

MEASURING THE EFFECT SIZE  
OF *TOOLBOX FOR LITERACY*

Paula Louchart

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Arts

Department of Psychology

Central Michigan University  
Mount Pleasant, Michigan  
February, 2011

Accepted by the Faculty of the College of Graduate Studies,  
Central Michigan University, in partial fulfillment of  
the requirements for the master's degree

Thesis Committee:

Michael D. Hixson, Ph.D.

Committee Chair

Jane Ashby, Ph.D.

Faculty Member

Sharon Bradley-Johnson, Ed.D.

Faculty Member

February 11, 2011

Date of Defense

Roger Coles, Ph.D.

Dean  
College of Graduate Studies

July 25, 2011

Approved by the  
College of Graduate Studies

•

This is dedicated to my family. To my husband and children who supported me, ate many frozen dinners and picked up the slack in household chores throughout this endeavor. I will always be grateful to those, who helped me with my own reading disability and consistently encouraged me to step out of my comfort zone and to always reach farther. To all my friends that waited for me while I secluded myself from the outside world.

## ACKNOWLEDGEMENTS

I wish to thank Dr. Michael Hixson, Thesis Committee Chair, for his contagious excitement, tireless enthusiasm and dedication to teaching struggling readers to read. His patience, support and advice created a safe environment for all students, young and not so young, to learn and grow. I have the deepest regard for his guidance and the many contributions he made toward my educational and personal growth.

In addition, I wish to thank the members of the ad hoc Thesis Requirements Committee: Dr. Jane Ashby and Dr. Sharon Bradley-Johnson for their many contributions to this study. These faculty members provided valuable direction when this project began, support during the data collection phase, reviewed the draft of the document. Throughout the process, these faculty members drew on their experiences of working with many graduate students as a faculty professor, thesis advisor and thesis or dissertation committee members.

I am grateful for the opportunity to work with the families and students involved in the Learning Acceleration Clinic. Thank you for the experiences that sparked my passion for reading interventions and working with struggling readers.

## ABSTRACT

### MEASURING THE EFFECT SIZE OF *TOOLBOX FOR LITERACY*

by Paula Louchart

The *Toolbox for Literacy* (Maloney, Brearley, & Preece, 2002) is a reading intervention designed to teach children to read using effective teaching methods. The pillars of this program are based on direct instruction, precision teaching and behavior modification. Although the components align with researched recommendations, it is important to determine if the components packaged together will sufficiently increase reading skills for struggling readers. Examining the effectiveness of a particular program with a controlled study is important to students, schools and the community because ineffective interventions waste precious resources, time and money. Therefore, *Toolbox for Literacy* was administered to eight students identified as struggling readers at the end of their first-grade year. A combination of general outcome measures and a norm-referenced test were used to measure gains in reading ability. Approximately 58 hours of intervention were provided in a one-to-one setting by either school psychology graduate or undergraduate psychology students. Results yielded significant increases in all five general outcome measures, and a moderate effect size on the norm-referenced test. The results of this small scale study indicate that *Toolbox for Literacy* is worthy of further evaluation using more rigorous empirical methods.

## TABLE OF CONTENTS

LIST OF TABLES .....	vii
CHAPTER	
I. INTRODUCTION.....	1
Problem Statement.....	1
II. REVIEW OF LITERATURE .....	3
Skills and Strategies of Good Readers.....	3
Research-based Teaching Methods .....	5
<i>National Reading Panel Recommendations</i> .....	5
Putting Research into Practice.....	8
<i>Direct Instruction</i> .....	8
<i>Precision Teaching</i> .....	9
Summer Reading Programs.....	11
III. METHODOLOGY .....	13
Participants and Setting.....	13
Independent Variables .....	14
<i>Toolbox for Literacy Program</i> .....	14
<i>Treatment Integrity</i> .....	15
Dependent Variables.....	16
Measurement Instruments .....	16
<i>Norm-referenced Test</i> .....	16
<i>General Outcome Measures</i> .....	18
Instructor Training .....	20
Initial Assessment.....	21
Intervention .....	22
IV. RESULTS .....	24
V. DISCUSSION .....	28
APPENDICES .....	32
REFERENCES .....	42

LIST OF TABLES

TABLE	PAGE
1. <i>easyCBM® Pre- and Post-test Scores</i> .....	26
2. <i>DIBELS® Next Pre- and Post-test Scores</i> .....	26
3. <i>Phonics-Based Reading Test Decoding Subtests Pre- and Post-test Scores</i> .....	27
4. <i>Mean Scores and Mean Differences for General Outcome Measures</i> .....	27
5. <i>Mean Quotients and Mean Differences for the Phonics-Based Reading Test</i> .....	27

# CHAPTER I

## INTRODUCTION

### Problem Statement

There has been an abundance of research in the field of reading for good reason. Reading is the foundation on which education is placed. Without a solid foundation, the acquisition of new material becomes increasingly difficult over time. Academic difficulties translate to a plethora of possible problem behaviors during the school years that follow the individual well into adulthood. Therefore, studies abound in reading; including identifying the skills good readers possess (Stanovich, 1991). However, these studies do not always directly translate to classroom procedures and teaching methods. The gap between research and practical application has to be bridged. For example, two early predictors of how well children will learn to read during the first two years of instruction are phonemic awareness and letter knowledge. Of these predictors only phonemic awareness instruction yields increased reading results (National Institute of Child Health and Human Development (NICHD), 2000). In terms of practice, it is important to determine whether evidence-based reading programs produce the desired outcomes for the target population. Intervention programs that are ineffective are costly to schools, students and the community.

In this study, the effects of *Toolbox for Literacy* (Maloney, Brearley, & Preece, 2002) for remediating reading difficulties in first-grade students during summer break were measured. There was no evidence found that this program has been previously evaluated. *Toolbox for Literacy* was designed to teach reading skills using the National Reading Panel (NRP)

recommendations. Direct instruction, precision teaching and behavior modification methods are the foundation of this program.

## CHAPTER II

### REVIEW OF LITERATURE

#### Skills and Strategies of Good Readers

The reading literature offers many studies, developmental theories and process models with a multitude of variations. In response to the confusing array of information produced by the different perspectives, Perfetti (1986) suggested concentrating on the skills good readers possessed instead of the mental processes of reading. From that stance it was easier to sort out which skills are important to teach. Research from around the world has helped develop a better understanding of what skills support reading proficiency (Carlisle & Stone, 2005; Cunningham, 1998; Perfetti, 1986). Several patterns emerged that encompassed various aspects of language development, phonological awareness, decoding, comprehension and fluency as they contribute to reading proficiency. Robust evidence supports the importance of these skills in the development of reading (e.g. Hart & Risley, 1995; Lonigan, Anthony, Phillips, Purpura, Wilson & McQueen, 2009; Snow, Burns, & Griffin, 1998).

Instructional sequences in reading have been developed by drawing from the similarities of language and reading development models. For example, Kellogg (1970) proposed a four-step model of language development. The child first learns to understand language by listening (receptive oral language), followed by speaking, (expressive oral language), which supports learning to read (receptive written language) and writing (expressive written language).

In comparison, Chall's (1983/1995) model of reading development described six stages related to skill development, but not necessarily age related. This model begins at Stage 0 when children first develop language skills (listening and speaking). Stages 1 and 2 refer to the earlier

stages of reading (letter-sound relationships and applying that knowledge to words and stories). The next three stages reflect the level of comprehension complexity from literal to inferential. Not all literate adults reach the highest stage of reading development. Only readers who search for relevant texts to serve a particular purpose reach this fifth stage of reading development.

