives for each grade level. The choice
to focus on reading from specific word lists was supported by early research in the
development of CBM (Deno, Mirkin, & Chiang, 1982) and by more recent studies
regarding assessing decoding skills (L. S. Fuchs, D. Fuchs, & Compton, 2004; Fuchs,
Fuchs, et al., 2003). This process resulted in the development of an item bank of skill
specific words that can be used to randomly generate word lists in multiple forms to be
used for the assessment and progress monitoring process. These word banks were then
utilized to develop an assessment tool that involves reading words in each skill area.
The word banks can also be used to develop progress monitoring probes in specific
skills once instruction is initiated to address that specific weakness. The next step is to
research the effectiveness of the words for assessing skill deficits and to evaluate
whether the assessment can provide teachers with skill specific information that can be
used to provide targeted instruction.
Current Study

The purpose of this study was to assess the effectiveness of the iRead Inventory for accurately assessing specific decoding sub-skill weaknesses and for informing the development of targeted interventions that increase the rate of student improvement in reading. One hypothesis was that the iRead Inventory would distinguish specific decoding sub-skill weaknesses for those students who struggle with decoding. The second hypothesis was that using the iRead Inventory in the development of interventions to target specific decoding sub-skills would lead to a greater rate of reading improvement for participants once the interventions were initiated than they showed prior to the targeted decoding instructions being provided.

This study had both a screening and an implementation phase. Results from the screening phase were used to evaluate hypothesis one that the iRead Inventory can identify specific decoding deficits and to identify participants for the intervention phase. The intervention phase focused on evaluating the second hypothesis that the iRead Inventory can be used to design interventions that lead to reading improvement. During the screening phase, second grade students were assessed for difficulties with fluency and accuracy using fall benchmark scores on R-CBM. All participants who were dysfluent and inaccurate (i.e. below the 25th percentile and greater than 3 errors) in the fall benchmark assessments were then screened using a research version of the iRead Inventory. The data from this screening and the R-CBM words read and error scores was compared to assess the utility of the iRead inventory for identifying decoding difficulties. The screening process also involved using NWF and PSF measures with NWF scores being utilized to further evaluate potential decoding
weaknesses and PSF scores being used to rule out phonemic awareness weaknesses. Those students who were identified as having primarily decoding issues were identified as eligible for the intervention phase of the study.

The intervention phase of the study involved using a single participant, multiple baseline, randomization design with three participants who were identified as having specific decoding deficits. Each participant was provided, in a sequential fashion, with targeted direct instruction for those deficits while reading development progress was monitored. To select three participants for the intervention phase, all students identified as having decoding difficulties were ranked and the three students with the most significant decoding issues were selected for the intervention phase. The information gathered from the iRead Inventory part of the screening process was used to identify the specific decoding skills for which they were provided direct instruction. Direct instruction following a specific protocol was developed to target the needs of each of the students. All three students were progress monitored twice weekly using R-CBM and NWF throughout the experiment as well as having their progress with learning the specific sub-skills monitored daily using the iRead Inventory. The randomization design involved beginning intervention phases for each of the students at random points after establishing a baseline for all three students. The points were selected to allow for at least three weeks of interventions prior to ending the experiment and the start points were selected randomly to allow for statistical analysis of change between baseline and intervention phases without relying on response guided decision models which suffer from internal validity issues (Todman & Dugard, 2001, p. 17).
The first hypothesis was tested by the ability of the iRead measure to reliably assess specific decoding deficits for each child. This was done by comparing the R-CBM words read and accuracy results for the dysfluent students to the iRead Inventory results. The second hypothesis was tested by visual and statistical analysis of the change in reading improvement rate in R-CBM, and NWF, and iRead Inventory progress monitoring probes for the students as they received the targeted intervention developed using the iRead Inventory. As the intervention was implemented a significant change in reading skill for each child, both on the GOM measures (NWF and R-CBM) and the SSMM measure (iRead Inventory), was anticipated to support the hypothesis that the assessment and intervention were responsible for the improvement. In addition, the iRead Inventory was used as a pre-post measure with each student being assessed at the end of the study using the full iRead Inventory to measure improvement in sub-skills. Lastly, the number of skills learned was monitored as well and that information utilized to conduct a qualitative analysis of the iRead Inventory intervention design process.
CHAPTER II
METHODS

Participants and Setting

The participants were three students from a second grade in a rural elementary school in Central Massachusetts. The students in this school were 93.6% Caucasian and 23% received free or reduced lunch. Given grade level makeup, all of the participants selected were Caucasian. The second grade has two classrooms that use the same curriculum and approach to instruction for reading. In addition, the two teachers use flexible grouping for reading class whereby similar students of similar skills are grouped during the second half of reading class. The participants were all identified for Tier 2 strategic intervention based on the Fall benchmark screening assessment results. Those students below the fall benchmark screening target score received small group guided reading instruction from both their general education teacher and the Title 1 teacher as the secondary intervention. All of the participants were enrolled in Title 1 and were receiving similar intervention in the general education and Title 1 settings, namely additional guided reading time. Thus, they were expected to show comparable improvement in reading prior to intervention. Students receiving intensive services in special education were not eligible for the study as the interventions they received were markedly different from those received by the other students. The screening phase of the study involved additional assessment of all of the dysfluent students in second grade. Phase two involved three participants being provided with a more specific, focused direct instruction intervention. This intervention occurred daily for approximately 15 minutes in a quiet office near their
classroom. The parents of participants in the intervention phase of the study were notified of their eligibility for the project via a letter and given the opportunity to withhold or provide consent to their child’s participation (Appendix A). Once the selection process was completed, the three participants selected ended up being from one classroom. This assures even greater consistency of their instruction given they had one teacher for general education and were in the same group for Title One.

Materials/Tools

For the screening phase, the initial screening tool was the Fall benchmark results using the AIMSweb Reading Curriculum Based Measurement (R-CBM) probes as part of a school wide reading improvement process (Shinn, 1989; Simmons et al., 2002). R-CBM involves the participant reading a passage calibrated to their grade level for one minute. The total words read correctly are counted and used as the primary measure of reading fluency. In earlier studies, the test-retest reliability of R-CBM measures range from .82 to .97 and alternate-form reliability coefficients range from .84 to .96 (Marston & Magnusson, 1985; Tindal et al., 1983). Both forms of reliability had most correlations above .90 and inter-observer agreement coefficients of .99 (Tindal et al., 1983). The AIMSweb R-CBM passages had test-retest reliability scores ranging from .80 to .90 and alternate-form reliability scores ranging from .79 to .90 (Howe & Shinn, 2002). Correlations of the AIMSweb passages within and across grades were between .78 and .98 with a median correlation coefficient of .90 (Howe & Shinn, 2002). With respect to validity, an initial study by Deno, Mirkin and Chiang (1982) found correlations between word reading and oral reading of passages ranging
from .73 to .91, with most coefficients above .80. R-CBM is, therefore, a reliable and valid tool for screening participants for reading difficulties.

A subsequent screening step involved using AIMSweb Nonsense Word Fluency (NWF) and Phoneme Segmentation Fluency (PSF) probes to select for participants whose primary difficulty is with decoding and not phonemic awareness. NWF is a measure of the alphabetic principle, namely the ability to engage in letter-sound correspondence and in blending letters into simple words. Students are asked to read from a list of consonant-vowel-consonant (CVC) and vowel-consonant (VC) nonsense words for a period of one minute. They are asked to either say each individual sound or blend the sounds into words. Alternate form reliability coefficient for the use of NWF with kindergarteners is .83 (.67-.88) (Good, Wallin, Simmons, Kame’enui, & Kaminski, 2002). Concurrent validity coefficients with the Woodcock Johnson Readiness Cluster are .51 and predictive validity coefficients with measures of R-CBM at the end of first and second grade range from .73 to .77 (Good, Wallin et al., 2002)

Phoneme Segmentation Fluency is a measure of the student’s phonological awareness. PSF involves the student listening to a word and being asked to verbally reproduce the three or four phoneme segments of the word. The alternate form reliability coefficient for PSF is .74 (.66 - .79) for kindergartners and .67 (.60-.70) for first graders (Good, Wallin et al., 2002). One year predictive validity coefficients for PSF with R-CBM and NWF yield median coefficients from .40 to .52 (Good, Wallin et al., 2002).
The next step in the intervention phase involved evaluating each of the participants using the decoding assessment tool of the iRead Inventory. The inventory is still in the research and development stage but lists of words have been developed that assess all of the vowel and consonant decoding sub-skills that are taught by the four curricula in the kindergarten to second grades. The focus for this study was on vowel combinations as most curricula and intervention programs focus first on vowels. In addition, the short time period for the study precluded teaching both vowel and consonant combinations. The initial assessment involved five words being selected randomly from each of the vowel related sub-skills and being listed on pages for the participants to read. Participants were asked to read the words following a set protocol (see Appendices B and C). As there were a large number of words, participants were asked to read 25 words at a time on a flip chart. The words were listed with specific sub-skills in columns to minimize the possibility of participants perceiving a pattern and using that knowledge to score well. Participants were given the opportunity to take a short break between each set of words. Once this broad assessment was complete, errors on the five words for each sub-skill were evaluated. If participants could read 4 or 5 words correctly in a given sub-skill they were assumed to have mastered that skill. If participants could read only 0 or 1 of the words of a specific sub-skill type, it was assumed they were struggling with that skill. If 2 or 3 words were read correctly, the participants were given an additional five words from that sub-skill list to read. (see Appendices D and E). A score below 80% correct (7 words or less correct) was then used to define a weakness in that particular sub-skill. The ability of this measure to differentiate between specific sub-skill weaknesses was
assessed by a review of how many specific skills were identified for the development of the intervention. The tool was also compared to R-CBM words correct and accuracy results to assess the utility of the iRead Inventory for identifying decoding weaknesses.

**Procedures**

*Design.* The screening phase of this investigation involved all of the students who scored low on the R-CBM benchmark assessment in the Fall. Those students all received further evaluation with the NWF and PSF assessment as well as the iRead Inventory. The results of this screening process were used both to assess the utility of the iRead Inventory and to screen participants for the second phase of the study. Those participants who are identified as having decoding difficulties, as evidenced by inaccuracy on the R-CBM assessment (4 or more words read incorrectly) and by decoding weaknesses on the iRead (5 or more specific decoding sub-skill weaknesses identified) and NWF (a score below the 25th percentile score for the beginning of second grade or more than 3 errors), were selected for the second phase of the study. To do so, scores were rank ordered and the three participants with the weakest decoding skills were eligible for the intervention phase of the study. The decoding weaknesses identified by the iRead Inventory were utilized to design the intervention by providing a focus for the direct instruction.

The intervention phase of the investigation involved a multiple baseline, randomization design (Todman & Dugard, 2001) across individuals with three participants from second grade involved in the treatment phase. Todman and Dugard (2001) advocate the use of randomization designs in single case research as an adjunct
to visual analysis. A randomization design involves setting a specific number of data collection points for each participant, establishing a minimum baseline, defining a minimum number of progress monitoring points for each intervention phase, and then randomly selecting the point for initiating each participants’ intervention phase. For example, in this study, the participants were progress monitored 24 times, the minimum baseline points were 4 and the minimum intervention points were 6. Therefore, the three points to begin interventions were selected between the 5th and 18th progress monitoring points. The intervention phases for each participant started on the day of the 6th, 12th and 17th progress monitoring sessions. The data were then graphed for visual inspection but also were analyzed statistically to determine if a change occurred in the intervention phases.

