ADVANCED PHONICS AND MULTISYLLABIC WORD FLUENCY MEASURES: A RELIABILITY AND VALIDITY STUDY

Sara J. Doty

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Department of Psychology

Central Michigan University
Mount Pleasant, Michigan
April 2012

Accepted by the Faculty of the College of Graduate Studies,

Central Michigan University, in partial fulfillment of

the requirements for the doctoral degree

Dissertation Committee:

Michael Hixson, Ph.D.	Committee Chair
Sandra Morgan, Ph.D.	Faculty Member
Dawn Decker, Ph.D.	Faculty Member
April 5, 2012	Date of Defense
Roger Coles, Ed.D.	Dean College of Graduate Studies
July 12, 2012	Approved by the College of Graduate Studies

Committee:

Michael Hixson, Ph.D., Chair

Sandra Morgan, Ph.D.

Dawn Decker, Ph.D.

This dissertation is dedicated to my loving family.
I am grateful for all of their love, support, and encouragement throughout the years.

AKNOWLEDGEMENTS

I wish to thank my dissertation chair Dr. Michael Hixson. This dissertation would not have been possible without his guidance, support and dedication in assisting me with this study. I consider it an honor to have worked with him and am grateful for the many opportunities he has given me that truly enhanced my graduate education.

In addition, I would also like to thank Dr. Sandra Morgan and Dr. Dawn Decker for being a part of my dissertation committee. Their contributions, suggestions, and expertise enhanced the quality of the study to make it better than what I could have accomplished on my own.

I would also like to thank the staff and students at Heritage Elementary, who with their time and effort helped to make this dissertation possible. Finally, I wish to acknowledge the support of Central Michigan University in producing this work.

ABSTRACT

ADVANCED PHONICS AND MULTISYLLABIC WORD FLUENCY MEASURES:
A RELIABILITY AND VALIDITY STUDY

by Sara J. Doty

The National Reading Panel (2000) identified phonics as being one of the five areas that are important for becoming a good reader. However, currently there is no fluency measure to assess student's advanced phonics skills. A measure of advanced phonics skill may be beneficial as a progress monitoring tool for students who are below ORF benchmarks and are receiving reading interventions. Since phonics skills can be assessed using nonsense or real words and assessing the ability to read multisyllabic words is another way to measure students' reading skills, these different methods should be investigated to determine which is the most valid and reliable. The purpose of the present study was to develop and evaluate the reliability and validity of two different fluency probes assessing complex phonics skills, as well as a multisyllabic word probe. Participants from second and third grade were administered three different one minute phonics fluency probes containing real words, a combination of real words and nonsense words, and a multisyllabic word probe. The TOWRE, DIBLES NEXT NWF and ORF were administered to each student to assess the concurrent validity of the fluency measures being developed. According to the results, test-retest and interrater reliability for all three fluency probes were good. However, the reliability for the multisyllabic word probe was slightly lower than required for screening measures for third graders. The fluency probes correlated well with the TOWRE and DIBELS measures. Multiple regressions indicated the phonics probes increased the amount of variance explained in student's overall reading scores above what Oral Reading Fluency alone was able to explain. This was true for all students, as well as only the

students who were below their ORF benchmark. Overall, the probe containing multisyllabic words had the least advantages of the three probes. The fluency probes testing real words or a combination of real and nonsense words appeared to have the most advantages. However, the two fluency probes had their own unique advantages for assessing advanced phonics skills. These probes should be researched further and should be examined to determine if they are useful progress monitoring measures.

TABLE OF CONTENTS

TABLE OF CONTENTS continued

LIST OF TABLES

LIST OF TABLES

TABLE PAGE

CHAPTER I

INTRODUCTION

Learning to read is one of the most important academic skills that students develop in school (McCoach, O'Connell, Reis, & Levitt, 2006). In 2000, the National Reading Panel suggested five areas that are essential to becoming a good reader. The first two of the five areas are related to alphabetics. Phonemic awareness is the first area that is important for reading. Instruction in this area concentrates on teaching children to manipulate phonemes (i.e., the smallest units of speech) in spoken syllables and words. The second area of alphabetics is phonics. In phonics instruction, children are taught letter-sound correspondences, spelling patterns, and how to apply that knowledge in their reading. Fluency is the third area recognized by the National Reading Panel. This includes the ability of a child to read connected text orally with speed, accuracy, and proper expression. Fluency has often been overlooked during reading instruction, which is unfortunate because if text is read in a laborious and inefficient manner, it is difficult for the student to comprehend what was read. For this reason, fluency is also necessary for reading comprehension, which is the National Reading Panel's fourth important area. Reading comprehension is the ability of the reader to obtain meaning from the text and involves intentional and thoughtful interactions between the reader and the text. Comprehension instruction includes teaching students techniques to recall what they have read, answer questions, generate questions about what they have read, and summarize the text. The fifth area recognized by the National Reading Panel, vocabulary, is also related to reading comprehension. The National Reading Panel stated there are two types of vocabulary-oral and print. If a reader can decode a new word into speech, they can understand it if it is in their oral vocabulary. However, if the word is not in the reader's oral vocabulary, they must determine the meaning of the word in another way (National Reading Panel, 2000).

According to the National Reading Panel, children should receive instruction in all five of the areas discussed above. Not only is instruction in these areas important, but it is also important to assess each area. Assessment in each area can help to determine what skills the child still needs instruction on and which skills they have learned. Furthermore, while children are being instructed in these areas, assessments can be used to monitor students' progress of the skills being taught.

Humans possess a phonetic reservoir that contains all of the sounds they are physiologically capable of making. A particular language system consists of a selection of distinct sounds from the phonetic reservoir to make up the phonemic base. Sounds from this phonemic base are arranged in a variety of patterns to produce meaningful utterances. While the estimated average number of phonemes used in known languages around the world is 45, the phonetic basis of American English consists of approximately 46 sounds. A total of 24 of the phonemes are consonantal, 9 are vocalic, 3 are semivocalic, and 10 are the suprasegmental phonemes of pitch, stress, and juncture. In American English, pronounceable units of sound are normally formed around a vowel nucleus. When consonant phonemes are combined to these vowels, syllables are formed. Combining these syllables creates the words of American English (Hanna, Hanna, Hodges, & Rudorf, 1966).

Reading Development

The phonemes that make up the English language can be combined to make the words we read. The National Reading Panel (2001) recognized the importance of alphabetics in reading.

Ehri (2005) proposes a theory that suggests four phases in how people learn to read words by sight which focuses on the importance of alphabetics in how children begin to read words. The four phases are based on knowledge of the alphabetic system, grapheme–phoneme relations, and phonemic awareness. The four phases are: pre-alphabetic, partial alphabetic, full alphabetic, and consolidated alphabetic. In the pre-alphabetic phase, children know very little about the alphabetic system. Because of this they cannot use form letter-sound connections to read words. Any words they do read are from remembering selected visual features. In addition, print they read is from contextual cues, such as a popular logo rather than reading the letters. Because of this, children in the pre-alphabetic phase are effectively nonreaders.