Foundational skills of receptive and expressive oral language are crucial to the development of reading skills because reading and writing are intertwined with oral language development (e.g. Dickinson & McCabe, 2001; Hart & Risley, 1995; Leong, 1979). The importance of remediating skill deficits as early as possible is widely recognized (e.g., Adams, 1990; NICHD, 2000; Stanovich, 1986). Language deficits may be particularly difficult to remediate when gaps of a year or more exist. As part of Project Upgrade (U.S. Department of Health & Human Services, 2009), pre-literacy intervention programs were developed for pre-school children who were, on average, a year behind the national norm in oral language development. This intensive year-long program yielded modest increases in phonological awareness skills (the ability to identify and manipulate onset, rimes, phonemes and syllables) but only small effects for oral language remediation. It is evident as time passes, demand, practice and motivation widen deficit gaps in reading (Stanovich, 1986), as well as language skills.

Other pre-literacy skills such as book orientation, page turning, and pretend reading (i.e., saying words as they look at the pages) also contribute to reading development (Adams, 1990) and demonstrate a growing awareness of language patterns found in books (Dickinson & McCabe, 2001). Numerous studies from around the world (e.g. Carlisle & Stone, 2005; Leong, 1979; Moreno & Valenzuela, 2004; Valente and Martins, 2004) consistently report strong correlations between good readers and solid phonological and phonemic awareness skills. Fortunately, teaching phonemic awareness skills improve early reading skills (Bradley & Bryant,

1983) especially when combined with systematic phonological instruction (Adams, 1990; NICHD, 2000).

In order for students to meet the demands of higher level reading their decoding skills must evolve in order to address multisyllabic and morphologically complex words efficiently. Adams (1990) suggested that the ability to divide words into manageable parts develops as children are exposed to written words. Good readers pull together letters that typically go together, (e.g., /st/, and /dr/) and separate letters that do not typically go together in the same syllable, (e.g., /dn/). Chunking strategies are robust contributors to proficient reading in many languages including Dutch (Verhoeven, Schreuder, & Baayen, 2003), Spanish (Moreno & Valenzuela, 2004) and even Chinese (Taft & Zhu, 1995) which is different from an alphabetic system.

Fluent word identification skills give good readers an advantage over poor readers because they contribute substantially to reading comprehension (e.g. Adams, 1990; Carnine, Silbert, Kame'euni, & Tarver, 2004; Snow, Burns & Griffin, 1998; NICHD, 2000; Perfetti, 1986). Fluency is developed as readers practice and repeatedly encounter familiar words in text (Stanovich, 1986). Missing any one of the reading components (i.e., phonemic awareness, vocabulary, letter-sound correspondence, fluency, and comprehension) could result in a domino effect on reading ability (Perfetti, 1986).

## Research-based Teaching Methods

### *National Reading Panel Recommendations*

The National Reading Panel (NRP) compiled and analyzed the existing reading methods literature to provide the best possible information about which skills to teach, and which

instructional approaches effectively teach those skills. The relevant topic areas to *Toolbox for Literacy* (due to age of participants) were phonics instruction, fluency, guided oral repeated reading, and vocabulary instruction.

Phonics instruction teaches students to decode words based on letter-sound correspondences. Phonics instruction varies with the length of the word being taught. Beginning readers are taught to sound out small words letter by letter then blend, whereas, more advanced readers are taught to identify letter combinations, morphemes and larger word parts to decode multisyllabic and morphologically complex words (e.g., /con/, /struc/, /tion/). First grade students benefit significantly when phonics is taught systematically (in a developmentally appropriate scope and sequence) and explicitly (NCHID, 2000).

Fluency is recognized as an important reading skill, yet it is widely neglected in school instruction. When students spend so much mental energy on decoding, mental energy is robbed from comprehension and they fail to understand what they read (Perfetti, Goldman & Hoagboam, 1979). Rasinski (2009) suggested that effective reading fluency is taught by focusing on decoding accuracy, modeling, assisted reading (hear reading by a good reader), practice (including repeated reading), prosody and synergy (putting it all together in a meaningful way). Perfetti (1986) warned that instruction focused on fluency only is ineffective. Speed is a skill to be taught indirectly with accuracy training and viewed as a by-product of improved reading or word representation.

Guided oral repeated reading provides students with crucial repeated practice and immediate feedback. Combining repeated reading procedures with guided oral reading yielded fluency outcomes that were not evident with guided oral reading alone (NCHID, 2000). Immediate and frequent feedback reinforces accurate decoding and provides immediate

correction for all errors (Bos & Vaughn, 2006). When students receive systematic and explicit instructor feedback while reading out loud, their word recognition, fluency and comprehension increase significantly (Carnine, Silbert, Kame'euni & Tarver, 2004). Praise encourages the student's effort and is meant to reinforce behaviors conducive to attaining reading goals. Praise needs to be specific so that the student understands exactly what behavior is being reinforced and varied for sustainability.

Learning the definition of new words before encountering them in text improves vocabulary acquisition and text comprehension (NICHD, 2000), and weaving vocabulary instruction into the curriculum yields better results than isolating vocabulary instruction (U.S. Department of Health & Human Services, 2009). The NRP recommended that vocabulary instruction should include a variety of methods both direct and indirect. Instruction should be based on the reader's ability, age and needs. Repetitive exposure to new words in a variety of contexts increases word recognition.

Studies suggest that strengthening all these skills with a variety of teaching methods best supports comprehension, the ultimate goal of reading. Instructionally, no one method appears to be the magic bullet. In fact, the NRP recommended using a combination of several research-based methods. However, the use of direct instruction and frequent progress monitoring using data-driven assessments was consistently recommended for teaching all skill areas (NICHD, 2000).

The features of effective instructional methods are: 1) assessing progress frequently, objectively, and recording progress on charts or graphs; 2) designing instruction so it is goal-driven, adaptable, appropriately paced, and broken down into manageable steps and taught with a model-lead-test approach; 3) delivering instruction at a quick pace, allowing for ample

opportunity for frequent student responding and specific, immediate error correction procedures and feedback (Bos & Vaughn, 2003).

## Putting Research into Practice

### *Direct Instruction*

Direct Instruction programs, developed by Engelmann and colleagues, include many features of effective instruction. The instructional material is carefully sequenced and organized so it is not confusing to the student. Key concepts, rules and expectancies are repeated frequently to facilitate the mastery of concepts that are difficult to learn. Apart from Englemann's specific program, basic direct instruction methods are supported by an abundance of evidence (NICHD, 2000). By facilitating the acquisition of new material faster than traditional teaching methods, direct instruction methods improve learning outcomes in all children. They are particularly helpful for children with learning difficulties (Carnine, Silbert, Kame'enui, & Tarver, 2004) and a variety of at-risk students including those from low socioeconomic levels (Umbach, Darch & Halpin, 1989).

Researchers found direct instruction to be effective across curriculum areas such as vocabulary (Kieffer & Lesaux, 2007), math (Baker, Gersten & Lee, 2002), and especially reading (e.g., Shippen, Houchins, Steventon, & Sartor, 2005). Direct instruction demonstrated positive long-term effects that included increases in academic achievement (Ross, Nunnery, Goldfeder, McDonald, Rachor, Hornbeck, & Fleischman, 2004), moderate but consistent impact on college acceptance (Gersten, Keating, & Becker, 1988), increased reading motivation (Umbach, Darch & Halpin, 1989) and much higher reading levels than control groups (Gersten, Keating & Becker, 1988).