Todman and Dugard (2001) advocate this approach for two reasons. First, using a response-guided approach to determining when to initiate interventions has been shown to be biased. Second, data can be analyzed statistically by comparing the actual data of the selected intervention points to a distribution developed by repeatedly selecting multiple possible combinations of intervention points from the actual data. The hypothesis being that if the intervention had an effect the result would differ from the distribution created by selecting from random possible intervention points (Todman & Dugard, 2001). The probability distribution based on randomly assigned intervention points is constructed by repeatedly resampling from the data, with replacement, a series of results that are consistent with the experimental design (i.e. repeatedly selecting data that reflects three possible intervention points between the 5th and 18th progress monitoring sessions). The repeated selection occurs for a minimum
of 2000 iterations and the statistic of interest (e.g., difference between means of the baseline and intervention phases) is calculated multiple times to generate the distribution. Once the distribution is developed based on the assumption of no intervention effect, the actual observed distribution is statistically compared to this distribution. If the observed results of the actual intervention points are significantly discrepant from the probability distribution based on randomly selected intervention points, the assumption is that the intervention had an effect beyond chance occurrence (Todman & Dugard, 2001). The advantage of this approach is that one does not need to assume that the sample reflects a hypothetical population that is normally distributed; instead one is assuming that the observed data are the best estimate of the underlying distribution of the population (Todman & Dugard, 2001). This allows for the comparison of observed results to the exact probability distribution based specifically on the data from that sample. It is not necessary to assume that the population is normally distributed nor is it necessary to be concerned about whether the sampling method is actually reflecting that population (Todman & Dugard, 2001, p. 31). Critical to this data analysis approach is the process of setting up the experiment so it can be analyzed and conducting the analysis to parallel the experimental design.

The experimental design for this study was developed to meet the requirements for a randomization design. The study consisted of 12 weeks of bi-weekly progress monitoring for the dependent variables of interest, i.e. R-CBM and NWF, yielding 24 data points for graphing and statistical analysis. During the intervention phase a baseline of at least two weeks (four progress monitoring sessions) was obtained prior
to initiating any intervention in order to assure sufficient quantity and stability of data for analysis (Todman & Dugard, 2001). Interventions were then provided on a schedule randomly selected for each of the three participants with the limitation that at least three weeks (six progress monitoring sessions) of intervention were provided for each child. Intervention start points for each child were randomly selected to begin somewhere between the 5th and 18th progress monitoring sessions (between the third and ninth weeks of the study). For example, one participant may have begun the intervention after the 6th progress monitoring session, one after the 9th progress monitoring session and another after the 11th progress monitoring session. Subsequent to the randomly selected intervention start point, each participant’s performance on R-CBM and NWF was progress monitored and the mean differences with and without intervention were compared statistically.

In addition, the full iRead inventory was administered to each participant at the end of the study and compared to decoding skills pre-intervention. Lastly, the number of sub-skills in which participants exhibited weakness at the beginning of the study (those skills for which they could decode less than 80% of the words presented) were compared to the number of sub-skills they improved on during the study (skills which previously had been weaknesses and they were now able to read 80% or more of the words presented). These data were used to develop a qualitative analysis of the iRead Inventory intervention design process (i.e. how many skills are they able to learn on average over the course of the study with this approach?)

*Screening.* The initial screening focused on R-CBM results. Participants were eligible for the screening phase of the study if they scored below the AIMSweb
established twenty-fifth percentile, fall benchmark of 28 WCPM for second grade (AIMSweb, 2007b). The error rates for those participants were assessed as well. Those participants with greater than 3 errors were classified as dysfluent/inaccurate while the participants with 3 or less errors will be classified as dysfluent/accurate. All of the participants who scored below the 25th percentile score were also assessed using the iRead Inventory. The number of decoding difficulties (skills for which the participant read correctly less than 80% of the words provided) were identified through that assessment. In addition, participants were assessed using the AIMSweb NWF and PSF assessments. Eligible participants needed to have a score below the beginning of second grade 25th percentile score of 35 for NWF or an error rate of greater than 3 sounds incorrect on NWF and above the end of first grade 25th percentile score of 41 for PSF (AIMSweb, 2007a). The purpose of this screening step was to identify those participants for whom decoding was the primary issue and for whom phonological awareness skills were at or above target for their age.

The results of the R-CBM, NWF and iRead assessments were then used to identify participants for intervention and prioritize them for inclusion in the study. Participants were eligible if they (1) scored below the Fall 25th percentile on R-CBM, (2) had more than 3 errors in R-CBM, (3) scored below the Fall second grade 25th percentile or had greater than 3 SC errors on NWF, and (4) exhibited 5 or more decoding weaknesses on the iRead Inventory. Eligible participants were then prioritized based on R-CBM and NWF scores, numbers of R-CBM errors and number of decoding difficulties identified in the iRead assessment. The parents of the three participants with the most significant decoding difficulties were contacted for
permission to participate in the intervention phase of the study at this point (see Appendix A). The participants who qualified were all males. All of the parents of the three weakest participants agreed to their participation in the study. In addition, a fourth participant was identified and included in the study to account for possible attrition. This decision proved to be fortuitous as one of the participants was withdrawn part way through the study. The results of this screening are summarized in Table 1.

Table 1

*R-CBM, iRead, NWF, and PSF Screening Results*

<table>
<thead>
<tr>
<th>Part. #</th>
<th>R-CBM errors</th>
<th>error rate</th>
<th>iRead (difficulties /45)</th>
<th>NWF errors</th>
<th>error rate</th>
<th>PSF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>5</td>
<td>24%</td>
<td>34</td>
<td>56</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>7</td>
<td>30%</td>
<td>22</td>
<td>44</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>7</td>
<td>27%</td>
<td>28</td>
<td>65</td>
<td>4</td>
</tr>
</tbody>
</table>

*Progress Monitoring and Data Collection.* Once the screening phase was complete and the intervention phase had begun, bi-weekly progress monitoring was initiated using AIMSweb NWF and R-CBM progress monitoring probes. Progress monitoring continued throughout the study and was completed by the primary investigator and a school psychology graduate student. Although the reliability and standardization of the R-CBM and NWF measurement process should minimize lag-1 autocorrelation, a procedure advocated by Todman and Dugard (2001, p. 36) was also
adopted whereby the progress monitoring data collection was alternated to minimize the potential of serial dependency. As such, the scores from the bi-weekly NWF and R-CBM assessments served as dependent variables that were graphed for visual analysis and analyzed statistically to determine whether there was a change between the baseline and intervention phases for the group and for each participant.

In addition, as part of the intervention outlined below, each participant was progress monitored using iRead Inventory word lists targeted to the specific sub-skill they were being instructed on in the intervention. These word lists were used to monitor a participant’s mastery of a given specific sub-skill and to plan for when instruction in the next sub-skill should begin. The total number of sub-skills that each participant mastered during each intervention session as well as during the course of the study was collected. Lastly, post-test performance on the iRead Inventory was compared to pre-test performance to determine if there was a significant change in specific decoding abilities over the course of the intervention.

*Intervention Procedures.* The initial iRead, NWF and PSF screening assessments were administered by the primary investigator and a school psychology graduate student. Training in the iRead inventory was provided prior to that assessment process. Training in the intervention was provided by the investigator prior to the implementation of the lessons. The emphasis was on the specifics of the intervention steps and the importance of using a consistent protocol. The training focused on the nature of direct instruction of reading, the structure and focus for the lesson plans, and the critical behaviors of signaling, modeling, pacing, monitoring and correcting mistakes (Carnine, Silbert, Kame’enui, & Tarver, 2004, pp. 64-68). The
intervention was varied based on the specific needs of each participant. Training focused on the types of lessons the interventionists needed to provide and the process of administering the iRead progress monitoring list. The lessons varied depending on which specific vowel sound, a vowel or a vowel-consonant-e (VCe) pattern the participant needed to learn (see Appendix F for examples of the lesson structure and the three types of lessons).

Interventions were conducted in a quiet room separate from the participants’ classroom and lasted 15 minutes per session. This daily sub-skill mastery intervention utilized the direct instruction approach outlined in Direct Instruction Reading (Carnine et al., 2004). A school psychology graduate student and an aide trained by the researcher provided the daily interventions for all of the subjects. In addition, the special education teacher provided some lessons on those rare occasions when both of the interveners were absent. Each of the daily 15-minute intervention sessions followed a lesson plan developed by the researcher based on the Carnine et al. (2004) guidelines and the specific needs identified by the iRead Inventory (see Appendix D). Interventions focused on learning vowel sound, vowel combination, or VCe skills (Carnine et al., 2004, pp. 62-64, 154-159). Daily lesson plans were developed by the investigator in consultation with the special education teacher to assure consistency and clarity for the implementers. The lessons followed a consistent format with the primary variation being the specific sub-skill being addressed. If a participant required assistance in more than one vowel sub-skill, the ordering of those sub-skills was based on recommendations made by Carnine et al (pp. 152-154) and on the order of the skills as outlined in the iRead as that order was determined based on a review of the
curricula and several intervention programs. Each lesson involved an initial assessment of the specific sub-skill taught during the previous day’s lesson using the iRead Inventory. If the participant successfully read 18, 19 or 20 of the words presented in that skill area on two consecutive days, the lesson was changed to focus on another sub-skill for the following day. If the participant failed to read greater than 17 words correctly on two consecutive days, the interventionist continued to provide lessons in that sub-skill. It should be noted that lessons included judicious review of previously learned skills by including them in the lessons that focused on a new sub-skill. So, for example, once a participant had mastered the short e sound, that sound was included in the process of teaching the next skill, e.g. short u. Specific interventions were tailored to each participant and were developed based on their progress in learning the sub-skill. The skills taught to each subject are outlined in Appendix F with an indication of the number of lessons provided per skill.

_Treatment Integrity and Inter-observer Agreement._ The review of lesson plans by the special education teacher was a check on whether the parameters for instruction, developed by Carnine et al (2004), were being outlined clearly enough so they could be followed with integrity. Twenty-six percent of the intervention sessions were videotaped at random to assess treatment consistency and integrity. In addition, an intervention report was completed after each session regarding the completion of the intervention and the participant’s progress with the task to remind interveners of the expectations and procedures (see Appendix E). Periodic reviews of those sessions occurred whereby the primary investigator observed the tape, and completed an intervention report based on that observation. The investigator and interventionist then
reviewed the intervention reports and discussed the fidelity of the intervention. The agreement between intervention reports (self completed and videotape observation completed) was used as a measure of intervention fidelity. Point-by-point agreement percentages were used to analyze the differences, if any, between the integrity checklists. The mean for the integrity checks was 80% (N=19, Range = 56% - 100%, SD = 13%). It was noted that percentage of agreement increased over time with follow up discussions, i.e. the lowest integrity percentage occurred in the second review while the highest percentage occurred in the second to last review with 94% agreement occurring the three other times in the last four reviews. An analysis of the integrity results also indicated that the most frequent discrepancies were on items 5 and 12 which involved pointing to letters and waiting for 2-seconds for a response and on items 6 and 13 which involved using the out-in motion and sweeping the letters. Often the interventionist did not wait for 2-seconds because the participant responded before 2-seconds had passed. Therefore, the video rater frequently rated these items one point lower than the intervener. It was concluded that this discrepancy reflected participant response time rather than intervener fidelity. With respect to the out-in motion and sweeping the words, the interventionists were using a piece of paper on a table rather than words on a black board as delineated in the direct instruction directions (Carnine et al., 2004, p. 64). Interventionists tended to point at the word or letter combination so their hand did not block the participant’s vision. Again, this modification was not deemed as being significant to the integrity of the intervention and may have, in fact, facilitated the process of learning.
In addition, twenty-six percent of the progress monitoring sessions were audiotaped and scored by an independent observer familiar with the administration of the R-CBM and NWF passages. Point-by-point agreement between the scores of the actual progress monitoring and the audiotape review were used for measuring progress monitoring fidelity. Each word was compared in a point-by-point manner and the total number of words on which both raters agreed was divided by the total of the number of words on which they disagreed and agreed and multiplied times 100%. The inter-rater agreement for R-CBM was 97% (N = 18, range = 89% - 100%, SD = 3%) and for NWF was 94% (N = 18, range = 87% - 100%, SD = 3%).
CHAPTER III
RESULTS

Data Screening

The study occurred over a twelve-week period with participant performance being monitored twice weekly with R-CBM and NWF probes. As noted above, one of the four participants was withdrawn from the study due to parent request because mother felt being removed from the class for the study was too much for him at the time. Given random assignment of intervention phases, this participant was the second student to begin receiving the intervention (at the 9th progress monitoring point). Therefore this participant’s data could not be used and the last two participants began the intervention phase later (at the 12th and 17th progress monitoring points). As discussed later, the brevity of the intervention for these two participants may have affected the results.