The second phase is the partial alphabetic phase. In this phase, children begin to learn the names or sounds of alphabet letters and use them to remember how to read. At this point, children have made connections with only some of the letters and usually focus on the letters at the beginning and end of words. When they read, they often confuse words that have similar spellings. In addition, in this phase children lack full knowledge of the alphabetic system, especially vowels. The result of this partial knowledge is that they are unable to segment many words into all of their phonemes, they have much difficulty decoding unfamiliar words, and will often invent partial spelling of words by writing only the more salient sounds.

When a child is able to form all connections between letters in spelling and phonemes in pronunciations, they are in the third phase called the alphabet phase. In this phase, the child knows the major grapheme-phoneme correspondences and can segment the pronunciations into phonemes that match up to the graphemes they see. This means that when a child sees letters in a word they connect them to the corresponding sounds in their memory and begin to remember words by sight instead of having to decode every word. Furthermore, word reading becomes much more accurate and readers are able to decode unfamiliar words.

The last phase is the consolidated phase. This phase emerges as alphabet phase readers retain increasingly more sight words in their memory. In this phase, the child becomes more familiar with recurring letter patterns and consolidates these grapheme-phoneme connections into larger units. These units include syllables, morphemes, and whole words that have become unitized. This knowledge of letter units is important for helping students to read multisyllabic words. This is because they are required to remember and recall fewer grapheme-phoneme connections from their memory. Instead of having to recall every the sound for every letter, they can recall the syllables in the word (Ehri, 2005).

Decoding

When reading, there are four different ways that people read words (Ehri, 2005). The first way is decoding in which one can either sound out and blend graphemes into phonemes, or work with larger chunks of letters to blend syllabic units into recognizable words. A second way of reading words is called analogizing which involves using words we already. Using context and letter cues to guess unfamiliar words is the third way people read words. This is known as predicting. The fourth way to read words is by memory or sight in which one can look at the words and our brain recognizes them. However, this only applies to words that we have read before. Of these four ways to read words, the first three (decoding, analogizing, and predicting) can be used to help us read words that are unfamiliar that we have not read before, whereas memory or sight reading only applies to words that are familiar. The present study focuses on students' reading words using decoding. There has been much research focused on decoding skills in children.

Compton (2000) studied the relationships between the rate of growth in phonemic awareness and rapid naming speed (cognitive processes), measures of print knowledge (letter name/sound knowledge, advanced graphophoneme knowledge, and orthogonal awareness), and decoding skills in first graders. Since phonemic awareness skills and rapid naming speed have been identified as important predictors of word reading development, they may also be independent predictors of decoding growth in first grade. Measures of print knowledge have also been suggested to help students develop word recognition skills.

It has been argued that as students develop word recognition skills, their graphophoneme connections move from simple letter-sound correspondences to more complex connections that include the sounds associated with multiletter grapheme units such as consonant blends and vowel diagraphs. Hence, it is suspected that various levels of graphophoneme knowledge (i.e., letter name, letter sound, and advanced graphophoneme) knowledge should be significant predictors of students' decoding skill growth. Since early reading skills can be predicted by gains in orthographic awareness, it may also serve as a predictor of students' growth of decoding skills in first grade (Compton, 2000).

To investigate the relationships among these skills and decoding nonword and real words, the skills growth was measured for 75 first-grade students. Each student was tested seven times in first grade, once a month from October to April, on their letter name knowledge, letter sound knowledge, advanced graphophoneme knowledge, word reading ability, phonemic awareness, orthographic awareness, rapid naming speed, and nonword reading. According to Compton's (2000) results from hierarchical linear modeling (HLM), students' initial scores in phonemic awareness skill, letter sound knowledge, advanced graphophoneme knowledge, and rapid naming speed all emerged as independent predictors of the variance in final nonword reading scores. The HLM model for word reading was able to account for over 70% of the variance in students' end of the year word reading scores, and 63% of the variance in their end of the year nonword reading. Results suggest that each of the factors make a unique contribution to early decoding skill acquisition and support the theory that students who have a mix of cognitive processing skills and print knowledge will tend to have the greatest success in beginning reading, whereas those who have less of an optimal mix of these abilities will tend to have less success (Compton, 2000).

When predicting growth rates of real word and nonword reading, only initial letter sound knowledge emerged as a unique predictor of variance beyond that accounted for by the group of variables in both word and nonword reading growth variables. Furthermore, students' initial scores in cognitive-processing abilities and measures of print knowledge combined accounted for only 42% of the word reading growth variance and 31% of the nonword reading growth variance. This suggests that there are other variables involved in predicting growth rates in students' decoding skills. When regression analyses were conducted on the growth of advanced phoneme knowledge, orthographic awareness, and rapid naming speed, each explained significant amounts of unique variance in the individual differences of students' real word reading growth rates, whereas growth rates in advanced graphophoneme knowledge and orthographic awareness explained significant amounts of unique variance in the growth rate of nonword reading. In addition, when these growth rate variables were added into the word and nonword reading growth regression equations, they were able to explain a significant amount of variance, 29% for real word reading and 17% for nonword reading, beyond what was explained by students' initial scores in these variables alone. These results suggest that there is a connection between the rate of growth associated with the acquisition of print knowledge and growth of decoding skills during the early stages of word reading development.

Results from this study also suggested that the role of rapid naming in predicting word reading skills changes during decoding skill development. Rapid naming speed is an independent predictor of both word and nonword reading performance during the early stages of decoding skill development. This may be because there is very little difference between the processes and strategies students employed to decode unfamiliar words and nonwords that contained similar spelling patterns. As students began to develop an extensive repertoire of sight words and move past the most basic stage of decoding skill development the process and strategies employed to read words and nonwords may be different depending on the students' relative familiarity of the words. During this point of decoding development, rapid naming speed is only related to the reading of real words. Finally, when students' decoding development reaches more advanced levels, once the effects of phonemic awareness have been removed rapid naming speed no longer independently predicts word or nonword reading performance (Compton, 2000).

Before students develop a decoding strategy they often use phonetic cue reading or visual cue reading which entails guessing the word based on the initial consonant or a contextual feature. As they become more skilled readers, they begin to accomplish word recognition by accessing the spelling features of the word which are more specific instead of using a letter-by-letter decoding procedure. This type of word recognition is often referred to as sight word reading (Aaron, Joshi, Ayotollah, Ellsberry, Henderson, & Lindsey, 1999; Ehri, 2005).