### *Precision Teaching*

Precision teaching is derived from Skinner's use of the cumulative recorder (used to record frequency of response) and Pavlov's use of frequency (used as a standard unit of measurement of observable behavior). Precision teaching is a sensitive indicator of learning and mastery as it is a system of charting observable, measurable behaviors and recording data that are directly relevant to instruction. It was developed so that decisions about instructional effectiveness could be made quickly and efficiently (Lindsley, 1991).

Monitoring progress by collecting and recording observable behavior provides objective data and preferable over the subjective judgment of performance rating scales. Teachers tend to overestimate student progress when traditional subjective methods are used as compared to using precision teaching methods (Graney, 2008). When a student's performance is frequently monitored this way, precious instructional time is not wasted on interventions and methods that are not productive (Fuchs & Fuchs, 2004). Student achievement outcomes increase when students and teachers are more accurately aware of the student's performance and instruction is changed in response to data (Fuchs, Deno & Mirkin, 1984).

Students have benefited from precision teaching in many learning situations. Precision teaching methods improved achievement for individuals with traumatic brain injuries (Chapman, Ewing & Mozzoni, 2005) as well as struggling readers (Houtveen & van de Graft, 2007). When combined with direct instruction methods, early elementary Dutch students made reading gains twice that of their peers whose performance was not systematically monitored and charted (Houtveen, & van de Grift, 2007).

The *Toolbox for Literacy* (Maloney, Brearley, & Preece, 2002) incorporates effective teaching methods, (i.e., direct instruction, precision teaching and behavior management) to teach

important areas of reading (phonics, vocabulary, fluency and comprehension). Instructors are guided through the model-lead-test format with scripted lessons that are carefully sequenced building upon and revisiting core concepts. The scripts require frequent student responding, the use of signals to immediately prompt responses, and specific feedback to immediately correct errors. The program specifies mastery criteria and provides standardized protocols to be followed in response to the student's performance on each reading task.

*Toolbox for Literacy* is divided into 4 levels designed for kindergarten through eighth-grade instruction. Each level contains 60 lessons that teach the most common words in the English language as compiled and cross-referenced from several lists available via web search (Maloney, 2010). Skills are taught in a sequence that corresponds to reading development stages (e.g. learning continuous sounds before short sounds). The learning objectives are divided into six tasks that teach a particular skill in each task. Each lesson provides review and the opportunity to preview new words, sounds and vocabulary before encountering them in the reading task as supported by research (Ebbers & Denton, 2008; NICHD, 2000). The lessons feature a mixture of fictional and non-fictional stories that are carefully structured and sequenced to match lesson objectives and become progressively longer and more difficult. The criterion for moving on to a new task specifies that the student should be able to complete the task quickly, completely and accurately before moving on. To ensure the student masters the material, lessons include timed fluency checks for sounds, words, and story reading. Timed assessments as such, are more consistent measures of reading skills (i.e., fluency and accuracy) than casual, untimed assessments (Perfetti, 1986).

*Toolbox for Literacy* lessons are identical to those in *Teach Your Children to Read Well* (Maloney, Brearley, & Preece, 2002). *Teach Your Children to Read Well* has been used in

previous years in the Learning Acceleration Clinic on the campus of Central Michigan University. The only differences between the two programs are the absence of workbook exercises in *Toolbox for Literacy* and the split of level 1 into two books (30 lessons in each volume of *Teach Your Children to Read Well*). The Learning Acceleration Clinic served a variety of students with reading difficulties during summer and after-school sessions. When pre- and post-test scores from a variety of norm-referenced tests were compared, the students demonstrated on average an effect size of 1.44. Although this may be considered a large effect size, (Hill, Bloom, Black & Lipsey, 2007) not much could be concluded from this datum because the clinic lacked the controls of an empirical study. The individual effect sizes varied with the students' age, initial reading levels, severity of reading deficits and how long they received instruction. However, the effect sizes are interesting and indicate a systematic investigation of the reading program was warranted because no other studies examining the effect size of *Toolbox for Literacy* were found.

### Summer Reading Programs

Over the last decade, summer programs have become increasingly popular because they offer academic remediation when students would otherwise experience loss or regression in knowledge and skill (Cooper, 2001). The regression is particularly noticeable in students from disadvantaged families and in the areas of math and reading. The most effective summer programs provide instruction in small groups or on an individual basis and focus on accelerated learning (Boss & Railback, 2002). The best time to intervene for reading deficits is during the summer before or after first grade because children are in the beginning stages of learning to read and deficit gaps are smaller (Denton & West, 2002).

Skill gaps can be narrowed or closed in as little as 8 hours of instruction in summer programs (Luftig, 2003); however the size of the gap and age of the student must be considered when designing summer programs. Research suggests that deficits become more difficult to remediate because gaps widen as students age, thereby requiring longer program hours. After 90 hours of remediation, the reading scores of incarcerated youth remained within or below the first percentile (Malmgren & Leone, 2000). Whereas, Cornelius and Semmel (1982) found that 24 hours of instruction during summer break was adequate for offsetting learning regression that often occurs in academically challenged and economically disadvantaged students. Considering these results, providing 60 hours of intensive reading instruction with an effective reading program, should be sufficient to yield significant reading gains.

The purpose of this study was to examine the effects of the *Toolbox for Literacy* when designed as a summer program to remediate deficits in students at the end of their first-grade year. As an evidence-based intervention, participants with reading difficulties were expected make significant reading gains after the implementation of the program. Reading gains were measured as score increases in phonemic awareness, reading accuracy (decoding), fluency and comprehension on general outcome measures as well as by performance on a norm-referenced test.

## CHAPTER III

### METHODOLOGY

#### Participants and Setting

Ten first-grade students ( $M = 7.6$   $SD = .52$ ) identified as struggling readers through school-wide screening at the end of the school year were recruited from the Mount Pleasant Public School district. The schools were instructed to send invitations to all students who scored below the 25<sup>th</sup> percentile, because these students would not have established benchmark. Benchmark indicates the lowest level of performance at which a student is likely to develop skills necessary to be a successful reader (Kaminski, Cummings, Powell-Smith, & Good, 2008). However, the schools did not share the actual screening scores for the participants and pre-test assessments revealed that some students scored in the average range on some of the measures used in the study.

Two participants did not complete the program. One student demonstrated interfering behaviors and chose to withdraw from the study. The other student did not attend after baseline data were collected. Therefore, their data were not included in the data set, leaving eight participants for the data set. The student population included three students who were receiving supplemental educational support services and six students who received free or reduced lunches. Seven of the students were Caucasian and one was mixed race. Three boys and five girls participated. Two students missed more than 20 days of school during the previous school year; the group median was 3.75 days.

Due to the limited number of participants recruited, a control group was not established. Two intervention sessions were offered. The first session began June 21, 2010 and the second

session began July 12. Participants were assigned to the first or second session according to the preference of their parent or guardian. Participants received approximately 58 hours of individual instruction, Monday through Thursday, 3 hours per day, for 6 consecutive weeks. Instruction was administered in private treatment rooms located in the Carls Center in the Health Professions Building on Central Michigan University's campus. The first-session group had one day less because of the July 4th holiday. There were no charges for the intervention sessions.

Instructors were recruited from undergraduate psychology classes and the school psychology graduate program at Central Michigan University. Instructors were offered the opportunity to take independent study credits for participating. Grades for those taking credits were based on performance as observed and rated on the treatment integrity form and a final paper summarizing the experience of working with their student.