In addition, in order to analyze the data for the randomization design it was necessary to only use data collection days for which data were collected for all three participants. On three of the data collection days one of the participants was absent and data from those days were not used. In addition, on one of the data collection days the data collector used the wrong type of probe and thus those results could not be used. This resulted in twenty usable data points spread evenly over the course of the study and included all phases. Missing data days did not occur twice in any week nor did they occur consecutively across weeks. Therefore, at least one datum point was collected for each week of the study with four weeks of only one data point and eight weeks of both data points being collected.
Summative Analysis

Pre- and post-intervention data were initially reviewed to analyze summative improvements in R-CBM, NWF and iRead sub-skills. The results are summarized in Tables 2, 3, and 4. The R-CBM and NWF data involved administering three probes to each participant and using the median score for WRC, SC, and errors. These assessments were conducted as part of the Fall and Winter benchmarking process and occurred 15-weeks apart from each other. The iRead data were collected for each participant during a single assessment session following the experimental administration guidelines. Total number of sub-skills mastered and identified as weaknesses were counted. The percentage of mastery of sub-skills taught was calculated as each participant was taught a different number of sub-skills. This provided a score for the number of sub-skills in which each participant exhibited improvement. In addition, some sub-skills that were not directly taught showed improvement as well and those were counted. A detailed summary of the iRead specific skills assessment results for each participant is provided in Appendix F.

Table 2

Pre-Intervention and Post-Intervention NWF results

<table>
<thead>
<tr>
<th>Part. #</th>
<th>pre-NWF errors</th>
<th>error rate</th>
<th>Blend</th>
<th>post-NWF errors</th>
<th>error rate</th>
<th>Blend</th>
<th>NWF-ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56</td>
<td>9</td>
<td>16%</td>
<td>Partial</td>
<td>93</td>
<td>5</td>
<td>5%</td>
</tr>
<tr>
<td>2</td>
<td>44</td>
<td>5</td>
<td>11%</td>
<td>Partial</td>
<td>84</td>
<td>4</td>
<td>5%</td>
</tr>
<tr>
<td>3</td>
<td>65</td>
<td>4</td>
<td>6%</td>
<td>Partial</td>
<td>126</td>
<td>2</td>
<td>2%</td>
</tr>
</tbody>
</table>
Table 3

**Pre-Intervention and Post-Intervention R-CBM results**

<table>
<thead>
<tr>
<th>Part. #</th>
<th>R-CBM--pre errors</th>
<th>error rate</th>
<th>R-CBM-post errors</th>
<th>error rate</th>
<th>R-CBM-ROI 15weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>5</td>
<td>24%</td>
<td>50</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>7</td>
<td>30%</td>
<td>48</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>7</td>
<td>27%</td>
<td>61</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4

**Pre-Intervention and Post-Intervention iRead results**

<table>
<thead>
<tr>
<th>Part. #</th>
<th>pre # weak</th>
<th>pre # mastery</th>
<th>post-# weak</th>
<th>post # mastery</th>
<th>taught skill ROI</th>
<th>Additional skills mastered*</th>
<th>Additional skills improved**</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>34</td>
<td>11</td>
<td>12</td>
<td>33</td>
<td>100%</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>23</td>
<td>10</td>
<td>35</td>
<td>100%</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>27</td>
<td>25</td>
<td>20</td>
<td>38%</td>
<td>5</td>
<td>8</td>
</tr>
</tbody>
</table>

*Skills not taught directly but for which each participant showed mastery in the post-intervention assessment.

**Skills not taught directly but for which each participant showed improvement in the post-intervention assessment.
**Visual Analysis**

Data were analyzed visually through plotting each participant’s NWF and R-CBM results graphically. Mean differences in level for each phase were calculated and plotted. The results for NWF and R-CBM are presented for each participant in Figures 1 through 6 consecutively.\(^1\)

As can be seen, each participant had a minimum of 4 data points (2 weeks) during baseline and a minimum of 6 data points (3 weeks) during the intervention phase. Participant 1 began intervention on the sixth day of data collection (end of the third week of the study), Participant 2 began intervention on the twelfth day of data collection (end of the sixth week of the study), and participant 3 began intervention on the seventeenth day of data collection (beginning of the ninth week of the study).

Given the randomization process, all of the participants received more than an adequate number of baseline and intervention weeks (Participant 1: baseline = 2.5 weeks, intervention = 9.5 weeks, Participant 2: baseline = 5.5, intervention = 6.5 weeks, and Participant 3: baseline = 8 weeks, intervention = 4 weeks). As noted previously, the participant who was withdrawn had begun intervention on the fifth week of the study and thus would have had a longer intervention phase to analyze. This issue may have been significant in that, on visual inspection, it appeared that for each participant there was a delay prior to the data indicating an upward trend. This is particularly pronounced in the data for Participant 1 with regard to NWF. An implication of this may have been that for Participant 3, though there were four weeks of intervention, there may not have been enough time to see a change.

---

\(^1\) Note that on these graphs there are 21 data points as it was necessary to add an additional data point for each subject between the baseline and intervention phases.
The mean scores for each of the participants on both NWF and R-CBM are summarized in Table 5. For each participant N = 20 and the Standard Deviation for each mean is in parentheses.

Table 5

*Mean Differences Results*

<table>
<thead>
<tr>
<th></th>
<th>Nonsense Word Fluency</th>
<th></th>
<th></th>
<th></th>
<th>Reading-CBM</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline M (SD)</td>
<td></td>
<td></td>
<td></td>
<td>Baseline M (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intervention M (SD)</td>
<td></td>
<td></td>
<td></td>
<td>Intervention M (SD)</td>
<td>Difference</td>
<td></td>
<td>Difference</td>
</tr>
<tr>
<td>Part. 1</td>
<td>53.6 (6.5)</td>
<td></td>
<td></td>
<td>17.9</td>
<td>39.4 (4.9)</td>
<td>41.1 (11.9)</td>
<td></td>
<td>1.7</td>
</tr>
<tr>
<td>Part. 2</td>
<td>60.7 (16.1)</td>
<td></td>
<td></td>
<td>14.5</td>
<td>38.7 (10.5)</td>
<td>43.9 (14.4)</td>
<td></td>
<td>5.2</td>
</tr>
<tr>
<td>Part. 3</td>
<td>87.4 (18.8)</td>
<td></td>
<td></td>
<td>18.1</td>
<td>54.7 (13.1)</td>
<td>60.5 (20.5)</td>
<td></td>
<td>5.8</td>
</tr>
</tbody>
</table>
Figure 1: Participant 1: Nonsense Word Fluency mean differences results.

Figure 2: Participant 2: Nonsense Word Fluency mean differences results.

Figure 3: Participant 3: Nonsense Word Fluency mean differences results.
Figure 4. Participant 1: Reading Curriculum Based Measurement mean difference results.

Figure 5. Participant 2: Reading Curriculum Based Measurement mean differences results.

Figure 6. Participant 3: Reading Curriculum Based Measurement mean differences results.
**Figure 7.** Participant 1: Nonsense Word Fluency slope differences results.

**Figure 8.** Participant 2: Nonsense Word Fluency slope differences results.

**Figure 9.** Participant 3: Nonsense Word Fluency slope differences results.
Figure 10. Participant 1: Reading Curriculum Based Measurement slope difference results.

Figure 11. Participant 2: Reading Curriculum Based Measurement slope differences results.

Figure 12. Participant 3: Reading Curriculum Based Measurement slope differences results.
Statistical Analysis

Randomization Analysis. Given the potential issues with bias and inaccuracy of visual analysis (Brossart, Parker, Olson, & Mahadevan, 2006; Kratochwill & Levin, 1992; Parker & Brossart, 2003, 2006; Todman & Dugard, 2001) data were analyzed statistically as well. A randomization test design was used in constructing the experiment and statistical analyses were conducted as per the procedures outlined in Todman and Dugard (2001). The primary focus of the analyses was on differences between the means of the baseline and treatment phases for each of the participants involved in the treatment. Todman and Dugard (pp. 164-168) provide a syntax for the analysis of a multiple baseline, randomization across individuals design using SPSS. That syntax was modified for the data that were collected in this design and used to determine if there were statistically significant differences for both R-CBM and NWF between the baseline and intervention phases. Todman and Dugard recommend resampling the data 2000 times in such a design to provide the best result for the randomization distribution used for comparison. If the difference between the hypothesized distribution and the observed data are significant at the .05 or lower level, the hypothesis that the use of the iRead Inventory to develop the intervention had a significant effect on improving reading skills acquisition would be supported beyond chance. The analysis of the NWF results yielded a one-tail probability of .98 (n = 60, p > .05) which indicates no difference between the observed NWF scores of the participants and those that were hypothesized and randomly generated from the data (minimum of 1967 randomly generated combinations). The analysis of the R-CBM results yielded a one-tail probability of .61 (n = 60, p > .05) which indicates no
difference between the R-CBM scores of the participants and a curve randomly generated from the data set (minimum of 1226 randomly generated combinations).

These results indicated no statistically different effect for the three participants on either R-CBM or NWF measures. Todman and Dugard (p. 45) suggest that if the data show an effect of trend or slope across phases the randomization design analysis may be confounded and suggest using alternative statistical analyses to explore possible effects. In examining Figures 1 through 6, it appears that there was a delay prior to participants exhibiting gains. This, combined with possible developmental growth effects, may have affected the results of the randomization tests and therefore alternative statistical analyses were conducted.

Effect Size Calculation. A review of the literature (Franklin, Allison, & Groman, 1996; Parker, 2006; Parker & Brossart, 2003; Parker et al., 2005; Parker, Hagan-Burke, & Vannest, 2007) indicated that the most reliable forms of statistically analyzing single case research involve calculating the Percentage of All Non-Overlapping Data (PAND) and computing a Phi coefficient as an expression of effect size (Parker et al., 2007) or calculating effect size using Allison’s MT formulation (Franklin, 1996; Parker & Brossart, 2003; Parker & Hagan-Burke, 2007). Parker and Brossart (2003) suggest using effect sizes to analyze single participant research for three reasons. First, effect sizes provide an index of the strength of association between intervention and outcome. Second, effect sizes provide a continuous (rather than dichotomous) index of treatment success. Third, effect sizes are not systematically affected by sample size, so a strong effect may be discerned even within a short data series. (Parker & Brossart, 2003). The authors state that, although
Cohen’s $d$ is the most common metric, regression based approaches offer more flexibility in interpretation. Moreover, when utilized with single case designs, Cohen’s $d$ does not effectively take into account issues associated with autocorrelation and the effect of slope or trend. The authors state that regression approaches may be more acceptable as they result in an $R^2$ or Cohen’s $d$ effect size that is accepted by most researchers, they use all data from both phases of the study, and they can be expanded to more complex analyses (Parker et al., 2007). Regression approaches, however, also have limitations, namely that the parametric assumptions may not be met in single case research, extreme outliers may unduly influence the regression model, and regression analyses require a level of expertise that may make them impractical for routine use.