There have been many disagreements over whether decoding and sight word reading are separate processes that develop independently of each other or are dual processes that are interdependent. It is reasoned that if these are dual processes and sight word reading skills depends on decoding skills, poor decoders will also be poor sight word readers. However, if the two skills are independent processes, some poor decoders will also be poor sight word readers whereas some poor decoders will not be poor sight word readers (Aaron et al., 1999).

In a study comparing the skills of students with poor reading skills, it was found that of the 26 students identified as having poor decoding skills 20 also had weak sight word reading skills. In addition of the 27 subjects who were identified as having weak sight reading skills, 22 had poor decoding skills. These ratios suggest there is a systematic relationship indicating that having poor decoding skills is associated with having poor sight word reading skills. According to the results of a 2×2 chi-square analysis the two reading abilities are not independent of each other. This suggests that having good decoding skills is necessary but not sufficient for sight word reading. Furthermore, in the absence of adequate decoding skills, sight word reading skills are not likely to flourish (Aaron et al., 1999).

The development of decoding skills also has an impact on students' comprehension skills. Decoding skills are viewed as being a prerequisite condition for reading comprehension, although not sufficient. It is believed that there is a direct causal link that decoding affects comprehension, but comprehension does not affect decoding (Shinn, Good, Knutson, Tilly, & Collins, 1992). A reading model by Samuels (1981) emphasized fluent, automatic decoding skills as a necessary mediator for comprehension to occur. According to the model, each student only has a limited amount of attention to devote to the reading process. If they have to allocate that attention to decoding, then limited or no attention can be devoted to comprehending the meaning of the text being read. However, if the student can decode words automatically without having to attend to the decoding process then they can attend to the meaning of the text. In order to examine whether reading is comprised of decoding and comprehension a confirmatory factor analysis of curriculum-based measurement (CBM) oral reading fluency has been conducted (Shinn, Good, Knutson, Tilly, & Collins, 1992). One hundred and fourteen third-grade students and 124 fifth-grade students participated in the study. Each student was administered two measures of decoding skills, four comprehension measures, and two CBM oral reading fluency measures. According to the results of this study, for third-graders a one-factor model of reading competence was found to most adequately fit the data. However, for fifth-graders a two factor model comprised of reading decoding and reading comprehension was the best. Reading decoding and comprehension were highly correlated in this study (r = .83), but could still be differentiated as separate constructs in fifth-graders (Shinn, Good, Knutson, Tilly, & Collins, 1992). For third graders there was no evidence of comprehension being a unique factor.

Phonics

Phonics instruction is one way students are taught to decode words. The National Reading Panel reviewed the research on the effectiveness of systematic phonics instruction and found that a variety of phonics programs have proven effective with children of different ages, abilities, and socioeconomic backgrounds. A meta-analysis conducted by the National Reading Panel concluded that systematic phonics instruction makes a bigger contribution to children's growth in reading than alternative programs. In 38 phonics intervention studies used in the meta-analysis, the mean effect size associated with direct phonics instruction was .44. Systematic phonics instruction typically involves explicitly teaching students a prespecified set of letter-sound relations in a sequential order and having the students read text that provides practice using these relations to decode words (National Reading Panel, 2000). When words become larger, it is more difficult to blend sounds together. To address this problem, some phonics programs have taught students to read larger word subunits, such as “st” and “br” as blends so that they do not have to remember as many separate parts when blending words.

Teaching students to use phonics skills while reading can help them to read words that are not familiar. It gives students a way to figure out an unfamiliar word rather than guessing or skipping over the word. However, in order for students to be able to use this strategy effectively, they need to be explicitly taught phonics rules. Even though phonics cannot be used to read irregular words, phonics instruction increased student’s ability to read irregular words (National Reading Panel, 2000). Not only does explicitly teaching phonics help with reading, but it is also effective for improving spelling skills. Phonics instruction is also effective in improving comprehension in Kindergarten and 1st grades. However, whether phonics instruction should continue beyond 2nd grade, and the goals of more advanced forms of phonics instruction still need to be researched (National Reading Panel, 2000).

Not only is it important to teach students phonics, but these skills also need to be assessed. Assessment of phonics can be conducted in several ways, such as testing letter-sound correspondence. In this type of assessment, students are presented with each letter in isolation and are asked to say the sound of the letter. This type of assessment allows the examiner to determine what letter-sound correspondences the student has learned and which they still need to be taught.

Phonics skills can also be assessed by having students read a variety of real words that contain an array of different phonetic elements. Students are asked to read the words and the examiner marks whether the student read the part of the word correctly containing the phonics skill being assessed. Even if the student does not correctly read the word, but accurately reads the phonics skill being assessed in the word, the item is counted as correct. This method of phonics assessment helps to determine which phonics skills the student needs to be taught. Phonics skills can also be assessed through the use of pseudowords (nonsense) words by having students read made up words. Pseudowords are pronounceable letter combinations which lack semantic meaning and cannot be read by using contextual cues or previous sight word knowledge. Because these are not real words, students have to use their alphabetic knowledge and phonics skills to decode the words (Thomson, Crewther, & Crewther, 2006). There is strong empirical support for the use of measures of pseudoword reading to assess the alphabetic principle. Of the 38 studies included in the NRP meta-analysis on phonics interventions, 18 included a measure of pseudo-word reading to determine intervention impact. More independent effect size calculations for pseudo-word reading (n = 40) than real word reading (n = 30) went into the analysis of impact (Good, Baker, & Peyton, 2011).

When assessing phonics skills, examiners typically look at two criteria. First, how accurate the student is at reading words. Specifically, they may look at whether the student reads the entire word correctly or if they correctly read the part of the word containing the phonics skills being assessed. Second, examiners may calculate the rate that the student reads the words. Not only should a student be able to use their phonics skills to read words, but for it to be effective, they should be able to do it at a quick pace. Therefore, they may assess both the accuracy and rate of which students are able to apply their phonics skills.

DIBELS Nonsense Word Fluency

A common way that students' decoding skills are assessed is through the use of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Nonsense Word Fluency (NWF) subtest. DIBELS is a standardized set of fluency measures of early literacy development. These measures are intended to identify students who are at-risk for developing reading problems and to monitor the development of their reading skills. DIBELS NWF is designed for students from the mid to end of kindergarten to the beginning of second grade (Good & Kaminski, 2002). NWF is a direct measure of the alphabetic principle and basic phonics. It assesses knowledge of basic letter-sound correspondences and the ability to blend letter sounds into consonant-vowel-consonant (CVC) and vowel-consonant (VC) words. There are two separate scores for NWF. The first is the number of letter-sounds they were able to correctly produce in a minute (CLS). The second score is new to the DIBELS NEXT edition. The second score students receive is for whole words read (WWR) which is the number of make-believe words read correctly as a whole word without being sounded out first (Good & Kaminski, 2011). Given the fact that it is fluency based, students who are phonologically recoding the words receive higher scores than students who are reading the letter sounds in isolation (Good & Kaminski, 2002).