### Independent Variables

#### *Toolbox for Literacy Program*

Participants were administered Level 1 (kindergarten to first grade) lessons from *Toolbox for Literacy* (Maloney, Brearley & Preece, 2002). The 60 lessons in Level 1 teach phonics and blending skills necessary to begin decoding. Oral reading practice was provided through word lists and sentences constructed from the previously taught words. Three hundred of the most common words in the English language are addressed by the end of Level 1. These words account for 66% of everything that is read (Maloney, Brearley, & Preece, 2002). Two books were used for this program, a Student Reader and the Instructor's Manual.

### *Treatment Integrity*

Treatment integrity data were collected weekly using the *Academic Therapist Observation Checklist-Revised* (Hixson, 2010; Appendix A). The observers were trained to implement the program to enhance the accuracy of rating. However, the raters did not have to meet specific criteria for reliability before beginning data collection. Observation data were collected via closed-circuit monitors. The closed-circuit monitors were located in each treatment room. The observers were able to switch between rooms without disturbing the instruction setting.

The observation checklist was divided into two sections because behaviors were recorded differently in each section. The first section was constructed with yes/no responses in relation to the observance of specific instructor behaviors (e.g. correct fluency administration, following scripts, and adhering to program criteria). The second section was constructed to record participant and instructor response frequency data during a 10-minute time frame (e.g., teacher reminders, praise, number of correct student responses).

Inter-rater reliability was collected three times during the study for all participants (5 percent of all data). The percentage of agreement for each section (teacher integrity and student response frequency) was calculated separately by dividing the smaller number of observations/frequency by the larger number then multiplying the dividend by 100. Agreement for the first section of the form was .94; and .86 on the response frequency section.

After data were collected, the researcher met with each instructor to discuss results and necessary improvements. Additional integrity checks were conducted if deemed necessary by the instructor's performance.

## Dependent Variables

Changes in reading ability were measured by comparing pre- and post-test scores from six reading assessment instruments that, when combined, measured the various skills associated with reading ability. The measures included: 1) the *Phonics-Based Reading Test* (PRT; Brownell, 2002) which yields a norm-referenced quotient for overall reading ability; 2) *Dynamic Indicators of Basic Early Literacy, Seventh Edition* (DIBELS® Next) Phoneme Segmentation Fluency (Good & Kaminski, 2009) measured phonemic awareness; 3) easyCBM® Letter Sounds (Alonzo, Tindal, Ulmer & Glasgow, 2006) measured letter-sound fluency; 4) DIBELS® Next Nonsense Word Fluency (Good & Kaminski, 2009) measured letter-sound fluency and blending skills; 5) easyCBM® Word Reading (Alonzo, Tindal, Ulmer & Glasgow, 2006) measured word recognition. and; 6) the median of three easyCBM® Passage Reading Fluency probes (Alonzo, Tindal, Ulmer & Glasgow, 2006) measured oral reading fluency (rate and accuracy).

## Measurement Instruments

### *Norm-referenced Test*

The PRT is comprised of three core subtests, Decoding, Fluency and Comprehension that yield a Total Reading composite ( $M = 100$ ,  $SD = 15$ ). For the first task, students are required to read aloud from a list of made-up words that represent specific word types (e.g., short vowel sounds, consonant blends, etc.). The Fluency subtest is comprised of two scores, accuracy and reading rate. As a student reads aloud, the number of errors and time it takes to read the passage are recorded. Their response to questions about the passages they read aloud comprises the Comprehension score.

Standardization data were collected from 1,282 participants between January 2001 and January 2002. The number of individuals tested at each age ranged from 102 to 130. Demographic characteristics of the sample were similar to the 2000 U.S. Census. The percentage of children with disabilities and those from urban/rural areas were similar to U.S. census data.

Reliability coefficients measuring internal consistency medians ranged from .92 to .98. Reliability coefficients were computed by age group for all individuals participating in the standardization study. Test-retest reliability data were collected from 37 examinees that were retested by the same examiner, an average of 20 days later. Correlations ranged from .87 to .95 on the subtests.

Content validity determines whether test contents cover representatively the behavior being measured (Anastasi & Urbina, 1997). The test employs a scope and sequence of phonics skills that are typically taught, as identified through a review of a number of supplemental phonics programs, basal reading tests, and professional resources. The passage reading tasks were constructed so they aligned with scope and sequence of phonics skills. The criterion-related correlations for the Total Reading quotient ranged from .61 to .83 (median of .77). The items appear unbiased in terms of sex, race, and ethnicity. Composite scores moderately correlate with the Stanford Achievement Test, Ninth Edition (SAT-9, 1996) and California Test Bureau Series (CTBS)/Terra Nova (1996). Construct validity coefficients ranged from .86 to .94 indicating items appear to align well the SAT-9 and CTBS/Terra Nova.

Item gradients on this measure are acceptable. Bracken (1987) suggests that item gradient problems exist when a change in raw score of one results in a change in standard score more than one-third of a standard deviation. Floor problems were noted in the Fluency (through

age 7-1) and Comprehension (through age 7-3) subtests. Floor problems indicate there are too few easy items for low performing children thereby overestimating their abilities. In this study, the floor problems were addressed by administering early literacy skill general outcome measures (Phoneme Segmentation Fluency, Nonsense Word Fluency and Letter-Sounds).

### *General Outcome Measures*

The general outcome measures used in this study are individually administered timed assessments. Twenty alternate forms are available for each measure within the same grade level and the measures ranged in difficulty from segmenting phonemes to reading passages aloud. First-grade level probes were selected for each measure. Benchmark assessments are typically administered three times per school year, beginning (September), middle (January), and end (May) of the school year.

DIBELS® Next data were collected in the middle of the year from thirteen schools across five states (N = Kindergarten: 27; First = 28). The DIBELS® Next Phoneme Segmentation Fluency (PSF) measure is administered orally. PSF benchmarks are typically administered at the middle and end of the kindergarten year and beginning of the first-grade year.

An assessor reads from a list of words and asks the student to say all the segments in each word presented. Students receive one point for each phoneme segmented correctly in 60 seconds. PSF measures were only evaluated on kindergarten students. Alternate-form reliability correlations were low (.44). However, three-form reliability was estimated at .70 using the given data and calculating with the Spearman-Brown Prophecy Formula. Inter-rater reliability was evaluated on first-grade students by shadow scoring. The reliability correlations were acceptable at .95. No test-retest data were reported (Powell-Smith, Good & Atkins, 2010).

DIBELS® Next Nonsense Word Fluency (NWF) probes measure letter-sound correspondence and blending sounds in randomly ordered VC and CVC nonsense words (e.g., vip, jek, hof). NWF assessments are typically administered beginning the middle of the kindergarten year through the beginning of the second-grade year. The total number of sound-letter correspondences correctly identified becomes the student's Correct Letter Sounds (CLS) score. The Whole Words Read (WWR) is the number of correctly blended pseudowords. NWF data were measured with first-grade students. Alternate-form reliability coefficients were acceptable on the Correct Letter Sounds (CLS) measure (.94) and better on the Whole Words Read (WWR) measure (.96). The estimated three-form reliability based on the Spearman-Brown Prophecy Formula was .90 for NWF/CLS and .88 for NWF/WWR. Test-retest correlations were .76 for NWF/CLS and .70 for NWF/WWR using a two-week interval. Inter-rater reliability coefficients were .99 for both measures. Criterion-related validity for NWF/CLS with GRADE Total Test ranged from .43 to .56. Criterion-related validity for NWF/WWR with GRADE Total Test ranged from .30 to .56. No other data are available on DIBELS® Next measures (Powell-Smith, Good & Atkins, 2010).