As an alternative to regression based approaches, Parker et al. (2007) suggest that Percentage of All Non-Overlapping Data (PAND) may have additional promise as a method for analyzing single participant research. Parker, et al (2007) propose the PAND approach as an alternative and state that the approach allows for use of all of the data and can be translated to Pearson’s Phi, a universally accepted effect size. The disadvantages of PAND are that it will not be sensitive to data where there is no overlap at all between phases, it does not control for positive baseline trend, and it may have reduced statistical power in cases where there are not a large number of data points, i.e. less than 25 (Parker et al., 2007). PAND is, therefore, not recommended for single case designs or for designs for which there is no overlap between baseline and intervention phases (Parker et al., 2007). However, it is described as potentially very useful for statistical analysis of multiple baseline studies where there is data overlap.
that impact visual and other statistical analyses. The PAND approach is, therefore, very well suited for the analysis of this study across all three participants, i.e. to answer the question, “Was there an effect within the multiple baseline design for NWF and R-CBM improvement?”

The PAND approach, however, does not allow for analysis of the data for each of the participants individually because there were only 20 points per participant. In addition, visual analysis indicated that there seemed to be an effect of slope on the data (see Figures 7 to 12 where slopes of baseline and intervention phases show clear trends) and so regression analysis was conducted for each participant in order to determine the magnitude of effect for each participant. Parker and Brossart (2003) reviewed seven regression based statistical methods and suggested that, for single case research with typical numbers of data points (in the range of 20), the Allison approach to calculating effect sizes has the most power. In addition, the Allison approach allows the researcher to account for the effect of slope where necessary or to calculate only mean differences in those cases where slope will not affect the effect size calculation. As such, a combination of the PAND and regression based approaches (which accounts for baseline trend) appears most suitable to examine the current data.

Percentage of All Non-Overlapping Data Results. PAND involves determining the percentage of data points that overlap in both phases, calculating the percentage of data points unique to each phase, and putting this information into a 2 x 2 contingency table. This 2 x 2 table can then be analyzed and Pearson Phi effect size can be calculated. The PAND was calculated for the complete NWF and R-CBM data set for all three participants. The PAND calculations for NWF for the multiple baseline
design yielded a significant \( \Phi \) of .53 (\( n = 60, p = .00 \)). The PAND results for Reading-CBM yielded a non-significant \( \Phi \) of .13 (\( n = 60, p = .31 \)).

**Allison Regression Approach Results.** The Allison approach (Faith, Allison, & Gorman, 1996) involves using the residuals of the regression equation for the baseline to predict the regression equation for the intervention phase. The residuals of the actual intervention phase data are then compared to this equation and an \( R^2 \) can be calculated. The \( R^2 \) can be readily translated to a Cohen’s \( d \) and used to compare effect sizes in a commonly used fashion. This approach allows the researcher to determine if the slope affects the effect size calculation by using baseline residuals to predict intervention phase data. In addition, the analysis tests whether an effect size using mean performance is better suited for analysis as opposed to one that uses slope in the calculation of the effect size.

The effect sizes for NWF and R-CBM for each participant were calculated separately using the Allison regression approach (Faith et al., 1996). Table 6 summarizes the effect size results as calculated using the Allison regression approach for each participant. Included are the \( R^2 \) values, the Cohen’s \( d \) calculation, and the size of the effect based on Cohen’s recommendations for describing effect sizes, i.e. small = .20, medium = .50, and large = .80 (Howell, 2004, p. 340).
### Table 6

*Effect Size Results for Multiple, Baseline Data for All Three Participants*

<table>
<thead>
<tr>
<th>Part. #</th>
<th>Nonsense Word Fluency</th>
<th>Reading-CBM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>Cohen’s $d$</td>
</tr>
<tr>
<td>1</td>
<td>.06</td>
<td>.59</td>
</tr>
<tr>
<td>2</td>
<td>.17*</td>
<td>.92</td>
</tr>
<tr>
<td>3</td>
<td>.13*</td>
<td>.84</td>
</tr>
</tbody>
</table>

* Means of actual data were compared as slope did not impact effect size calculations (Franklin et al., 1996, p. 263).
CHAPTER IV
DISCUSSION

CBM tools are general outcome measures to assess weakness in an overall skill (i.e. reading, spelling, computation) and to monitor progress in the participant’s learning of that skill once an intervention has been implemented (D. Fuchs et al., 2004; Fuchs & Fuchs, 2004; Hintze et al., 2006). As general outcome measures they do not provide information regarding specific skills to be addressed for remediation (Hintze et al., 2006). While efforts to develop ways to use CBM and Curriculum Based Assessment (CBA) approaches for diagnoses of specific skill deficits have been explored (Burns, 2004; Fuchs, Fuchs, et al., 2003; Hosp & Fuchs, 2005; Stecker et al., 2005) the results of these efforts have largely suggested that further research and development is needed in this area. Hintze, et al. (2006) suggested that standardized specific skills mastery measurement tools need to be developed to assist teachers in targeting and remediating the skills that lead to weakness in the general skill. In the area of reading assessment R-CBM has been shown to be an excellent measure of overall reading ability (Hintze & Shapiro, 1997; Marston & Magnusson, 1985; Tindal et al., 1983). The DIBELS early reading measures have been demonstrated to be a reliable measure of early reading skills such as phonemic awareness and decoding (Good, Kaminski, Simmons, Kame'enui, & Oregon School Study Council, 2001; Good, Wallin, et al., 2002). These general outcome measures, however, do not allow for the assessment of specific decoding difficulties beyond the level of CVC words. Therefore, students may perform poorly with R-CBM but be fine in NWF and the teacher may not know where to target intervention. The iRead Inventory has been
developed to address this gap in assessment of decoding skills. The iRead Inventory has been developed to be a specific reading sub-skill measurement tool for assessing decoding skills with vowel and consonant combinations. The current study has been conducted to assess the ability of the iRead Inventory to identify specific skill weaknesses with decoding vowel combinations and to guide the development of instruction to address those identified weaknesses.

The current study was designed to address two hypotheses regarding the utility of the iRead Inventory. The first hypothesis was that the iRead Inventory would distinguish specific decoding sub-skill weaknesses for those students who struggle with decoding. The second hypothesis was that using the iRead Inventory in the development of interventions to target specific decoding sub-skills would lead to a greater rate of reading improvement for participants once the interventions were initiated than they showed prior to the targeted decoding instructions being provided. The following is a discussion of the results of the study regarding these two hypotheses.

The iRead Inventory as a Tool for Identifying Specific Decoding Weaknesses

The iRead Inventory was used to assess specific decoding weaknesses with vowel combinations for three participants in second grade who exhibited R-CBM weaknesses (Fall benchmark below 28 WRC and greater than 3 errors) and high errors rates in NWF (error rates above 3 SC). As indicated in Table 1, Participant 1 had difficulty with 34 of 45 specific vowel combinations assessed, Participant 2 had difficulty with 22 of 45 vowel combinations, and Participant 3 had difficulty with 28 of 45 vowel combinations. In addition, the level of difficulty with each specific skill
could be measured by whether the participant was only able to read 0 or 1 of the original 5 words correct or could read some number of words out of 10 correct. Appendix F summarizes the specific skill weaknesses for each participant and includes the ratio of correct words to words presented for each skill. This allowed for prioritizing of skills for intervention and for a measure of improvement over time. Table 4 and Appendix F provide information regarding the improvement for each participant in the targeted skill areas. In each case the iRead Inventory was able to distinguish specific vowel combination decoding weaknesses for remediation and to identify those skills for which there was improvement. Each participant showed increased mastery in those skills that were considered for intervention. In addition, each participant showed increased mastery of some skills that were not specifically targeted for intervention and demonstrated some improvement in other skills (i.e. the ratio of words correct to total words improved) as well. The iRead was able to distinguish and was sensitive to these changes over time. These results support the hypothesis that the iRead Inventory can distinguish specific decoding sub-skill weaknesses for students who struggle with decoding. Not only was the hypothesis supported but the iRead Inventory was able to produce ratios that indicate the degree of difficulty for each sub-skill for which there was a weakness and to show improvement over time.

The iRead Inventory as a Tool for Supporting Reading Improvement

The second hypothesis of this study was that the iRead Inventory combined with a direct, targeted reading instruction would lead to improved reading skills for the targeted participants. The results support this hypothesis in several ways. The
following discussion outlines how this hypothesis has been supported by the study results through qualitative analysis, visual analysis, and statistical analysis. Some weaknesses of the study are subsequently outlined. The discussion concludes with a review of possible implication for future research, both with respect to the general practice of assessing and teaching reading and with respect to the development of the iRead Inventory.

*Summative Findings.* All three participants were assessed pre- and post-intervention for NWF, R-CBM, and the iRead Inventory specific skills. When level of performance and rate of improvement were measured in a manner similar to the methods recommended by most authors to measure improvement from a progress monitoring perspective (Deno, Espin, Fuchs, Shinn, Walker, & Stoner, 2002; D. Fuchs et al., 2004; Fuchs, 1989; Hall, 2006; Hosp, Hosp, & Howell, 2007) the participants all showed improvement in NWF and R-CBM.

The improvements for NWF and decoding were most dramatic with respect to changes in level and rate (see Table 2). All three participants increased their number of Sounds Correct (Participant 1 = 37 SC/15 weeks, Participant 2 = 40 SC/15 weeks, and Participant 3 = 61 SC/15 weeks) and did so at a rate far greater than the average rate of improvement (1.7 SC per week) for first grade students at the ninetieth percentile (2.5, 2.7, and 4.1 SC per week for Participants 1 through 3, respectively) (AIMSweb, 2007a). Their NWF scores continued to be above the benchmark scores for the end of first-grade as well. The error rates for each participant decreased as well (Participant 1 = 9 (16%) to 5 (5%), Participant 2 = 5 (11%) to 4 (5%), and Participant 3 = 4 (6%) to 2 (2%)). All three participants also went from being partial blenders (blending less
than 90% of the words) to being full blenders (blending greater than 90% of the words). So, from the perspective of standard progress monitoring approaches to determining responsiveness to intervention, all three participants improved in decoding skills.

The growth in R-CBM scores for the participants showed improvement as well, though the improvement was not as dramatic as that of their NWF scores (see Table 3). All three participants increased their number of Words Read Correctly (Participant 1 = 29 WRC/15 weeks, Participant 2 = 24 WRC/15 weeks, and Participant 3 = 35 WRC/15 weeks) and did so at a rate greater than the average rate of improvement (1.1 SC per week) for second grade students at the ninetieth percentile (1.9, 1.6, and 2.3 for Participants 1 through 3, respectively) (AIMSweb, 2007b). The rates of improvement for Participants 1 and 3 were in the range of ambitious growth rates for second grade students (2 WRC/week) as recommended by Fuchs, Fuchs, Hamlett, Walz, & Germann, (1993). Participant 2 had a R-CBM improvement rate comparable to realistic growth rates for second grade students ( Fuchs, et al., 1993) and his R-CBM scores for the Winter benchmark were above the 25th percentile score of 53 (AIMSweb, 2007b). The error rates for each participant decreased as well (Participant 1 = 5 (24%) to 3 (6%), Participant 2 = 7 (30%) to 3 (6%), and Participant 3 = 7 (27%) to 3 (5%)). The changes in error rates were fairly dramatic and suggest that, although their R-CBM scores did not improve as dramatically as NWF scores, the improvement in decoding did seem to carry over into R-CBM with higher observed rates of accuracy in decoding words (i.e. each participant made fewer errors). Improvement in fluency does not seem to occur as immediately as improvement in
decoding, which makes intuitive sense as the focus of the intervention was on decoding. The participants still need to continue to develop their reading skills to make benchmark during subsequent assessments. Furthermore, the rate of improvement for the participants seems to indicate that they would continue to progress at a rate higher than, or at least equal to, their peers.