Rouse and Fantuzzo (2006) examined the validity of DIBELS probes for kindergarteners, which include Letter Naming Fluency (LNF), NWF, and Phoneme Segmentation Fluency (PSF). Specifically the researchers looked at the relationship between the DIBELS probes and standardized scores from nationally normed tests of district literacy constructs and the relationship between the probes and measures of other cognitive and behavioral competencies. A random sample of 330 kindergarteners participated in the study. In addition to the DIBELS subtests, each student was given several achievement tests which included the Test of Early

Reading Ability-Third Edition (TERA-3; Reid, Hresko, & Hammill, 2001), Developmental Reading Assessment (DRA; Beaver, 1997), TerraNova-Second Edition (CBT/McGraw-Hill, 1997), and the Test of Early Mathematics Ability-Second Edition (TEMA-2; Ginsburg & Baroody, 1990). In addition, teachers were asked to complete the Learning Behavior Scales (McDermott, Green, Francis, & Stott, 1999) to measure the students' typical learning-related behaviors, and the Penn Interactive Peer Play Scale (Fantuzzo & Hampton, 2000) to assess the students' peer-play interactions at the end of kindergarten. Convergent and discriminant validity was assessed for the three kindergarten literacy indicators of DIBELS: LNF, NWF, and PSF. The strongest significant positive relationships of each were with indicators of overall reading ability (i.e., DRA and the TERA-2). These results corroborate previous findings on the validity of DIBELS. Evidence of discriminant validity was found with different reading constructs. For example, NWF showed the strongest relationship to the construct of using Conventions of Print, which is a direct assessment of students' knowledge of the process and basic principles of learning to read. Results of predictive validity showed a significant relationship with reading, vocabulary, and language constructs measured at the end of first grade. LNF evidenced the largest correlation followed closely by NWF. Convergent and divergent validity analyses also showed a relationship between DIBELS LNF, NWF, and PSF with other cognitive constructs, such as early mathematical abilities (measured by the TEMA-2), and social-behavioral constructs. These convergent and divergent validity results underscore the important developmental overlap among cognitive and behavioral competencies for young children.

A longitudinal study evaluating the validity of DIBELS LNF and NWF as measures of early reading skills was conducted (Speece, Mills, Ritchey, & Hillman, 2003). Thirty-nine students who were tested in the spring of kindergarten and first grade participated in the study.

A battery of language, prereading, and reading measures were administered to the students that included the Peabody Picture Vocabulary Test-Revised (Dunn & Dunn, 1981), Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999), Letter-Sound Identification measure from the Texas Primary Reading Inventory (Foorman, Fletcher, Francis, Carlson, Chen, Mouzaki, et al., 1998), DIBELS LNF and NWF, Woodcock-Johnson –Revised (WJ-R; Woodcock & Johnson, 1989) Letter Word Identification and Word Attack subtests, and CBM ORF. The concurrent validity coefficient for the WJ-R Letter Word Identification was .91 for NWF. Predictive validity for NWF was also strong. The study found predictive validity coefficients of .59 for both WJ-R subtests and .71 for first-grade ORF. In terms of construct validity, NWF in kindergarten was the only significant predictor of kindergarten WJ-R Letter Word Identification. In addition, a correlation of .69 was found between kindergarten NWF and first-grade ORF. Furthermore, NWF was highly sensitive and identified 87.5% of the poor readers in first grade. Overall, the NWF fluency provides similar information to that provided by the WJ-R Letter Word Identification subtest as seen by the concurrent validity coefficients. This study indicates that NWF and to a lesser extent, LNF, are valid measures of early reading and poor reader status.

Research has also been conducted exploring the relationship of NWF to later measures of reading. Harn, Stoolmiller, and Chard (2008) investigated the relationship between NWF and ORF across the school year in 938 first grade students. They found moderate to strong correlations between the two across the school year. In addition, they found a positive linear relationship across the year as scores on NWF and ORF increased. The authors noted however, that when students' scores on NWF were initially high at the beginning of the year, the linear effect was dramatically reduced. This effect may be due to the fact that students that are performing high at the beginning of first grade have most likely mastered initial phonetic skills necessary for word reading and are using their more developed skill in phonological processing and unitization to read. Because of this, their skills in NWF should not be expected to improve as much as ORF scores.

Harn, Stoolmiller, and Chard (2008) also evaluated four different strategies students used on NWF and its relationship with ORF. One, the sound-by-sound strategy where students only provided the letter sounds of the nonsense words and did not attempt to recode or blend the sounds. Second, some students used a recoding strategy where they initially said the sounds of each letter and then recoded the word (e.g., /t/ /e/ /m/ /tem/). Third, students who used the partial blending strategy produced a larger unit of sound within the word beyond letter sounds (e.g., /t/ /em/, /wa/ /f/). Fourth, were students who used the unit strategy and simply read the nonsense word as a unit (e.g., /tem/, /waf/). When comparing students using the different strategies results indicated that students who used the unit strategy not only scored significantly higher on NWF but they also read significantly more fluently on both the winter and spring ORF assessments than students using other strategies. Furthermore, students who used a partial blending strategy did not read more fluently in winter but were reading more fluently in the spring compared to students who used the sound-by-sound and recode strategies. Interesting to note is the students who went sound-by-sound and then recoded the words correctly not only scored lower on ORF at the end of the year, but they also had the lowest mean of the four groups. The authors stated that even though these students are unitizing, they seem to lack automaticity even though they can recode the words. These students may not be attending to the larger units that they are producing, even within simple words (Harn, Stoolmiller, & Chard, 2008).

A study looking at the ability of NWF to predict later reading outcomes and its sensitivity and specificity for English Language Learners (ELLs) has also been conducted (Vanderwood, Linklater, & Healy, 2008). The study involved 134 ELL students who were tested with DIBELS NWF three times during first grade and curriculum-based measurement in reading and reading comprehension at the end of third grade. All of the students were classified as being ELLs in first grade according to the CaliforniaS English Language Development Test (CELDT; California Department of Education, 2002) and 73% of these students were still considered to have limited-English proficiency at the end of third grade. Data incorporated in the analysis from first grade included NWF from the beginning, middle, and end of first grade, scores from the reading and reading comprehension sections of the Stanford Achievement Test, Ninth Edition (SAT9; Assessment Committee, 2002) given at the end of first grade, and AIMSweb curriculum-based measurement reading (CBM-R) probes from the end of first grade. In third grade end of the year AIMSweb CBM-R and maze probes, as well as the California Achievement Test, Sixth Edition (CAT6; California Department of Education, 2004) normal curve equivalence reading composite scores were used in the analysis. The study found that the ELL students' growth in NWF over the course of first grade was approximately 1.43 sounds per week. In addition, the NWF in first grade had significantly large correlations to CBM-R and Maze, while the growth of NWF over time was moderately correlated with these measures. NWF throughout first grade as well as students' growth on NWF were all moderately correlated with the CAT6.