The easyCBM ® measures were tested in the spring of 2009 on first-grade classes from a mid-sized K-8 Pacific West school (N = 48). To administer Letter-Sounds, students are presented with randomly arranged upper and lower case letters along with some common digraphs (e.g., ph, th, sh, etc.). Students were given 30 seconds for each measure. However, instructions indicate that students are given 60 seconds to name as many letter sounds as they can while reading across the paper from left to right. Alternate-form reliability was tested at the 1:3 (first grade, third month) and 1:5 (first grade, fifth month) levels. Results ranged from .76 to .88.

Test stability for the first grade measures ranged from .64 to .68 on a one-week interval (Alonzo & Tindal, 2009).

The easyCBM® Word Reading Fluency is a measure of decoding and sight word knowledge. Word Reading Fluency probes are available for grades 1 through 3. Students are presented with a list of decodable and sight-words and are instructed to read words aloud. Errors and skipped words are counted as incorrect while self-corrections and words read correctly are counted as correct. The student receives one point for every correct response in 60 seconds. The results yielded strong correlations for alternate-form reliability ranging from .95 to .97, and test-retest reliability at a one-week interval ranging from .94 to .95 (Alonzo & Tindal, 2009).

Oral reading fluency was measured using easyCBM® Passage Reading Fluency probes. Passage Reading Fluency probes are administered on an individual basis under timed conditions. Passage Reading Fluency probes are available for grades 1 through 8. As the student reads a passage the examiner follows along on the examiner's copy marking errors and the last word read. Alternate-form correlations were .95 to .97. Test-retest for Passage Reading Fluency measure was evaluated with a one-week interval. Performance was correlated on two forms of the measure in two testing sessions with a result of .96 to .97 (Alonzo & Tindal, 2009). (

### Instructor Training

Instructors attended four half-day training sessions to learn the *Toolbox for Literacy* program (Appendix B). Instructors were familiarized with the Instructor's Manual that outlines fluency and mastery criteria for each task, lesson advancement or review, and how to use the charts to record data. Training manuals that were provided by the company were used to train

the instructors. The manuals highlighted and explained the different tasks that were to be mastered before implementing the program. The first several tasks in the Instructor's Manual are very detailed and become shorter as the instructor and student become more familiar with the format. Methods and tasks were modeled for the instructors and opportunity for practice with each other was provided. Instructors were taught how to effectively use signaling, praise, and pacing. Some of the letter-sounds that are traditionally taught as short sounds (e.g., *y* and *w*) are taught as continuous sounds in the program. A short sound (e.g., *b* or *c*) can not be held, whereas continuous sounds (e.g., *m* or *n*) can be held for as long as one has breath making them easier to blend than short sounds. The program CDs were used to help teach the all the letter-sounds to the instructors. Instructors were expected to master the delivery of all tasks and error correction procedures in a quick and fluent manner before they began teaching to ensure instructional consistency. Criteria for mastery were delivering the silent e rule lesson in less than 3 minutes and passing letter-sound fluency that was expected of the student (30 correct sounds in 30 seconds).

The instructors had to meet specific mastery criteria before working with the students. All instructors completed the online Health Insurance Portability and Accountability Act (HIPAA) training prior to working with students as required by federal regulations.

#### Initial Assessment

All measures were administered by school psychology graduate students from the university in private treatment rooms at the Carls Center in the Health Professions Building. Participants were assessed with the PRT and general outcome measures to establish a pre-test measure the week before the start of the program. In addition, the placement test on the back of

the *Toolbox for Literacy Student Reader* was administered to determine what the appropriate instructional level and starting point was for each participant.

### Intervention

Intervention sessions were held Monday through Thursday from 9 to 12 each morning. Instructors followed the scripted lessons and correction procedures directly from the Instructor's Manual of *Toolbox for Literacy*. Lessons are grouped into a set of five lessons. New fluency lists (i.e. sounds, words and stories) are provided for each new lesson set. The criteria for sound and word fluencies were 30 sounds or words in 30 seconds with two or fewer errors. The fluency criteria indicated whether lessons or tasks needed to be reviewed or skipped. The participants were given at least three opportunities to pass fluency checks during each lesson. If the fluency criterion was not met within five lessons, the instructor reviewed the five lessons, starting with the first lesson of the set, until the criteria was met. When criteria were met on all three fluency measures, the instructor proceeded to the next lesson set. When fluency criterion was met during the first three or four lessons of a set, the fluency check was skipped for the remainder of the lessons in that set. The instructors recorded fluency scores and words with which the student struggled for additional practice. Behavior modification methods are woven into *Toolbox for Literacy* by prompting frequent responses, delivering appropriate and frequent feedback and pacing instruction to the student's needs (Maloney, Brearley & Preece, 2002). In addition, the program provides a reward system option in which points that the student earns for completing tasks appropriately are charted and can be redeemed for pre-arranged tangibles or privileges. This option was used for students who exhibited interfering behaviors that were not

remedied by instructional methods (i.e. frequent responding, praise, and feedback). This option was used for two students in the program.

## CHAPTER IV

### RESULTS

Data for the pre- and post-test measures appear in Table 1. Data were aggregated using SPSS Graduate Pack 16.0 for Windows (2007) and analyzed using paired-sample one-tailed *t*-tests for each measure. One-tailed tests were used because this study evaluated reading gains. The participants scored significantly higher on all general outcome measures after the program.

For the Phoneme Segmentation Fluency (PSF) pre-test the beginning of first-grade year norms were used to gauge performance of the sample because norms for later in the year were not reported. The benchmark for the beginning of the first-grade year was 40. The mean pre-test PSF score of the sample was 43.38 with a standard deviation of 12.92. Four students met or exceeded benchmark on this measure upon pre-test and all students met benchmark at post-test ( $M = 58.38, SD = 16.77$ ). The average number of phonemes identified per minute was significantly higher after the reading program than before the program,  $t(7) = 4.09, p = .002$ .

The mean pre-test Letter-Sound Fluency (LSF) score for this sample was 32.25 with a standard deviation of 4.59. In comparison, the average student in the spring of their first-grade year reads 44 letter sounds per minute. No student from the current sample met benchmark on this measure, whereas all but one student met benchmark at post-testing ( $M = 58.88, SD = 16.02$ ). The student who failed to meet benchmark on this measure missed benchmark by one point. The participants read significantly more letter-sounds per minute after the program than before the reading program,  $t(7) = 4.99, p < .001$ .

Nonsense Word Fluency (NWF) yielded two measures, one reporting the number of letter-sound combinations read in the correct sequence (CLS) per minute and the number of letter-sound combinations that were correctly blended into pseudowords per minute (WWR).

The mean of the pre-test scores on the Nonsense Word Fluency (Correct Letter Sequences) was 49.12 ( $SD = 9.72$ ). The benchmark for students in the beginning of the first-grade year was 54. Only two students from the current sample met or exceeded benchmark upon pre-test and seven met this benchmark on post-test ( $M = 65.00$ ,  $SD = 21.88$ ). The mean of Whole Words Read (WWR) was 8.62 ( $SD = 7.72$ ) for the current sample at pre-test compared to the 13 WWR benchmark. Three students met this benchmark on the pre-test and four met the benchmark on the post-test ( $M = 14.88$ ,  $SD = 10.20$ ). Again, beginning of year norms were used because later norms were not reported for these measures. There was a significant increase on both measures after implementation of the reading program, NWF/CLS;  $t(7) = 2.62$ ,  $p = .02$ ; NWF/WWR;  $t(7) = 2.09$ ,  $p = .04$ .