Lastly, a summative evaluation of each participant’s specific vowel combination decoding skills was administered using the iRead Inventory as a pre- and post-assessment. The number of weaknesses for two of the participants decreased quite dramatically (Participant 1 from 34/45 to 12/45, and Participant 2 from 22/45 to 10/45). Participant 3 did not exhibit the same levels of improvement with respect to specific decoding skills. He, however, had significantly fewer intervention sessions than the other two participants, which may partially explain the differences. Participants 1 and 2 also mastered 100% of the skills they were taught during the intervention phase, decoding all of the words in the skills they covered during the interventions. Participant 3 showed less mastery, only mastering 38% of the skills he was taught during intervention lessons. This may, however, again be explained in part by the fewer number of intervention sessions. His scores for the remaining 62% of the skills he had been taught did show some improvement, as they were higher than they had been on initial assessment. A very interesting result was that all three participants showed some generalization of decoding skill improvement to vowel combinations they had not been taught. This was exhibited in two ways. Each participant showed mastery of some vowel combination words that they had not been taught (Participant 1 = 9 additional sub-skills, Participant 2 = 8 additional sub-skills, and Participant 3 = 5
additional sub-skills). In addition, all three participants showed improvement in their scores for a number of sub-skills in that their iRead score was higher than on the initial assessment (Participant 1 = 8, Participant 2 = 4, and Participant 3 = 8). Decoding skills were not taught in any other context (e.g. regular language arts or Title one instruction) during the time of the study increasing the likelihood that improvement in decoding skills could have been primarily attributed to the intervention. It seems that targeting decoding instruction to specific sub-skills may generalize to other decoding abilities. This is one area for further research.

Visual Inspection Findings. While the summative results were promising, further analysis of the dependent variables needed to be conducted to determine if those results were supported formatively over the course of the study. When reviewing time-series graphs of the data several conclusions can be drawn. The NWF results were quite clear from a visual analysis perspective. For each participant, the change in level between baseline and intervention phases is readily apparent when looking at the graphs. The means for each phase differed by a large number of sounds correct (17.9, 14.5, and 18.1 for Participants 1 through 3, respectively).

As noted previously, for each participant there seemed to be an initial period (5 data collection points or 2.5 weeks) where his NWF scores did not vary significantly from the baseline phase. Then, each participant exhibited a noticeable increase in NWF scores. It seems that, at least with this intervention process, there is a learning curve during which participants needed to receive instruction for some time prior to showing improvement in the given skill. There is considerable support for the concept of a learning curve in cognitive psychology (Anderson, 1982; Fitts & Posner,
Fitts and Posner (1967) reviewed a number of studies and postulate that there is a curve to learning whereby skill acquisition is best described as a power curve with the rate of learning being best described as $f(x)$ where the dependent variable is raised to some power, e.g. $y = x^a$ where $a$ is a function of practice and repetition. Ninio (2006, pg. 41) illustrates a hypothetical upward learning curve for skill acquisition. This observation and the possibility of a learning curve with positive power (Ninio, 2006) has potentially significant implications for the process of monitoring student progress under an RTI model where generally the expectation for student improvement is assumed to be linear (Batsche et al., 2005; Fuchs & Fuchs, 2001; Stecker, 2006) and changes in intervention are predicated on student performance relative to this expectation. If the delay in learning observed in this study occurs for other students, then conventional rules for determining a change in intervention must be called into question (e.g. changing an intervention when the four most recent data points fall below the expected rate of progress or aim line). It may well be likely that some students exhibit a learning curve where their scores initially fall below a linear aim line and a call for a change in intervention may be made, when what was needed is more time for the intervention to gain effect. Further research with respect to this question is warranted.

The graphic results for R-CBM are less equivocal than the results for NWF. The differences in level between phases are less apparent and there seems to be more overlap of data points than occurred in the NWF results, particularly if one looks at the NWF data after the initial period of learning. The mean differences in words read correctly are relatively small between phases (Participant 1 = 1.7, Participant 2 = 5.2,
and Participant three = 5.8) for R-CBM. The R-CBM graphs did not as clearly exhibit the delay in score improvements noted in the NWF graphs. It seems, from an initial analysis of the visual data, that targeting decoding skills did not lead as directly to R-CBM improvement as it did for NWF improvement.

These results for R-CBM may be explained by two possible hypotheses. The first is that an improvement in decoding skills may not immediately translate to an improved ability to read fluently. The development of reading skills may be a stepwise process whereby each skill needs to be addressed separately and sequentially. Certainly, this is supported by research (Beck, Osborn, & Lehr, 1998; Birsh, 2004; NRP, 2000; NRC, 1998; Schumm, 2006; Torgesen, 2002). This may be particularly true for students who are struggling to learn to read. As the current participants were second grade students who were below the 25th percentile for R-CBM scores, they may need focused instruction in decoding, followed by fluency instruction, and then subsequent instruction in vocabulary, and comprehension. The second hypothesis is that focusing on one skill, i.e. decoding, may limit the students’ ability to focus on another. In other words, as students focus on addressing a particular weakness, their cognitive capacity to address other skills may be diminished while they focus on the particular weakness. It could be the case that the current participants might show improvement in their R-CBM scores after their decoding skill gains were consolidated. As a result, they may not need fluency instruction but rather need time to integrate their decoding skills and apply them prior to exhibiting improvement in their fluency.
The graphic results for NWF and R-CBM slope differences indicate that the data may have a trend in the baseline phase that will impact the statistical analysis of the results (see Figures 6 through 12). It is, therefore, necessary to utilize a statistical analysis that accounts for slope differences as part of the statistical analysis.

**Randomization Analysis Findings.** The findings for the randomization analysis indicate that there is no significant change for the three participants as a group for either NWF or R-CBM scores (NWF = .98, \( p > .05 \), R-CBM = .61, \( p > .05 \)). These results seem to contradict the visual inspection. However, as noted previously, Todman and Dugard (2001) acknowledge that the randomization analysis can be confounded by a trend in the baseline data. Clearly, the review of the graphs for trend differences indicates that there are trends in the baseline phases for each participant. It appears that the randomization analysis may not work for studies for which there is learning that occurs cumulatively and where rate of improvement confounds can be anticipated. It is, however, still a very valid way to structure studies as the randomization of intervention points allows for other statistical approaches to be used that require this (Parker et al., 2007) and minimizes the impact of baseline bias (Todman & Dugard, 2001).

**Percentage of All Non-Overlapping Data Findings.** A PAND analysis was conducted for both NWF and R-CBM for all three participants to determine if there was a statistically significant difference for either of these abilities. The analysis yielded results that indicated that the changes in NWF scores were significant (\( N = 60, \Phi_i = .53, p = .00 \)) but that changes in the R-CBM scores were not significant (\( N = 60, \Phi_i = .13, p = .31 \)). This supports the conclusion reached from the visual analysis, namely
that the use of the iRead Inventory to target instruction to specific decoding skills weaknesses yields a significant improvement in decoding skills but not a concurrent improvement in fluency skills. The PAND analysis provides a statistical analysis of the multiple baseline results for all three participants. However, this analysis cannot be applied to the individual participants to assess which participant showed significant improvements for two reasons. The first of these is that there were only 20 data points for each participant and the PAND process is not recommended for data sets of that size (Parker et al., 2007). The second reason is that the PAND process is not sensitive to effects of trend (Parker et al., 2007) and we have seen from visual analysis that there appears to be a possible confound from slope. It was determined, therefore, to complete a regression based effect size analysis for each participant’s results to determine for whom and to what degree there was an impact of the intervention. This analysis was conducted for both NWF and R-CBM as the R-CBM group results may have been affected by slope. There may be particular participants for whom a significant effect did occur with respect to R-CBM and all of the prior analyses may not have been sensitive to that change.

Allison Effect Size Calculation Findings. The effect size calculations using the Allison (Faith, 1996, p. 263) approach yielded some very interesting results (see table 6). As predicted from the visual and PAND analyses, there were effects for each of the participants with respect to NWF (Participant 1 Cohen’s $d = .59$, medium effect size; Participant 2 Cohen’s $d = .92$, large effect size; and Participant 3 Cohen’s $d = .84$, large effect size). The analysis of the R-CBM effect sizes for each participant yielded some interesting results that contradicted the visual and PAND analyses. When trend
was taken into account, an effect was found for Participants 1 and 2 R-CBM abilities (Participant 1 Cohen’s $d = 1.04$, large effect size; and Participant 2 Cohen’s $d = 1.13$, large effect size). For these two participants, the intervention seemed to have affected their reading fluency once trend was accounted for. This result was not apparent in any other analysis. For Participant 3 the effect size analysis yielded a small and negative effect size when slope was taken into account (Cohen’s $d = -.17$). This result may have been influenced by the relatively short period of intervention for this participant. These results lend some credence to the second hypothesis for the R-CBM results, namely that while improvements in fluency may not be apparent immediately if decoding is addressed, those improvements may occur as decoding skills are consolidated. The first two participants had time to consolidate their decoding gains and showed a large effect for R-CBM scores between phases when trend was accounted for. Participant three did not show the same effect and that may have been due to lack of opportunity to consolidate gains. These results at the individual level also speak to the validity of doing an analysis that accounts for the effect of slope on the data in order to get an accurate assessment of student progress, an assessment that was not available from visual and PAND analysis.

The results from this analysis and prior analyses support the second hypothesis of this study, namely that the iRead Inventory can be used to support the development of specific skills instruction that can lead to student improvement in reading. The primary effect is on decoding ability which is to be expected given the focus of the iRead Inventory assessment, i.e. vowel combination skills, and the intervention used in the study (Carnine et al., 2004).
Implications

The results of the study support both hypotheses regarding the iRead Inventory, namely that the iRead Inventory can be used to identify and target specific decoding sub-skills for intervention and that the use of the iRead Inventory to target interventions will lead to an improvement in reading ability for students above and beyond the improvement in their reading ability without the targeted instruction. The iRead Inventory assisted in identifying specific decoding sub-skills as evidenced by the list of sub-skills identified as weaknesses for each participant. A measure of the comparative weakness of those skills was also developed using iRead Inventory scores. In addition, the iRead Inventory was used to monitor mastery of each of the skills as students were learning and was used as a post-intervention assessment of progress. For two of the participants, the iRead Inventory results indicated mastery of all of the taught skills after intervention was complete while one participant, who had the briefest intervention phase, did not show as consistent a mastery of taught skills. In addition, all of the participants showed some improvement in decoding with sub-skills they were not taught directly. The iRead Inventory was able to assess that generalization of decoding skills as well.