Regression analyses were run to determine whether knowledge of phonics contributes above and beyond broad language performance in predicting reading outcomes. In a four step hierarchical regression ELL level, end of first grade NWF, growth in NWF, and SAT9 were entered as variables. When predicting scores on CBM-R, after ELL was entered, NWF still accounted for a significant amount of variance. However, adding the growth of NWF over time did not add a significant amount of variance explained (Vanderwood, Linklater, & Healy, 2008).

Even though the results indicated that NWF at the end of first grade had moderate to strong correlations with third grade literacy outcomes, the authors argued that knowledge about the relationship between a screening measure and outcome measures is not sufficient to warrant its use. For additional evidence, the authors also looked at the predictive accuracy of NWF with third grade CBM-R, maze, and CAT6 reading outcomes. Students below the DIBELS NWF benchmark were considered to be "at-risk" and on the outcome measures students below the 25th percentile were considered being "below expectations" in reading. The specificity of NWF and the outcome measures were all above 70%. However, the sensitivity and positive predictive values were at 50% or below. Even though the overall identification rate of those at or above expectation was high (above 70% on two outcome measures, and above 65% on the third), in order to identify future poor readers that are ELL students, improvement in NWF's accuracy is needed. The authors noted that NWF had a higher than desired false-negative rate in which over 80% of these were students that were classified at the lowest ELL level. Because of the high false-negative rate for the lowest ELL students additional information besides NWF should be used to determine whether these students are not at-risk. Overall more empirical support for using NWF with ELL students is needed (Vanderwood, Linklater, & Healy, 2008).

Another study with ELL students compared levels and patterns of growth in ORF and NWF at the end of first grade (Graves, Plasencia-Peinado, Deno, & Johnson, 2005). The researchers were interested in whether measuring growth in students' nonsense word reading would be a useful complement to the growth in reading text (ORF). Seventy-seven first graders from nine different classrooms classified as being low, middle or high readers based on their ORF scores participated in the study. Each student was given NWF and ORF once a week for six consecutive weeks. Results of the study found a strong correlation of .86 between the end of the year first grade ORF and NWF scores. However, the correlation between students' gain on ORF and gain on NWF across the 6 weeks was only .302. When comparing the low, middle, and high readers, students in the low achieving group gained significantly fewer nonsense words per minute than the students in the high achieving group. When looking at the standard deviation in NWF scores, the deviation was larger for students in the middle and high reader groups. This indicates that there is a larger degree of variability in these students' performances than the students in the low reader group. When looking at the slopes of the low readers, the ORF and NWF appeared to be similar. The authors noted that in previous studies, growth in ORF was the steepest for beginning readers and tended to slow as they became more proficient. However, the authors found the opposite for NWF scores. The study found that on NWF, the highest achieving group outperformed the lower groups. The steep slope for high performing readers in NWF suggests that at some point the task is quite simple for strong decoders.

Several studies have evaluated performance on NWF in the fall of first grade and its importance of prediction spring ORF outcomes. A recent study investigated the relationship among students' NWF initial skill level, NWF growth across the school year, and end of year ORF and reading comprehension skills (Fien, Park, Baker, Smith, Stoolmiller, Kame'einui, 2010). A sample of 3,506 first grade students was assessed during the school year with DIBELS NWF, ORF, and the Stanford Achievement Test 10th Edition Reading Comprehension subtest. After administering NWF, students were categorized into five strata based on their score. Using a hierarchical linear modeling for analysis, one major finding of the study was that for the students in the first four strata (95%) positive NWF trajectories in the beginning of the year were strongly predictive of end-of-year ORF and reading comprehension outcomes. However, findings indicated that once students reach a threshold scoring above 70 correct sounds per minute, subsequent gains on NWF are not meaningful for improved reading outcomes. At this point, the student's automaticity would cap any additional comprehension benefits. The authors noted that this may because NWF measures only vowel consonant and consonant-vowel-consonant words. It is unknown whether further growth on more advanced phonics elements such as multisyllabic words or more complex phoneme-grapheme relations would be important for students who have reached the threshold in predicting and accounting for their further improvements in overall reading proficiency. Looking at growth on the more advanced phonic elements may be important for students who have already met the threshold. However, this has not yet been investigated.

Word Reading Assessment

Even though DIBELS NWF is a common method of assessing students' early phonics skills, as noted earlier, phonics skills can also be tested by using real words. One study looked at the relationship between pseudoword, real word reading, and their relationships with the phonological processing subskills of phonological awareness and rapid automatic naming (RAN) in the first 4 years of schooling (Thomson, Crewther, & Crewther, 2006). A total of 71 Australian students in the first four grade levels (similar to kindergarten through third grade) participated in the study. These students were given the Woodcock Language Proficiency Battery-Revised (Woodcock, 1991) Word Identification subtest as a measure of real word reading and the Word Attack subtest as the measure of pseudoword reading. In addition they were given the Lindamood Auditory Analysis Test (Lindamood & Lindamood, 1979) as a measure of phonological awareness and digit naming from the Phonological Assessment Battery (Fredrickson, Frith, & Reason, 1997) was used for the RAN test. Results showed a strong correlation of .94 between pseudoword and real word reading. Across the grade levels correlations were .91 for the first 2 years of school (kindergarten and first grade), .93 for year 3 of schooling (second grade), and dropped to .70 in year 4 of schooling (third grade). Correlations also showed that across the 4 years of schooling phonological awareness was more strongly correlated with real word than with pseudoword reading, and there were no differences in the correlations with RAN. However, when year 4 of schooling was not included (third grade) correlations of pseudoword reading and real word reading with phonological awareness, pseudoword reading was not significantly different from that of real word reading. This suggests that during the first 3 years of schooling, pseudoword reading and real word reading are measuring the same construct. Results from the study also demonstrated that pseudoword reading tasks provide no more useful diagnostic information regarding phonological awareness or phonological processing ability in young readers than do real word reading tests.