The sample pre-test scores on the Word Reading Fluency (WRF) yielded a mean of 24.75 with a standard deviation of 11.81. In comparison, the average reader at the end of the first-grade year reads 42 words per minute. One student performed as well as or better than the 50<sup>th</sup> percentile on pre-test and five students met or exceeded this goal at post-test ( $M = 38.75$ ,  $SD = 19.91$ ). The participants read significantly more words per minute from the *easyCBM*® Word Reading Fluency probes after the program,  $t(7) = 4.33$ ,  $p = .001$ .

Each participant was administered three *easyCBM*® Passage Reading Fluency probes pre- and post-test. The median of those three probes is represented in this data set. One student met or exceeded the 50<sup>th</sup> percentile goal from the norm sample (58) on the Passage Reading Fluency (PSF) measure at the time of the pre-test ( $M = 27.00$ ,  $SD = 18.39$ ). Upon post-testing, two students meet or exceeded the 50<sup>th</sup> percentile ( $M = 35$ ,  $SD = 23.50$ ). There was a significant increase in the number of words read correctly per minute after the implementation of the reading program,  $t(7) = 2.09$ ,  $p = .04$ .

All but one student scored within the average range on the Phonics-Based Reading Test at the time of pre-testing ( $M = 91.57, SD = 5.44$ ) and one student scored within the below average range at post-test ( $M = 93.86, SD = 8.13$ ). There were significant increases on the Decoding subtest,  $t(7) = 2.17, p = .03$ <sup>1</sup>.

Table 1. *easyCBM® Pre- and Post-test Scores*

Participant	LSF 50 <sup>th</sup> Percentile Cut Score: 44		WRF 50 <sup>th</sup> Percentile Cut Score: 42		PRF 50 <sup>th</sup> Percentile Cut Score: 58	
	Pre-	Post-	Pre-	Post-	Pre-	Post-
1	23	43	16	21	20	27
2	33	86	31	54	36	65
3	38	48	17	25	9	9
4	36	72	25	44	25	45
5	33	73	48	77	64	70
6	29	48	9	19	6	9
7	34	45	24	27	21	20
8	32	56	28	41	35	35

Table 2. *DIBELS® Next Pre- and Post-test Scores*

Participant	PSF Benchmark: 40		NWF/CLS Benchmark: 54		NWF/WWR Benchmark: 13	
	Pre-	Post-	Pre-	Post-	Pre-	Post-
1	57	57	44	48	15	3
2	66	83	45	68	15	19
3	33	40	53	56	0	2
4	36	51	55	93	9	20
5	50	83	69	98	19	32
6	41	50	41	54	0	14
7	32	46	47	34	0	8
8	32	57	39	69	11	21

<sup>1</sup> The Total Reading Composite, Fluency and Comprehension subtests were administered but did not have an adequate floor. Only the Decoding subtest had an adequate floor for this age group of students.

Table 3. *Phonics-Based Reading Test Decoding Subtest Pre- and Post-test Scores*

Participant	Pre-test	Post-test
1	99	101
2	90	98
3	91	97
4	95	100
5	97	105
6	94	97
7	99	94
8	91	91

Table 4. *Mean Scores and Mean Differences for General Outcome Measures*

Measure	Pre-test	Post-test	Difference	Standard Deviation of Mean Difference
PFS	43.38	58.38	15.00**	10.38
LSF	32.25	58.88	26.63**	15.08
NWF/CLS	49.12	65.00	15.88*	17.16
NWF/WWR	8.62	14.88	6.25*	8.46
WRF	24.75	38.75	14.00**	9.15
PRF	27.00	35.00	8.00*	10.85

\* $p < .05$ . \*\* $p < .01$ .

Table 5. *Mean Quotients and Mean Differences for the Phonics-Based Reading Test*

Sub-test	Pre-test	Post-test	Difference	Standard Deviation of Difference
Decoding	94.50	97.88	3.38*	4.41

\* $p < .05$ .

## CHAPTER V

### DISCUSSION

Overall, results indicate significant improvement on all general outcome measures, which are, by design, sensitive to change in the skill areas they assess. Non-significant results on the overall PRT reading composites are not surprising because norm-referenced tests are less sensitive to change and the test's age range, the overall result did not have an adequate floor for these students. Significant results were found for the PRT Decoding subtest which was the only subtest that had an adequate floor. Inadequate floors occur when there are not enough easy items on a test for children at the lowest age levels thereby overestimating a student's performance. An adequate floor extends down at least two standard deviations below the mean with a raw score of one (Bracken, 1987). Perhaps the other subtests did not indicate significant results because the scores at the time of pre-testing were an overestimate of the students' abilities.

Other problems with this study were the small sample size, large variability and possible limited experience of instructors in managing children. The variability in some cases was larger than the difference between tests. Instructors who are better prepared to deal with student behaviors such as special education teachers or academic therapists may obtain better results. Although the instructors were supported by experienced school psychologists, there would be potentially different outcomes if the students were managed by experienced instructors. For example, one student exhibited behavior issues that interfered with her performance on the PRT during the post-test. She scored well within the average range on the pre-test and below average (significantly lower when using the 95 percent confidence interval) on the post-test. Having a larger sample size and having experienced instructors manage the student's behaviors would potentially offset this sort of anomaly.

The program is advertised as a home-schooling or home-intervention tool and not meant to be a stand-alone (basal) program to teach reading (Maloney, 2010). In this regard, the scripted lessons and detailed instructions would be beneficial to parents or caregivers that implement this program at home because it guides those who do not have an educational background through the process of using effective teaching methods. Parents are likely to not receive formal training and need to rely solely on the products offered for training. Training materials including CDs to develop consistent letter sound awareness and skills are available. In addition, program support is offered via telephone by Mike Maloney and his staff at no additional charge.

Comments from the instructors were collected at the end of the program regarding the ease of implementation and student receptiveness to the instruction. In general, the instructors appreciated the scripted format of the program, yet some found errors in the scripts relating to page number references between the teacher's manual and the student reader. Most students found repeating lessons aversive when they did not meet their goals. Repeating lessons and performing fluency tasks seemed to be what triggered most of the behavior issues that were presented. Most instructors felt the story reading fluency goal was set too high for struggling readers. They were also surprised when their students reached their goals. The issue regarding the story reading goal of 125 words per minute can be controversial because a student who is in a program for a limited amount of time and has to review lessons repeatedly misses out on exposure to other letter-sound combinations. The counter argument is that it does not make sense to proceed when the student has not fully mastered what has been taught. The defining issue then becomes: What speed of repeated reading constitutes mastery? When this program was used in the Learning Acceleration Clinic at CMU, in one case the story reading fluency goal was lowered so the student was exposed to more material during the summer sessions. As the

student progressed and repeatedly met the adjusted goal, the goal was raised until the student met the goal set in the program. Evidence from the student's pre- and post-test scores suggested this process was equally successful.

Future studies should consider using phoneme segmenting and nonsense word fluency measures from a source other than DIBELS Next® if the participants are from first-grade. Norms for first-grade students are not collected beyond the fall benchmark period for phoneme segmenting fluency and Nonsense Word Fluency measures. Comparing the sample to these norms is problematic because it is likely to overestimate the student's ability. Phoneme segmenting measures from easyCBM® would be a good option because spring norms are provided for first-grade students and alternate-form reliability coefficients are stronger than DIBELS Next ® probes.