For this study the iRead Inventory was used in a diagnostic assessment format, i.e. all vowel combination skills were assessed. For use in the classroom, it may be more beneficial to use the iRead in an intervention assessment format, i.e. identify several skills at a time for remediation, target those skills for intervention, then assess for additional weaknesses once those skills are mastered. This would allow for real time monitoring of decoding skills as intervention is being implemented. The repeated
assessment by the teacher would monitor evolving skills so that if the student was
generalizing, i.e. applying learning to skills other than the specific ones being taught,
those improved skills would be detected. This would allow the teacher to provide
instruction for only those skills that continued to be a weakness for the student. The
results of this study indicate that the iRead Inventory can be used to identify specific
decoding weaknesses and progress related to instruction.

The second hypothesis was that the iRead Inventory could be utilized to inform
and target instruction for decoding skills and lead to an improvement in student
reading skills. The results indicate that the iRead, in combination with targeted
instruction, can lead to improvement in decoding skills. The summative, visual,
PAND, and Allison effect size calculations all support this conclusion. Use of the
iRead Inventory to target instruction also seems to support the development of reading
fluency, though these results were less equivocal. The summative results indicated
some improvement in R-CBM but the visual and PAND analysis seemed to indicate
that there was little, if any, improvement in R-CBM scores between the baseline and
intervention phases. However, when an effect size analysis was conducted that
accounted for trend in the data, the results indicated that a large effect was present for
R-CBM scores for the two participants who had the longest period of instruction. The
third participant who had four weeks of instruction showed a small, negative effect
size for R-CBM scores between baseline and intervention phases. This may be
accounted for by the short duration of the intervention or by the fact that the gains may
not have had enough time to show in the assessment process. The iRead Inventory
seems to be an effective assessment for specific decoding sub-skill instruction and
developing interventions targeted toremediating those skills. In fact, the iRead Inventory in combination with intervention, supports generalization of decoding skills and can measure that generalization so the teacher can target skills that the student needs to learn.

An interesting result of this study was that all of the participants seemed to exhibit a delay in performance improvement on the dependent variables after intervention was initiated. With respect to NWF improvement there was a delay of approximately 2.5 weeks for each participant between the point when intervention was initiated and gains were shown in the NWF scores. This delayed acquisition effect was not as apparent on visual analysis of the R-CBM data for the three participants. It is reasonable to expect that if the intervention focus is decoding, the improvement in fluency may take longer to appear in assessments. Given that the R-CBM data showed consistent trend effects for all three participants, it was necessary to conduct further analysis of these data. When effect sizes were calculated that accounted for the effect of trend, participants one and two showed a large effect size between baseline and intervention phases for R-CBM. This indicates that, in fact, there was an effect of intervention on R-CBM for these participants. The effect was just not readily observable with more commonly utilized analyses. This result supports the hypothesis that there is a learning curve whereby students may not immediately exhibit gains in skills that are being targeted for intervention. Some amount of time, in this case approximately 2.5 weeks of daily direct instruction, may need to pass prior to students exhibiting gains in a skill for which they are receiving instruction. It may be even longer for a higher order skill to show improvement, i.e. fluency. This hypothesis is
also supported by the results on R-CBM for participant three. He had only had a brief
time to benefit from instruction and did not show the same fluency gains, as measured
by regression analysis effect sizes, within the time frame of the study.

Study Limitations and Directions for Future Research

This study has several limitations. The first of these is that the iRead Inventory
is still in the research development phase. There are limited reliability, validity, or
administration analyses of the iRead Inventory as an assessment tool. Several studies
are being conducted regarding the psychometric properties of the iRead Inventory.
Preliminary results of a study by McGurl, Farrell-Meier, Wells, and Hintze (2008)
indicate that there is strong reliability of the word lists and the words work well to
discriminate between poor, average and strong readers. Farrell-Meier, McGurl, Wells,
and Hintze (2008) also have obtained preliminary data regarding the validity of the
iRead Inventory. The iRead Inventory seems to have strong validity when compared to
several standardized reading assessments. A final analysis of the data from those
studies is necessary before we can assess with confidence whether the results for the
iRead Inventory’s utility are accurate. Ongoing research needs to be conducted on the
utility of the words being used to assess specific sub-skills, on the administration and
scoring procedures, on the number of words being read, and on the utility of the
assessment process for informing instruction. Considering the process that was utilized
to develop the iRead Inventory the author is fairly confident in its utility, however,
additional empirical research needs to continue to standardize and develop this tool.

The participants in the study were from a rural community with a fairly
traditional, literature-based approach to reading instruction. The weaknesses exhibited
by these participants may be instructionally based and the gains may be associated with receiving sustained, targeted, direct instruction for the first time. The number of weaknesses and the gains exhibited may be different in a setting where students are receiving more empirically supported instruction and supplementary interventions. In addition, the participants were all Caucasian, males from a middle class background. Further studies need to be conducted regarding the use of the iRead Inventory with urban students from more diverse backgrounds.

The question of a learning curve is an area for future research. Given that planning for RTI progress monitoring focuses on comparing a student’s progress in a linear fashion (Batsche et al., 2005; Fuchs & Fuchs, 2004; Hosp et al., 2007), it may be significant that students may not exhibit learning of skills in this manner. Certainly, the cognitive psychology literature can provide some guidance with respect to this question (Fitts & Posner, 1967; Anderson, 1982; Logan, J. D., 1992; Ohlsson, S., 1992; Ninio, A., 2006). However, research should be conducted into whether this phenomenon is routine, whether the learning curve has a consistent form, and whether there are implications for intervention planning. This is particularly true if the heuristic of changing interventions after four data points fall below the aim line is being followed (Hosp et al. 2007). Hosp et al. (2007) do suggest that no decision be made regarding the significance of four consecutive data points until at least 8 data points have been gathered and they emphasize that they need to be consecutive data points below the trend line. This specific suggestion regarding the four points heuristic may be particularly important to use in guiding decision making if the learning curve phenomenon is consistently exhibited. It may be true that many students will exhibit a
learning curve that is likely to have initial data points below the aim line. Students may need a certain amount of time to integrate learning and be able to show gains on measurements of those skills. Research into whether this phenomenon is consistently exhibited and the actual nature of learning curves manifested when using curriculum-based measurement to track progress would be warranted and may inform the process of developing heuristics for decision making within an RTI framework.

Lastly, research into the question of generalization of learning decoding skills may be warranted. This study indicates that instruction in specific decoding skills may generalize to increased ability in decoding, even in skills not directly taught. Assessment and intervention of decoding skills acquisition may need to take into account this generalization process, if it proves to occur consistently. Research could focus on answering questions regarding which specific skills tend to generalize to other skills, the optimal order of teaching skills in order to increase the generalization potential, and ways to develop assessment tools that measure that generalization efficiently in order to allow teachers to tailor interventions to address only those skills for which a particular student is still exhibiting weakness.
APPENDIX A
CONSENT FORM

Dear Parent/Guardian,

I am conducting a research study on using a reading assessment tool to develop targeted interventions to help students learn to read more fluently. Your child is eligible for the study based on the school wide fall Oral Reading Fluency assessment results. The information collected will be used for my dissertation in the area of school psychology at the University of Massachusetts – Amherst.

• If you choose to participate, your child will be involved in three ways: 1) Your child will be asked to read a series of words that help me to assess the specific letter-sound combinations that are hard for him or her to sound out. 2) Based on the identification of specific letter-sound combinations to target, your child will receive a daily 15 minute lesson targeted to learning those combinations. 3) Your child will be asked to read a passage for one minute and to read a series of nonsense words for one minute twice a week for the duration of the study. The lessons will be video taped for review and the twice weekly readings will be audio-taped.

• Each child’s privacy will be protected by using alternate names on any written materials and when writing up the results. The audio and video tapes will be locked away securely and used only for analysis for the study. Upon completion of the study they will be destroyed. Your child will be leaving the classroom daily for the instruction but the purpose will not be shared with peers or staff members who do not need to know.

• There are no known risks involved in having a student read words or passages, or participating in the direct instruction regarding specific letter-sound combinations. However, if a student becomes tired or finds the activity challenging they are free to stop the activity at anytime. Your child will be told this before the study begins.

• The potential benefit of this study would be that your child would be contributing to research in the area of developing effective instruction for specific letter-sound combination difficulties. In addition, you child will receive specific instruction about letter-sound combinations that are difficult for him or her. A potential benefit is that your child’s decoding and fluent reading skills might improve. I am hopeful that this research will ultimately benefit young children by helping develop an assessment tool that can help teachers specifically target the skills they need to teach to help students decode more easily.

• Participation in this study is voluntary.
  o If you consent to your child’s participation, please return the slip and check the “DO ___” choice on the form below. Your child will be asked to participate, and at any time during this study if your child gets tired or does not want to continue they will simply return to their class.
  o If you’d prefer to NOT have your child participate, you do so without prejudice. Please return the slip and check the “DO NOT ___” choice.
  o Please complete the bottom of this page and return it to your child’s teacher by: ____________.

If you have any questions or concerns please contact: Andrew J. Koerner, Graduate Student Investigator at akoerner@educ.umass.edu or 413 477-6351 or John Hintze, Research Supervisor at 413 577-1470

Thank you very much for you time,

_____________________
_____________________
_____________________

Andrew J. Koerner, MA       John Hintze, Ph.D.       Catherine Brandon, MS

I, ______________________  DO   DO NOT   want my child, ____________________ to participate
                          Parent/Guardian Name                           Child’s Name
in this research study.

_____________________
Parent/Guardian Signature

_____________________
Date

*Please return this form with either consent or non-consent checked above*
APPENDIX B

IREAD INVENTORY
GENERAL ADMINISTRATION INSTRUCTIONS

The examiner follows these standardized procedures exactly.

The examiner says to the student:

When I say, ‘start,’ begin reading each word, beginning at the top of this page.
Read across the page [demonstrate by pointing]. Try your best to read each word.
If you come to a word you don’t know try to read it but if you can’t read it you may skip it.
Be sure to read whatever words you can read. When you get to the bottom of this page, stop.
Are there any questions?
[Pause] Start.

As the student reads the words, the examiner records any errors by marking a slash (/) through the incorrectly read word. The examiner circles any particular letter or letter combination the student reads incorrectly or struggles to pronounce. If the student hesitates for more than 5 seconds on any word, the examiner says “Try your best to read it. If you can’t you may skip it.” That word is marked as an error if the student cannot read it after the prompt.

The student may be provided with a break between pages being read. When the student completes a page the examiner may ask Are you ready for the next page or would you like a break? If the student would like a break give him or her a few seconds and up to a minute and then go on. If they do not want a break, continue to the next page.

When continuing to subsequent pages you may follow the directions above or, if the student clearly understands the directions, shorten them as follows:
When I say “start” begin reading the words and try to read each word. When you get to the bottom of the page stop.
Ready, “start”. This can be shortened even further if deemed necessary

Scoring and Discontinue Directions

Scoring

The examiner notes the number of words read correctly in the space below each column of words for a skill set. For any given column where the student reads 0 or 1 words correctly, the examiner circles the specific skill to indicate that the student has a weakness in that specific skill. No further assessment of that skill is necessary. If the student reads 2 or 3 words correctly the examiner turns to the “b” page of the assessment and follows the following protocol. If the student reads 4 or 5 words correctly, no further assessment of that skill is necessary.

Specific Sub-skill Extended Assessment: The examiner points to the column that corresponds to the sub-skill within which the student scored a 2 or a 3 and says, “Please read these words to me.” The examiner marks errors as above and then totals up the errors from both columns. If the total is less than 8 the examiner circles that skill as an area of weakness for that student.