A study that used real words investigated the concurrent validity of five simple measurements of reading (Deno, Mirkin, & Chiang, 1982). Participants in the study included thirty three general and special education students ranging from grades one to five. The reading behaviors that were measured in the study included: 1) Words in Isolation where students were timed on a list of 60 words, 2) Words in Context in which students were given a passage with every fifth word underlined and they were only to read the underlined words, 3) Oral Reading Fluency, 4) Cloze Comprehension where students were given a passage with every fifth word missing and had to guess what the missing word was, and 5) Word Meaning: in which every fifth word of a passage was underlined and students were asked to give the meaning of the word. Three 1 minute timings of each of these measures were given to each student. The Stanford Diagnostic Reading Test (SDRT; Karlsen, Madden, & Gardner, 1975) Reading Comprehension subtest, and the Woodcock Reading Mastery Test (WRMT; Woodcock, 1973) Word Identification and Word Comprehension subtests were given to each student to assess the concurrent validity of the reading behavior measures that were developed. Results indicated that the Words in Isolation, Words in Context, and Oral Reading Fluency measures correlations with the standardized reading measures ranged from .73 to .91. These correlations were similar regardless of whether students were in general or special education. This suggests that any of the three word reading measures (Words in Isolation, Words in Context, and Oral Reading Fluency), would be a valid index of a student's reading proficiency. In addition, it appears that any one of these measures, or a combination of the measures, can be used to estimate a student's proficiency in decoding.

To further research the three measures, a second study was conducted to determine whether the grade level of the materials or the duration of the test would alter the correlations. This study included 45 students in first through sixth grade. During this study, students were given two 30 second and two 1 minute timings on the word recognition measures that were at a third grade reading level. Results of the study indicated that the correlations between using third-grade and sixth-grade word reading measures remained high and were consistently in the .80's and .90's. In addition the 30 second tests correlated highly with the 1 minute tests (Deno, Mirkin, & Chiang, 1982).

An important validity issue is whether the reading measure accurately identifies students at-risk for reading difficulties. However, the accuracy of the measures used to determine which students are at-risk is relatively low when looking at the high rates of false positives and negatives. A study examined whether adding a word identification fluency (WIF) measure to a multivariate screening battery that included phonemic awareness, a rapid naming skill, and oral vocabulary or 5 weeks of progress monitoring data would add any predictive utility for determining which students are at-risk (Compton, Fuchs, Fuchs, & Bryant, 2006). Two hundred and six first grade students who had low reading skills participated in the study. Each of the students was given the multivariate screening battery and progress monitored once a week with WIF for 5 weeks. Progress monitoring data was fit to a line using an ordinary least square regression and a slope for number of words gained per week (WIF Slope). The estimated number of words read at week 5 (WIF Level) was calculated. The WIF progress monitoring probes contained 50 high-frequency words randomly sampled from the Dolch preprimer, primer, and first-grade level lists. At the end of second grade, the Test of Sight Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1997) and the Woodcock Reading Mastery Test-R/NU (Woodcock, 1998) word attack, word identification, and passage comprehension subtests were used to assess the students' reading skills.

A logistic regression was used to determine if the WIF measure added any predictive validity to correctly identifying at-risk students. The first model used the multivariate screening battery as predictors and found that the sensitivity rate was about 75% and the specificity rate was around 77.9%. The initial WIF score was added to the second model. This increased the sensitivity to 85% and specificity to 80.6%, but the improvement was not statistically significant. In the third model the WIF Slope and WIF Level (i.e., the estimated week 5 level based on linear regression) variables were added to the model. By adding these variables, there was a significant improvement in the sensitivity of classifying at-risk, in which 90% of the students were correctly identified. However, the specificity was still low (83.4%) meaning a large number of students were still being falsely identified as being at-risk. In order to increase the specificity of the test battery the researchers used a classification tree approach which allowed for the same set of predictors (scores on the multivariate battery) to interact in such a way that different combinations of risk factors could be used for classification across individuals. When this method was added as a fourth model, the specificity increased to 100% while the sensitivity rose to 93.5%. This suggests that even though adding additional word reading measures such as WIF can increase the ability to identify young students who are at-risk, different methods of classifying students are also needed to prevent over identifying students who are not at-risk (Compton, Fuchs, Fuchs, & Bryant, 2006).

There has also been a study that compared the validity of Word Identification Fluency (WIF) to NWF (Fuchs, Fuchs, & Comptom, 2004). Students had one minute to read a random selection of high frequency words on the WIF assessment. One hundred and fifty one first grade students participated in the study and were administered the Word Attack and Word Identification subtests of the Woodcock Reading Mastery Test-Revised (WRMT-R; Woodcock, 1987) and the Comprehensive Reading Assessment Battery (CRAB; Fuchs, Fuchs, & Hamlett, 1989) in the fall and the spring. In the fall and spring each student was given two measures of the WIF and NWF and the average was taken. An ordinary least-squares regression was calculated between the scores to get a slope of weekly improvement. Correlations for the WIF measure were significantly higher than for the NWF measure on the fall Word Identification subtest. In addition, WIF correlations for three of the four spring standardized reading measures were statistically higher than for NWF. The comparisons favored WIF except when the standardized measure was highly aligned with NWF (i.e., Woodcock Word Attack subtest), in which case the two measures were comparable. The researchers also compared the predictive validity of WIF and NWF to the predictive validity of the two Woodcock measures. When compared to the Word Attack subtest, NWF performed comparably when predicting scores on Woodcock Word Identification and CRAB comprehension, but outperformed both Woodcock measures when predicting CRAB fluency. WIF outperformed both Woodcock measures in predicting Word Identification fluency, CRAB comprehension, and CRAB fluency scores. Finally, a dominance analysis was conducted to determine if either the predictive validity of the WIF or NWF exceeds the other to compare the differing amount of unique variance attributed to the two predictors as they relate to the spring scores on the standardized reading measures. In these comparisons, WIF dominated NWF in 10 of the 16 comparisons. The slopes from WIF and NWF were added into the analysis to examine whether it would provide additional predictive value over the WIF and NWF performance. Adding the WIF slope added predictive value over the WIF and NWF in four of the eight comparisons. The NWF slope provided additional predictive value over WIF and NWF in two of the eight comparisons. When comparing the WIF and NWF slopes in predicting standardized reading scores, the WIF slope dominated in three of the four comparisons. Overall, results for the study favored WIF over NWF. One problem the authors noted with NWF is that competent phonological decoding is, especially as the year progresses, better represented by the capacity to decode a variety of phonetic patterns. However, DIBELS NWF only assesses consonant-vowel-consonant and vowel-consonant pseudowords. This is a concern because the students who perform well on nonsense word fluency's consonant-vowel-consonant pseudowords may or may not be skilled at reading consonant-vowel- consonant –e words, r-controlled words, dual vowel words, multisyllabic words, etc. They noted that the correlation between NWF and important criterion measures may be reduced because of the restriction of NWF to a single easy phonetic pattern. Because of this, there is a need to assess students' ability to decode words that contain complex phonetic patterns.