A norm-referenced test that offers alternate forms should also be considered so that the participants have relatively equal but different test items between pre- and post-testing to eliminate possible practice effect or it could be used if testing is spoiled and re-testing is necessary. This would have been particularly helpful in the case of the student who demonstrated interfering behaviors. Re-testing after behavior issues were resolved may have produced more accurate results in this study.

Results indicate the *Toolbox for Literacy* reading program effectively reduced reading deficits for these first-grade students who did not achieve spring benchmark reading goals at school. The results also indicate the reading program is worthy of further examination using a larger sample of first-grade struggling readers and a control group. The study should also include random assignment (after all names are collected and shuffled). A third group could be

added in which students are given time with an instructor who does not instruct reading (i.e., teaches math or just spends time with the students).

This program appears worthy of pursuing with a controlled study. As indicated earlier, examining the effectiveness of a particular program with a controlled study is important to students, schools and the community. Although the components align with researched recommendations, it is important to determine if the components that are packaged together will sufficiently increase reading skills for struggling readers. *Toolbox for Literacy* appears to be a promising program for teaching important literacy skills to beginning readers.

## APPENDICES

APPENDIX A

ACADEMIC THERAPIST OBSERVATION CHECKLIST

Academic Therapist: \_\_\_\_\_ Date: \_\_\_\_\_

Student: \_\_\_\_\_ Evaluator: \_\_\_\_\_

Rating scale

Y = observed N = not observed, not consistently observed

**1. Starts within 2 minutes of class time**

Y N

**2. Reads instructions from manual for each task**

Y N

**3. Teaches at a good pace**

Y N

**4. Uses correct procedures for reading words**

Y            N

**5. Catches 100% of errors and uses error correction procedure properly**

Y            N

**6. Correct fluency administration**

Y            N

**7. Gives specific praise = 2+/min.**

Y            N

**8. Transitions between lesson parts < 1 min.**

Y            N

**9. Behavior Report card completed and followed 100%**

Y            N

**10. Before each activity, AT discusses activity and learning skill and academic outcomes  
(Behavior Report Card)**

Y            N

**11. After each activity, AT gives student feedback about learning skills and academic  
achievement (Behavior Report Card)**

Y            N

**12. Talks no more than 20 seconds before student response**

Y            N

**13. Helps student to stay on task**

Y            N

**14. Student's materials were well organized**

Y            N

**15. Student 100% correct before moving to next activity**

Y                      N

**16. Student does at least 3 timings, if daily or ultimate goal not met**

Y                      N

**17. AT asks for help from Manager when student does not reach goal within 3 lessons and when 3 data points below goal line on CBM or DIBELS graph (check data)**

Y                      N

**Student**

**1. Talk-outs (10 min. observation)**

< 6	< 4	< 2
1	2	3

**2. Teacher reminders <2 per 10 minutes**

Y            N

**3. Student plots data on chart or writes data in book immediately after the activity.**

Y            N

**4. On task more than 90% of the time**

Y            N

**Response rate > 10/min. (\*Total student responses/by 10)**

Y            N

**Notes/Comments:**

## Recording Form

Typically for a 10 minute observation during the teacher presentation portion of the lesson.

*First time corrects	
*First time errors	
*Correct after first time	
*Error after first time	
Specific Praise	
General Priase	
Talk Outs	
Teacher Reminders	

First time correct = correct student response upon first presentation of an item (each new word, each lead or test)

First time errors = incorrect student response upon first presentation of an item (each new word, each lead or test)

Correct after first time = correct student response on an item student had just responded to incorrectly

Error after first time = incorrect student response on an item student had just responded to incorrectly

Specific praise = a general praise statement combined with a statement specifying why the praise is being given (e.g., good reading, I like how you sounded that out, yes, s-a-t)

General Praise = praise statement that is generic in nature (e.g., good job, nice work, that's it, alright)

Talk outs = any student statement besides a student response (e.g., I'm tired, what was that?, I have to use the bathroom)

Teacher reminders = teacher statement aimed at getting student on task (e.g., look here, pay attention, watch me, it is time to do work)

APPENDIX B  
INSTRUCTOR TRAINING SYLLABUS

**June 7 – 10, 2010**

**Instructor Training Meetings:**

Monday through Thursday, 9:00 a.m. to 12:00 p.m.

Health Professionals Building room 2141

Paula Louchart (989) 860-5108; Mike Hixson (989) 774-6462

**Course Objectives:**

1. Master letter sounds and sound combinations that are taught to students (60 sounds per minute with less than 2 errors or “learning opportunities.”)
2. Deliver tasks accurately and fluently (5-steps; 0-2 errors, in 4 minutes)
3. Recognize and use appropriate pacing
4. Use chart to track and monitor fluency performance

**Training Outline:**

June 7 Review syllabus; HIPAA training information; orientation to effective teaching methods, correction procedures, signals, teaching letter sounds and sound combinations, introduce tasks 1 – 5.

\*Please read through tasks 6-10 and academic therapist observation form for Tuesday.

June 8 Review and practice tasks 1-5; introduce tasks 6-10; review academic therapist observation form. This form will be used in conjunction with mastery criteria to test instructors.

\*Please read through tasks 11-15 for Wednesday

June 9 Review and practice tasks 6-10; introduce tasks 11-15

June 10 Review and practice tasks 11-15; Mastery tasks testing

June 17 Review; initial student/room assignments;

\*Complete HIPAA requirements before meeting this day.

### **Evaluation:**

Before working with student participants, instructors are required to:

1. Meet mastery criteria for delivering instruction and performing fluency tasks
2. Complete HIPAA training and pass the online test (submit results to Carla).

### **Dress Code:**

Carls Center is a professional service provider. Although professional dress is not required, they request adults to dress appropriately and respectfully (no jeans, cutoff shorts, tank tops, short shorts, etc.)

## REFERENCES

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Alonzo, J., & Tindal, G. (2009). *Alternate form and test-retest reliability of easyCBM® reading measures* (Technical Report No. 0906). Eugene, OR: University of Oregon Behavior Research & Teaching, Retrieved from <http://easycbm.com>.
- Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). *easyCBM® online progress monitoring assessment system*. Eugene, OR: University of Oregon, Center for Educational Assessment Accountability. Retrieved from <http://easycbm.com>.
- Anastasi, A., & Urbina, S. (1997). *Psychological Analysis* (7<sup>th</sup> ed.). Upper Saddle River, NJ: Prentice Hall.
- Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *Elementary School Journal*, 103, 51-73.
- Bos, C. S., & Vaughn, S. (2006). *Strategies for teaching students with learning and behavior problems*, (6<sup>th</sup> ed.). Boston: Pearson.
- Boss, S., & Railback, J. (2002). *Summer school programs: A look at the research, implications for practice, and program samples*. Washington, DC: Office of Educational Research and Improvement (ED).
- Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment*, 4, 313-326.
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read – a causal connection. *Nature*, 301(3), 419 – 421.
- Brownell, R. (2002). *Phonics-Based Reading Test*. Novato, CA: Academic Therapy Publications.
- Carlisle, J., & Stone, M. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly*, 40, 428-449.
- Carnine, D. W., Silbert, J., Kame'enui, E. J., & Tarver, S. G. (2004). *Direct instruction reading*, (4<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Chall, J. S. (1983/1995). *Stages of reading development*. (2<sup>nd</sup> ed.). Orlando, FL: Harcourt Brace College Publishers.