Discontinue Directions

Intervention Planning Assessment: If the iRead is being used to assess a student for instructional purposes only (i.e. the purpose of the assessment is to identify focused skills for a teacher to remediate), follow these discontinue rules:

Administer the Core Vowel section and then the Supplemental Vowel section: If, at the end of any page of words the student has had a TOTAL (including from prior vowel skill pages) of 5 vowel skill weaknesses (five specific skills circled) discontinue the vowel section and go on to the consonant section. Offer the student a break and continue administration as above.

Diagnostic Evaluation Assessment: If the iRead inventory is being used as part of a comprehensive diagnostic assessment (i.e. the assessment is being used to identify all of the weaknesses in a student’s decoding skills), follow these discontinue rules:

Administer the complete Core Vowel section regardless of number of errors.
APPENDIX C

ASSESSMENT SAMPLE PAGES

iRead Inventory
Core Vowel Section
Examiner Score Sheet
Page 2- Closed Syllables-Short Vowels (K)

2a

<table>
<thead>
<tr>
<th>ask</th>
<th>mess</th>
<th>bill</th>
<th>trot</th>
<th>stub</th>
</tr>
</thead>
<tbody>
<tr>
<td>lap</td>
<td>deck</td>
<td>dip</td>
<td>plop</td>
<td>hut</td>
</tr>
<tr>
<td>mash</td>
<td>tempt</td>
<td>with</td>
<td>bomb</td>
<td>grub</td>
</tr>
<tr>
<td>stab</td>
<td>fleck</td>
<td>rib</td>
<td>dock</td>
<td>such</td>
</tr>
<tr>
<td>flag</td>
<td>shell</td>
<td>hint</td>
<td>cop</td>
<td>rub</td>
</tr>
</tbody>
</table>

___/5 short a  ___/5 short e  ___/5 short i  ___/5 short o  ___/5 short u

If the student reads only 0 or 1 words correctly in a column, circle that skill; **do not** have them read more words from that skill set. If student reads 2 or 3 words correctly in a column, have the student read the words in the corresponding column from the same skill set on page 2b. If the student reads 4 or 5 words correctly, no further assessment of that skill is necessary.

2b

<table>
<thead>
<tr>
<th>tax</th>
<th>beck</th>
<th>hit</th>
<th>shot</th>
<th>fuss</th>
</tr>
</thead>
<tbody>
<tr>
<td>wag</td>
<td>send</td>
<td>twitch</td>
<td>slob</td>
<td>blush</td>
</tr>
<tr>
<td>rash</td>
<td>strep</td>
<td>slick</td>
<td>bog</td>
<td>plus</td>
</tr>
<tr>
<td>tap</td>
<td>fence</td>
<td>chip</td>
<td>flock</td>
<td>grunt</td>
</tr>
<tr>
<td>stack</td>
<td>dress</td>
<td>limp</td>
<td>clog</td>
<td>spun</td>
</tr>
</tbody>
</table>

___/10 short a  ___/10 short e  ___/10 short i  ___/10 short o  ___/10 short u

If a student reads a row or column of words from page 2b, total the number of words read correctly from the rows or columns on page 2a and 2b. If the total is less than 7, circle that skill.

Continue to page 3a.
<table>
<thead>
<tr>
<th>ask</th>
<th>mess</th>
<th>bill</th>
<th>trot</th>
<th>stub</th>
</tr>
</thead>
<tbody>
<tr>
<td>lap</td>
<td>deck</td>
<td>dip</td>
<td>plop</td>
<td>hut</td>
</tr>
<tr>
<td>mash</td>
<td>tempt</td>
<td>with</td>
<td>bomb</td>
<td>grub</td>
</tr>
<tr>
<td>stab</td>
<td>fleck</td>
<td>rib</td>
<td>dock</td>
<td>such</td>
</tr>
<tr>
<td>flag</td>
<td>shell</td>
<td>hint</td>
<td>cop</td>
<td>rub</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tax</td>
<td>beck</td>
<td>hit</td>
<td>shot</td>
<td>fuss</td>
</tr>
<tr>
<td>wag</td>
<td>send</td>
<td>twitch</td>
<td>slob</td>
<td>blush</td>
</tr>
<tr>
<td>rash</td>
<td>strep</td>
<td>slick</td>
<td>bog</td>
<td>plus</td>
</tr>
<tr>
<td>tap</td>
<td>fence</td>
<td>chip</td>
<td>flock</td>
<td>grunt</td>
</tr>
<tr>
<td>stack</td>
<td>dress</td>
<td>limp</td>
<td>clog</td>
<td>spun</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX D

LESSON PLAN SAMPLES

Vowel-Sound Correspondence

*Student has exhibited a weakness with a particular short vowel sound, in this case short e.*

Part 1

Teacher: Student has exhibited a weakness with a particular short vowel sound, in this case short e.

1. Teacher writes on the board: e. **When I touch under the letter, you say the sound. Keep saying the sound as long as I touch it.**
2. Teacher *models* the sound. Teacher holds her finger under the letter and says: **“My turn.”** Teacher moves finger out and in, touching under the letter for 2 seconds if it is a continuous sound and for an instant if it is a stop sound. Teacher says the sound while touching the letter, then quickly moves her finger away from the letter and immediately stops saying the sound.
3. Teacher *tests* by having the student say the sound several times with directions:
   - a. Teacher points under the letter and says: **What sound?** Signal by touching under the letter for about two seconds.
   - b. Teacher repeats step a several times, touching under the letter “eeeeeee” from 1 to 3 seconds.
   - c. Teacher corrects any incorrect pronunciation immediately.
      Listen: **“eeeee”** and then repeats a and b.
4. Teacher continues to part 2 when student correctly pronounces the letter sound at least 5 times.

Part 2

Teacher: Student writes on the board several letters that the student has mastered along with the letter that is the focus of the lesson. The letters used are initially those for which the student scored well on the iRead Inventory and in subsequent lessons are ones learned previously to allow for review. The focus letter appears several times in different positions (focus letter should be approximately every 5th letter).

1. Teacher says: **When I touch under a letter, you say the sound. Keep saying the sound as long as I touch under it.**
2. Teacher tests new sound. She points to the focus letter first, pauses 2 seconds, moves her finger out and in, touching under the letter for 2 seconds if it is a continuous sound and for an instant if it is a stop sound.
3. Teacher tests all of the letters. Points to a letter, pauses 2 seconds, then moves finger in and out, touching the letter.
   - Says the sound
   - Says the sound
   Teacher alternates the letters with the focus letter being repeated frequently at first and then diminishing over time (i.e. e t e c d e a h i e etc...)
4. Errors are corrected by modeling the correct sound and asking the student to repeat it.
   **Lesson ends when student pronounces target sound correctly 5 or more times.**
Letter Combinations

Student has exhibited a weakness with a particular letter combination, in this case O-controlled R.

Part 1

Teacher
Teacher writes on the board: or, e, ea, or, th, sh, ai, or, ing
Sounded like: or in short, e in bet, ea in beat, th in thank, sh in shop, ai in maid, ing in sing (The alternative letters and combinations are chosen based on three criteria: combinations or single letters that the student has read successfully in the iRead Inventory, has been taught in prior lessons and needs to review and sequencing considerations outlined in Carnine et al (1994).)

1. Teacher models by saying the sound of the new letter combination and tests by having the student pronounce it. Teacher points to or and says: These letters usually say /or/. (as in short) What sound? (signal) “or”

2. Teacher alternates between the new combination and other combinations. Teacher points to a letter combination, pauses 2 seconds, says “What sound?” and signals with an out-in motion.

3. Teacher presents the remaining letter combinations using an alternating pattern similar to this: or, e, ea, or, th, or, sh, ai, ing, or.

4. Teacher corrects errors by telling the student the correct sound, having the student say the sound, and then continuing to alternate with other sounds.

5. Once the student correctly identifies the focus combination at least 5 times, the teacher goes on to part 2.

Part 2

Teacher
Teacher writes on the board: born, lord, fort, boot, round, horn, stain, moon, out, cord
(The alternative words are selected based on the criteria outlined in part 1 with the first three words having the focus combination and one third to one half of the remaining words having the focus combination.)

1. a. Students identify the sound of the letter combination, then read the word. Teacher points under the underlined letters and asks, “What sound?” (Signal.) “or”

   b. Teacher points to the left of the word. “What word?” (Signal.) “born”

   c. Teacher repeats steps 1a and 1b with remaining words.

2. a. Students reread the list without first identifying the sound of the letter combination. Teacher points to born, pauses 2 seconds, and asks, “What word?” (Signal.) “born” sound and word

   b. Teacher repeats step 2a with remaining words.

3. Teacher corrects errors by asking the student to say the sound. If the error continues, the teacher models the correct sound and has the student repeat it. Then she asks the student to say the word and goes back 4 words and begins again.

Lesson ends when student can repeat all target sound words correctly.
VCe Pattern Words

Student has exhibited weakness with words that end in a VCe pattern.
The general rule is taught for all VCe words unless if the student shows a weakness with one particular vowel in which case this format is used but with an emphasis on that vowel.

Part 1

Teacher
Teacher writes on the board: game, rope, mine, tape, note, time.
(Words are chosen from a, i, o/Ce lists as they are most common.
If student has a weakness with a particular vowel, the list contains only that vowel in this part of the lesson.)
1. Teacher states the rule: “An e at the end tells us to say the name of this letter.”
2. Teacher guides students in applying the rule.
   a. Teacher points to game. “Is there an e at the end of this word?” (Signal.) “Yes.”
   b. Teacher points to a. “So we say the name of this letter.”
   c. “What’s the name of this letter?” (Signal.) “A”
   d. “Get ready to tell me the word.” Teacher pauses 2 seconds, then says, “What word?” (Signal.) “Game”
   e. Teacher repeats steps 2a to 2d with the remaining words.
3. Error correction involves repeating steps 2a to 2d and then returning to the beginning of the word list and re-presenting the words.
4. Once the student correctly pronounces all the words, the teacher proceeds to part 2.

Part 2

Teacher
Teacher writes on board: make, sit, hope, like, ram, hop.
The words are selected with one of each vowel (a, i, o) in a CVe pattern and one of each vowel in a CVC pattern. One or two minimally different pairs should be included. Words should not follow a pattern of presentation.
1. Teacher reminds student of the rule: “Remember, an e at the end of a word tells us to say the name (points to initial vowel) of this letter.”
2. Teacher guides students:
   a. Teacher points to make. “Is there an e at the end of this word?” (Signal.) “Yes”
   b. Teacher point to a in make. “Do we say /ā/ or /â/ for this letter?” (Signal.) “/â/”
   c. Teacher points to the left of make, pauses, then says, “What word? (S)” “Make”
   d. Teacher repeats step 2a to 2c with remaining words.
   e. Error correction occurs as in part 1.
3. Students read words without teacher guidance.
   a. “When I signal, tell me the word.”
   b. Teacher points to make, pauses 2 seconds, then asks, “What word?” (S) “Make”
   c. Teacher repeats step 3b with the remaining words.
   d. Error correction occurs as in part 1.

Lesson is complete when student can repeat the words correctly with no errors.
APPENDIX E
INTERVENTION FIDELITY FORM

Student _________________________  Teacher __________________  Day: M T W Th F

Date: __/__/____  Time of Lesson: __________________

Directions: Please complete this form after each lesson.