In an unpublished thesis, Doty (2008) compared the reliability and validity of three measures of advanced phonics skills. These were brief, one-minute probes containing real words (Real Word Fluency), nonsense words (Advanced Nonsense Word Fluency), a combination of the two (Combined Fluency), and high frequency words (High Frequency Fluency). Thirty-nine third grade students participated in the study. Using a two-week test-retest interval, the reliability for all the fluency probes were above .82, with the Real Word Fluency probe having the highest reliability. Interrater agreement for the fluency probes was also high with the average agreement being 95% or higher. The Wechsler Individual Achievement Test-Second Edition Pseudoword Decoding and Word Reading subtests (WIAT-II; The Psychological Corporation, 2002), the Test of Silent Contextual Reading Fluency (TOSCRF; Hammill, Wiederhold, & Allen, 2006), and Dynamic Indicators of Basic Early Literacy Skills Nonsense Word Fluency and Oral Reading Fluency (Good & Kaminski, 2002) were administered to each student to assess the concurrent validity of the fluency measures being developed. The phonics fluency probes correlated well with the WIAT-II and DIBELS measures, but often did not correlate significantly with the TOSCRF. Only the Advanced Nonsense Word Fluency and Combined Fluency probes were significantly correlated with the TOSCRF, which was used as an overall reading measure. Of the fluency probes, the High Frequency Fluency probe had the lowest correlations with standardized reading measures, many of which were not significant.

A series of one-tailed t-tests were conducted to test whether there were significant differences between the WIAT-II Pseudoword Decoding and Word Reading correlations for each fluency measure. Only the Combined Fluency probe showed a significant difference between the two subtests indicating that it is more of a phonics measure than a word reading measure. No significant differences were found with the other fluency probes. A series of hierarchical multiple regressions indicated few of the phonics probes increased the amount of variance explained above Oral Reading Fluency in students' overall reading scores. Overall, the fluency probes testing real words (Real Word Fluency) or a combination of real and nonsense words (Combined Fluency) appeared to have the most advantages. However, the two fluency probes had their own unique advantages for assessing advanced phonics skills and should be researched further (Doty, 2008).

Purpose

The Real Word Fluency and Combined Fluency measures were designed as an extension of the DIBELS NWF that can assess phonics skills of students ranging from the beginning of second to the beginning of fourth grade. The skills assessed by the fluency probes include consonant blends, the final e rule, r-controlled vowels, prefixes, and suffixes (See Appendix A). These skills were chosen because they are representative of the many skills students need to be successful readers. The BRIGANCE Diagnostic Comprehensive Inventory of Basic Skills-Revised (Brigance & Glascoe, 1999) and *Direct Instruction Reading* (Carnine, Silbert, & Kame'enui, 2004) were used as guides to determine the important phonics skills students need to learn as well as approximate grade levels for each skill. In addition, since students are beginning to learn how to read multisyllabic words in these grades a Multisyllabic Word Fluency probe has also been designed as a measure of students' reading skills.

Ultimately, the goal is for one of these measures to be a useful progress monitoring measure for students who are having reading difficulties. In the early elementary grades, students are learning advanced phonics skills and how to read multisyllabic words. Students who are not making benchmarks in oral reading fluency are sometimes given additional reading interventions. If these interventions are teaching the students the advanced phonics skills or how to read multisyllabic words, a progress monitoring measure assessing these skills would be beneficial. Currently, phonics progress monitoring measures such as DIBELS NWF only focus on basic letter-sound correspondences and would not be helpful for students learning more advanced phonics skills. Progress monitoring may be conducted using ORF, but for students with reading difficulties it may not be as sensitive to students' growth as a measure that explicitly targets advanced phonics skills. If only ORF is used, for some students it may appear as though the student is nonresponsive to the intervention and is not making progress. However, it may be that the student is making progress, but it is not enough to show growth in their ORF scores. Having a measure of advanced phonics skills or multisyllabic words may be more sensitive for monitoring the students' growth.

To reach this goal, the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes being designed would need to be evaluated for their use as progress monitoring measures. Fluency benchmarks indicating when the student has mastered the skill (phonics or multisyllabic word reading) would need to be established. However, before this is done the reliability and validity of the measures needs to be evaluated. This study will therefore be looking at the following questions to assess the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes to identify which one has the best technical adequacy. Doing so will help to determine which probe should be evaluated in the future for use as a progress monitoring measure.

Research Questions

1) Is the test-retest reliability for the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency adequate ($r \geq .80$; Salvia, Ysseldyke & Bolt, 2012) for 2nd and 3rd graders?
2) Is the interrater agreement for the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency adequate (≥ 80%; Salvia, Ysseldyke & Bolt, 2012) for 2nd and 3rd graders?
3) Which probe has the best reliability: Real Word Fluency, Combined Fluency, or Multisyllabic Word Fluency?
4) Do the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes correlate with other measures of phonics skills? Which measure has the highest concurrent validity with phonics measures?
5) Do the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes correlate with other reading skills? Which measure has the highest concurrent validity with other reading skills?
6) Do the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes correlate with overall reading scores? Which measure has the highest concurrent validity with overall reading scores?
7) Are the Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probe more of a measure of phonics or other reading skills?
8) For all students, individually how much variance does the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes explain in students' DIBELS ORF scores?
9) For students not above the DIBELS ORF benchmark, individually how much variance does the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes explain in students' DIBELS ORF scores?

10) For all students, does the Real Word Fluency, Combined Fluency, or Multisyllabic Word Fluency probe provide additional information about students' overall reading scores above DIBELS ORF?

11) For students not at DIBELS ORF benchmark, does the Real Word Fluency, Combined Fluency, or Multisyllabic Word Fluency probe provide additional information about students' overall reading scores above DIBELS ORF?

CHAPTER II

METHODOLOGY

Participants

Three hundred and forty-five second and third grade students from a Utah elementary school were invited to participate in the study. Overall, one hundred and sixty-two students, 46.7% of the students invited, participated in the study. Of the students invited to participate, 75% of second graders participated and 57.6% of third graders participated. Of the 162 students in the study, ninety were second graders and seventy-two were third graders. Ninety-nine (61%) were males and sixty-three (39%) were females. Of the students who participated, 95.1% were Caucasian, 3.7% were Hispanic, .6% were Pacific Islanders, and .6% were of Multiple Races. Twenty-five (15.4%) of the students participating in the study received special education services. Of these students, sixteen were classified under speech language impairment, five under specific learning disability, and four under other health impairment. At the elementary school, 17.5% of the students received free or reduced lunch.

Examiners

Six teacher assistants and one school psychologist served as examiners in the study. Prior to the study, all of the examiners received materials on how to score and administer each of the measures. During the training, each measure and scoring procedure was described to the examiners. Additionally, examiners were given practice administering and scoring each type of probe. Before the examiners were allowed to administer the probes to students, they were required to demonstrate the ability to score the all probes with at least 90% accuracy. All seven of the examiners met this criterion before the study began.