- Chapman, S. S., Ewing, C. G., & Mozzoni, M. P. (2005). Precision teaching and fluency training across cognitive, physical, and academic tasks in children with traumatic brain injury: A multiple baseline study. *Behavioral Interventions*, 20, 37 – 49.
- Cooper, H. (2001). Summer school: Research-based recommendations for policymakers. *SERVE Policy Brief*. Retrieved December 20, 2009, from <http://www.SERVE.org/Publications/Sumschool.pdf>.
- Cornelius, P. L., & Semmel, M. I. (1982). Effects of summer instruction on reading achievement regression of learning disabled students. *Journal of Learning Disabilities*, 15(7), 409-413.
- Cunningham, P. (1998). The multisyllabic word dilemma: Helping students build meaning, spell, and read “big” words. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 14, 189-218.
- Denton, K., & West, J. (2002). *Children’s reading and mathematics achievement in kindergarten and first grade*. Washington, DC: US Department of Education.
- Dickinson, D. K., & McCabe, A. (2001). Bringing it all together: The multiple origins, skills and environmental supports of early literacy. *Learning Disabilities Research & Practice*, 16, 186-202.
- Ebbers, S. M., & Denton, C. A. (2008). A root awakening: Vocabulary instruction for older students with reading difficulties. *Learning Disabilities Research & Practice*, 23, 90-102.
- Fuches, L. S., Deno, S. L., & Mirkin, P. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement and student awareness of learning. *American Educational Research Journal*, 21(2), 449-460.
- Fuchs, L. S., & Fuchs, D. (2004). Determining adequate yearly progress from kindergarten through grade 6 with curriculum-based measurement. *Assessment for Effective Intervention*, 29, 25-37.
- Gersten, R., Keating, T., & Becker, W. (1988). The continued impact of the direct instructional model: Longitudinal studies of follow through students. *Education and Treatment of Children*, 11, 318-327.
- Good, R.H., & Kaminski, R.A., (Eds.). (2009). *Dynamic Indicators of Basic Early Literacy Skills* © NEXT. Eugene, OR: Dynamic Measurement Group. Available at <http://www.dibles.org>.
- Graney, S. B. (2008). General education teacher judgments of their low-performing students’ short-term reading progress. *Psychology in the Schools*, 45(6), 537-549.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore: P. H. Brookes.

- Hill, C. J., Bloom, H. S., Black, A. B., & Lipsey, M. W. (2007). Empirical benchmarks for interpreting effect sizes in research. *MDRC Working Papers on Research Methodology*
- Hixson, M. D. (2010). *Academic therapist observation checklist - Revised*. Unpublished document, Central Michigan University, Department of School Psychology.
- Houtveen, T., & van de Grift, W. (2007). Reading instruction for struggling learners. *Journal of Education for Students Placed At Risk, 12*, 405-424.
- Kaminski, R., Cummings, K. D., Powell-Smith, K.A., & Good, R.H.III (2008). Best practices in using Dynamic Indicators of Basic Early Literacy Skills for formative assessment and evaluation. In A. Thomas and J. Grimes (eds.) *Best practices in school psychology V*. Bethesda, MD: National Association of School Psychologists.
- Kellogg, R. (1970). *Analyzing children's art*. Palo Alto, CA: National Press Books.
- Kieffer, M. J., & Lesaux, N. K. (2007). Breaking down words to build meaning: Morphology, vocabulary, and reading comprehension in the urban classroom. *The Reading Teacher, 61*, 134-144.
- Leong, C. K. (1979). Children's concepts of language in learning to read. *Bulletin of the Orton Society, 29*, 115-128.
- Lindsley, O. R. (1991). Precision teaching's unique legacy from B.F. *Journal of Behavioral Education, 1*, 253-266.
- Lonigan, C. J., Anthony, J. L., Phillips, B. M., Purpura, D. J., Wilson, S. B., & McQueen, J. D. (2009). The nature of preschool phonological processing abilities and their relations to vocabulary, general cognitive abilities, and print knowledge. *Journal of Educational Psychology, 101*, 345-358.
- Luftig, R. L. (2003). When a little bit means a lot: The effects of a short-term reading program on economically disadvantaged elementary schoolers. *Reading Research and Instruction, 4*(23), 1-13.
- Maloney, M., Brearley, L., & Preece, J. (2002). *Toolbox for literacy: Instructor's manual*. Bellesville, ONT: Teach Your Children Well Press.
- Maloney, M. (2010). Personal communication. June 18, 2010.
- Malmgren, L. W., & Leone, P. E. (2000). Effects of a short-term auxiliary reading program on the reading skills of incarcerated youth. *Education and Treatment of Children, 23*(3), 239-247.
- Moreno, M. D. P., & Valenzeula, M. J. G. (2004). Development of phonological awareness, reading experience and task difficulty. *Revista de Logopedia, Foniatria y Audiologia, 24*, 2-15.

National Institute of Child Health and Human Development. (2000). *Report of the national reading panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. (NIH publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Perfetti, C. A. (1986). Continuities in reading acquisition, reading skill and reading ability. *Remedial and Special Education*, 7, 11-21.

Perfetti, C. A., Goldman, S., & Hoagboam, T. (1979). Reading skill and the identification of words in discourse context. *Memory & Cognition*, 7, 273-282.

Powell-Smith, K. A., Good, R. H., & Atkins, T. (2010). *DIBELS Next*® *Oral Reading Fluency Readability Study* (Tech. Report No. 7). Eugene, OR: Dynamic Measurement Group.

Rasinski, T. (2009). Effective teaching of reading: From phonics to fluency. Chicago: National Reading Conference. Retrieved October 18, 2009, from <http://www.timrasinski.com/?page=presentations>.

Ross, S. M., Nunnery, J. A., Goldfeder, E., McDonald, A., Rachor, R., Hornbeck, M., & Fleischmann, S. (2004). Using school reform models to improve reading achievement: A longitudinal study of direct instruction and success for all in an urban district. *Journal of Education for Students Placed at Risk*, 9(4), 357-388.

Shippen, M. E., Houchins, D. E., Steventon, C., & Sartor, D. (2005). A comparison of two direct instruction reading programs for urban middle school students. *Remedial and Special Education*, 26, 175-182.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

*Stanford Achievement Test, Ninth Edition*. (1996). San Antonio, TX: PsychCorp

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 4.

Stanovich, K. E. (1991). Word recognition: Changing perspectives. In R. Barr, M. Kamil, P. Mosenthal & P. D. Pearson, (Eds.), *Handbook of reading research Vol. 2*, pp. 418-452). New York: Longman.

Taft, M., & Zhu, X. (1995). The representation of bound morphemes in the lexicon: A Chinese study. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 293-316). Hillsdale, NJ: Erlbaum.

*Terra Nova*. (1996). Monterey, CA: CTB/McGraw Hill.

Umbach, B., Darch, C., & Halpin, G. (1989). Teaching reading to low performing first graders in rural schools: A comparison of two instructional approaches. *Journal of Instructional Psychology, 16*, 112-121.

U.S. Department of Health & Human Services. (2009). *Project Upgrade in Miami-Dade County*. Administration for Children & Families. Retrieved on March 19, 2009, from [http://www.acf.hhs.gov/programs/opre/cc/upgrade\\_miami\\_dade/reports](http://www.acf.hhs.gov/programs/opre/cc/upgrade_miami_dade/reports).

Valente, F., & Martins, M. A. (2004). Metalinguistic competencies and learning to read in two primary school groups using different teaching methods. *Analise Psicologica, 22*, 193-212.

Verhoeven, L., Schreuder, R., & Baayen, H. (2003). Units of analysis in reading Dutch bisyllabic pseudowords. *Scientific Studies in Reading, 7*, 255-271.