<table>
<thead>
<tr>
<th></th>
<th>Didn’t Get To It</th>
<th>Partly</th>
<th>Completely</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Part 1 of the lesson:**

<table>
<thead>
<tr>
<th></th>
<th>Didn’t Get To It</th>
<th>Partly</th>
<th>Completely</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Part 2 of the lesson:**

<table>
<thead>
<tr>
<th></th>
<th>Didn’t Get To It</th>
<th>Partly</th>
<th>Completely</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## APPENDIX F

### DETAILED IREAD INVENTORY SUB-SKILL RESULTS

**FOR ALL PARTICIPANTS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Participant 1-D</th>
<th>Grade</th>
<th>Second</th>
<th>Date of Assessment: Yr. 2007 Mo. 10</th>
<th>Age: Yr. 7Mo. 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Pre-Intervention Data

- **M CaC**
  - 7/10 CeC

- **M short a**
  - 5/10 short e
  - 7/10 short i
  - 6/10 short u

- **5/10 long a-Ce**
  - 2/10 long i-Ce
  - 3/10 long o-Ce
  - 1/5 long u-Ce
  - 7/10 R-cont a

- **7/10 R-cont o**
  - 1/5 R-cont e
  - 1/5 R-cont i
  - 1/5 R-cont u
  - 3/10 Cons-LE

- **M Long e EE**
  - 1/5 Long e EA
  - 0/5 Long e IE
  - 1/5 Long e EY
  - 1/5 Long a AI

- **0/5 Long e Y**
  - 0/5 Long a El(GH)
  - 0/5 Long o OA
  - 5/10 Long a AY
  - M Long u OO

- **5/10 Long i IGH**
  - 0/5 Long u OU
  - 1/5 Long i Y
  - M Long o OW
  - 1/5 Long u UE

- **0/5 Long u EW**
  - 1/5 OY Diphthong
  - M OU diphthong
  - 4/10 OO variant
  - 0/5 AW variant

- **0/5 OI diphthong**
  - M OW diphthong
  - 0/5 AU variant
  - 0/5 AIL variant
  - 4/10 L-cont A

34 combinations exhibit difficulty, 11 mastery

<table>
<thead>
<tr>
<th>Name</th>
<th>Participant 1-D</th>
<th>Grade</th>
<th>Second</th>
<th>Date of Assessment: Yr. 2008 Mo. 3</th>
<th>Age: Yr. 7Mo. 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Post-Intervention Data

(taught skills are bold with number of lessons in parenthesis, generalized improvement is in italics)

- **M CaC**
  - M CeC (3)

- **M short a**
  - M short e (2)
  - M short i
  - M short o (3)
  - M short u (6)

- **M long a-Ce (4)**
  - M long i-Ce (4)
  - M long o-Ce (4)
  - M long u-Ce (3)
  - M R-cont a (2)

- **M R-cont o (4)**
  - M R-cont e (3)
  - M R-cont i (4)
  - M R-cont u (2)
  - 5/10 Cons-LE

- **M Long e EE**
  - M Long e EA
  - M Long e IE (5)
  - 4/10 Long e EY
  - M Long a AI

- **M Long e Y**
  - 2/10 Long a El(GH)
  - M Long o OA
  - M Long a AY
  - M Long u OO

- **M Long i IGH**
  - 0/5 Long u OU
  - 2/10 Long i Y
  - M Long o OW
  - M Long u UE

- **5/10 Long u EW**
  - 7/10 OY Diphthong
  - M OU diphthong
  - 7/10 OO variant
  - 0/5 AW variant

- **M OI diphthong**
  - 7/10 OW diphthong
  - 0/5 AU variant
  - M AIL variant
  - 5/10 L-cont A

12 combinations exhibit difficulty, 33 mastery,
100% of taught skills improved (14), additional 9 skill areas mastered.
<table>
<thead>
<tr>
<th>Name</th>
<th>Date of Assessment Yr.</th>
<th>Grade</th>
<th>Age Yr.</th>
<th>Pre-intervention data</th>
<th>Post-intervention data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 2-L</td>
<td>2007 Mo 10</td>
<td>Second</td>
<td>7 Mo. 7</td>
<td></td>
<td>(taught skills are bold with number of lessons in parenthesis, generalized improvement is in italics)</td>
</tr>
<tr>
<td>M CaC</td>
<td>M CeC</td>
<td>M CiC</td>
<td>M CoC</td>
<td>M CuC</td>
<td>M short a</td>
</tr>
<tr>
<td>M long a-Ce</td>
<td>M long i-Ce</td>
<td>M long o-Ce</td>
<td>M long u-Ce</td>
<td>5/10 R-cont o</td>
<td>5/10 R-cont e</td>
</tr>
<tr>
<td>M Long e EE</td>
<td>M Long e EA</td>
<td>0/5 Long e IE</td>
<td>7/10 Long e</td>
<td>M Long a AI EY</td>
<td></td>
</tr>
<tr>
<td>3/10 Long e Y</td>
<td>0/5 Long a El(GH)</td>
<td>6/10 Long o OA</td>
<td>M Long a AY</td>
<td>M Long u OO</td>
<td></td>
</tr>
<tr>
<td>6/10 Long i IGH</td>
<td>0/5 Long u OU</td>
<td>1/5 Long i Y</td>
<td>2/10 Long o OW</td>
<td>6/10 Long u UE</td>
<td></td>
</tr>
<tr>
<td>0/5 Long u EW</td>
<td>M OY Diphthong</td>
<td>7/10 OU diphthong</td>
<td>4/10 OO variant</td>
<td>0/5 AW variant variant</td>
<td></td>
</tr>
<tr>
<td>0/5 OI</td>
<td>M OW diphthong</td>
<td>0/5 AU variant</td>
<td>M AIL variant</td>
<td>M L-cont A</td>
<td></td>
</tr>
<tr>
<td>diphthong</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22 combinations exhibit difficulty, 23 mastery</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Date of Assessment Yr.</th>
<th>Grade</th>
<th>Age Yr.</th>
<th>Post-intervention data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 2-L</td>
<td>2008 Mo 3</td>
<td>Second</td>
<td>7 Mo. 11</td>
<td>(taught skills are bold with number of lessons in parenthesis, generalized improvement is in italics)</td>
</tr>
<tr>
<td>M CaC</td>
<td>M CeC</td>
<td>M CiC</td>
<td>M CoC</td>
<td>M CuC</td>
</tr>
<tr>
<td>M long a-Ce</td>
<td>M long i-Ce</td>
<td>M long o-Ce</td>
<td>M long u-Ce</td>
<td>5/10 R-cont o</td>
</tr>
<tr>
<td>M R-cont o (2)</td>
<td>M R-cont e (4)</td>
<td>M R-cont i (2)</td>
<td>M R-cont u (2)</td>
<td>M Cons-LR</td>
</tr>
<tr>
<td>M Long e EE</td>
<td>M Long e EA</td>
<td>M Long e IE (9)</td>
<td>7/10 Long e EY</td>
<td>M Long a AI</td>
</tr>
<tr>
<td>M Long e Y</td>
<td>0/5 Long a El(GH)</td>
<td>M Long o OA</td>
<td>M Long a AY</td>
<td>M Long u OO</td>
</tr>
<tr>
<td>1/5 Long i IGH</td>
<td>4/10 Long u OU</td>
<td>2/10 Long i Y</td>
<td>M Long o OW</td>
<td>7/10 Long u UE</td>
</tr>
<tr>
<td>M Long u EW</td>
<td>M OY Diphthong</td>
<td>M OU diphthong</td>
<td>M OO variant</td>
<td>M AW variant</td>
</tr>
<tr>
<td>M OI diphthong</td>
<td>3/10 OW</td>
<td>1/5 AU variant</td>
<td>M AIL variant</td>
<td>5/10 L-cont A</td>
</tr>
</tbody>
</table>

10 combinations exhibit difficulty, 35 mastery, 100% of taught skills improved (6), additional 8 skill areas mastered.
Name: Participant 3-N  
Grade: Second  
Date of Assessment Yr. 2007 Mo. 10  
Age Yr. 8 Mo. 6

<table>
<thead>
<tr>
<th>Pre-intervention data</th>
<th>Date of Assessment Yr. 2007 Mo. 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>M CaC</td>
<td>7/10 CoC</td>
</tr>
<tr>
<td>M CeC</td>
<td>M CuC</td>
</tr>
<tr>
<td>M short a</td>
<td>M short e</td>
</tr>
<tr>
<td>M short i</td>
<td>M short o</td>
</tr>
<tr>
<td>M short u</td>
<td></td>
</tr>
<tr>
<td>6/10 long a-Ce</td>
<td>1/5 long i-Ce</td>
</tr>
<tr>
<td>1/5 long o-Ce</td>
<td>0/5 long u-Ce</td>
</tr>
<tr>
<td>0/5 R-cont o</td>
<td>6/10 R-cont i</td>
</tr>
<tr>
<td>0/5 R-cont u</td>
<td>M Cons-LE</td>
</tr>
<tr>
<td>3/10 Long e EE</td>
<td>1/5 Long e EA</td>
</tr>
<tr>
<td>0/5 Long e IE</td>
<td>7/10 Long e EY</td>
</tr>
<tr>
<td>2/10 Long a Al</td>
<td></td>
</tr>
<tr>
<td>5/10 Long e Y</td>
<td>0/5 Long a El(GH)</td>
</tr>
<tr>
<td>0/5 Long o OA</td>
<td>M Long a AY</td>
</tr>
<tr>
<td>3/10 Long i Y</td>
<td>M Long o OW</td>
</tr>
<tr>
<td>0/5 Long i IGH</td>
<td>6/10 Long i Y</td>
</tr>
<tr>
<td>6/10 Long u OU</td>
<td>0/5 Long u UE</td>
</tr>
<tr>
<td>2/10 Long u EW</td>
<td>M OY Diphthong</td>
</tr>
<tr>
<td>0/5 OI diphthong</td>
<td>M OW diphthong</td>
</tr>
<tr>
<td>0/5 AU variant</td>
<td>0/5 AIL variant</td>
</tr>
<tr>
<td>1/5 L-cont A</td>
<td></td>
</tr>
</tbody>
</table>

28 combinations exhibit difficulty, 27 mastery

Name: Participant 3-N  
Grade: Second  
Date of Assessment Yr. 2008 Mo. 3  
Age Yr. 8 Mo. 10

<table>
<thead>
<tr>
<th>Post-intervention data</th>
</tr>
</thead>
<tbody>
<tr>
<td>M CaC</td>
</tr>
<tr>
<td>M CeC</td>
</tr>
<tr>
<td>M short a</td>
</tr>
<tr>
<td>M short i</td>
</tr>
<tr>
<td>M short u</td>
</tr>
<tr>
<td>M long a-Ce (2)</td>
</tr>
<tr>
<td>M R-cont o (2)</td>
</tr>
<tr>
<td>6/10 long o-Ce (2)</td>
</tr>
<tr>
<td>5/10 long u-Ce (d)</td>
</tr>
<tr>
<td>7/10 Cons-LE</td>
</tr>
<tr>
<td>M Long e EE</td>
</tr>
<tr>
<td>5/10 Long e IE (3-i)</td>
</tr>
<tr>
<td>5/10 Long a Al</td>
</tr>
<tr>
<td>1/5 Long e Y</td>
</tr>
<tr>
<td>0/5 Long o OA</td>
</tr>
<tr>
<td>2/5 Long i Y</td>
</tr>
<tr>
<td>0/5 Long i IGH</td>
</tr>
<tr>
<td>M Long u EW</td>
</tr>
<tr>
<td>6/10 OU diphthong</td>
</tr>
<tr>
<td>1/5 OI diphthong</td>
</tr>
<tr>
<td>1/5 AIL variant</td>
</tr>
</tbody>
</table>

(d) = lessons discontinued due to a problem with the stimulus words.  
(3-1) = 3 lessons provided but study concluded prior to student showing mastery.  
25 combinations exhibit difficulty, 20 mastery, 38% of taught skills improved (3), additional 5 skill areas mastered.
REFERENCES