Measures

Real Word Fluency Probe

Real Word Fluency probes were individually administered to each student in order to assess their ability to use phonics to decode words (See Appendix B). Each probe contained 100 real words and the students were asked to read as many words as they could in 1 minute. If student hesitated for more than 3 seconds on a word the examiner told the students what it was so they could continue. If the student was reading the words on the probe as whole words, the examiner told the students the whole word they hesitated on. If student was reading the words on the probe by word parts, the examiner told the students the word part they hesitated on. The examiner had a copy of the probe and recorded any incorrect responses students made. Incorrect responses included saying the wrong sound, hesitating for more than 3 seconds on a word or part of a word, saying sounds out of order, and omitting sounds. Inserting sounds, improper articulation, repeating sounds, and self corrections within 3 seconds were scored as correct.

Combined Fluency Probe

The Combined Fluency probe was also individually administered to each student in order to assess their ability to use phonics to decode words. Students were given a probe and asked to read as many words as they could in 1 minute. The Combined Fluency probe contained a selection of 50 real and 50 nonsense words with a total of 100 words on each probe (See Appendix C). Words were selected so that each probe measures a variety of phonics skills. The rules used to score the Real Word Fluency probes were used to score the Combined Fluency probes.

Multisyllabic Word Fluency Probe

The Multisyllabic Word Fluency Probes was individually administered to each student in order to assess their ability to use phonics to decode multisyllabic words. Students were given a probe and asked to read as many words as they could in 1 minute. The Multisyllabic Word Fluency probe contained a selection of 50 real and 50 nonsense multisyllabic words with a total of 100 words on each probe (See Appendix D). Words were selected so that each probe measures a variety of phonics skills. The rules used to score the Real Word Fluency probes were used to score the Multisyllabic Word Fluency probes.

Test of Word Reading Efficiency (TOWRE)

The Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999) is a measure of a student's ability to pronounce words accurately and fluently. The TOWRE is a standardized test for ages 6 through 24 years 11 months and was individually administered to each student. It consists of the Phonemic Decoding Efficiency subtest which consists of 63 nonwords and the Sight Word Efficiency subtest that includes 104 real words. For each subtest, the student had 45 seconds to read as many words as they can.

Standardization for the TOWRE was based on a normative sample of 1,507 individuals from 30 states, ages 6 through 24. The sample was representative of the 1997 U.S. Census data in terms of geographic area, gender, race, ethnicity, urban/rural residence, socioeconomic status, and parent's educational background.

In terms of reliability, for ages 7 through 10 (2^{nd} through 4^{th} grade) alternative-form reliability coefficients ranged from .93 to .97 on the Sight Word Efficiency subtest, .93 to .95 on the Phonemic Decoding Efficiency subtest, and .96 to .98 for Total Word Reading Efficiency.

These coefficients suggest that both forms are testing similar content. To assess test-retest reliability, 72 individuals in Florida took the test twice within 2 weeks. Data was analyzed for three age groups. For ages 6 to 9, test-retest reliability coefficients were .96 (Form A) to .97 (Form B) for the Sight Word Efficiency subtest, .90 for the Phonemic Decoding Efficiency subtest, and .95 for Total Word Reading Efficiency.

When looking at the concurrent validity, the TOWRE has been compared to the Phonemic Decoding Efficiency (PDE) and the Word Attack subtests of the *Woodcock Reading Mastery Test- Revised* (WRMT-R; Woodcock, 1987) in two separate studies. Results from the studies indicated that for students between second and fourth grade, correlations between the TOWRE Phonemic Decoding Efficiency subtest and the WRMT-R Word Attack subtest ranged from .87 to .91. For the Sight Word Efficiency subtest, correlation with the WRMT-R Word Identification subtest ranged from .89 to .94. These correlations indicate that the subtests are highly related to other tests measuring similar skills.

In support of construct validity, scores were shown to increase with age. In addition, mean scores were found for different gender and ethnic groups were all within the average range. Scores for different disability subgroups are as one would predict. Students with speech-language handicaps scored much higher than students with learning disability and were in the average range. Students with learning disabilities had the lowest TOWRE scores, which is what would be predicted.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS™) NEXT Edition

DIBELS NEXT™ Oral Reading Fluency (DORF) assesses students' ability to read passages accurately and quickly. This is measured by calculating the number of words read correctly during a one-minute sample of passage reading. It is scored by having the examiner calculate the number of words read correctly. Students were administered the spring benchmarks for their grade level during the study. Three probes were administered and the examiner calculated the number of words read correctly for each passage. The median score for the three passages were recorded for each student. These scores were correlated with the students' fluency probe scores.

The technical adequacy for DORF is based on studies from general Curriculum-Based Measurement (CBM) Reading procedures. For elementary students, the test-retest reliability of DIBELS NEXT ORF words read correctly ranges from .91 to .97. Alternate-form reliability for individual DORF passages is particularly strong. For words read correctly, an initial study on the technical adequacy of the new measure found alternate-form reliability using different passages to range from .95 to .98 (Good & Kaminski, 2011).

The second grade fall benchmark of DIBELS NEXT™ Nonsense Word Fluency (NWF) was individually administered to the students as part of the study. NWF is intended to be used for students from the middle of kindergarten through the beginning of second grade. Students are given one minute to sound out or read as many of the words they can. There are two separate scores for NWF. The first is the number of letter-sounds they were able to correctly produce in a minute (CLS). The second score is for whole words read (WWR) which is the number of make-believe words read correctly as a whole word without first being sounded out (Good & Kaminski, 2011).

The fall benchmark for second grade was chosen because it is the last benchmark students are assessed with NWF. The NWF scores were correlated with the standardized reading measures and these correlations were compared with the correlations from the fluency probes to determine which better predicts the students' reading ability**.** In terms of technical adequacy, a two-week alternate-form reliability for 55 kindergarten and first graders during the middle of the year benchmark was conducted. Correlations for correct letter sounds were .71 for kindergarten and .85 for first grade. For whole words read kindergarteners had a reliability of .92 and first graders had a correlation of .90. An additional sample of 27 first graders revealed an estimated three-form test-retest reliability of .90 for correct letter sounds and .88 for whole words read. Initial interrater reliability scores were all correlated .90 or higher for NWF (Good & Kaminski, 2011).

Concurrent validity correlations between DIBELS NWF with the Group Reading Assessment and Diagnostic (GRADE) ranged from .39 to .51 at the beginning of the year to .40 to .56 at the end of the year for students between kindergarten and second grade. Concurrent and predictive validity coefficients for NWF with later DIBELS scores fall in the moderate to strong range, with slightly higher correlations with DORF Words Correct than DORF Accuracy or Retell (Good & Kaminski, 2011).

Participants at the elementary school in the study had been given the DIBELS 6th Edition benchmark assessments in the years previous to when this study was conducted. During the study, DIBELS NEXT spring benchmark materials were used. The students' scores from these benchmark assessments were correlated with the fluency probes.

Procedures

The study was conducted at the end of the school year. Before the study began, a consent form was sent home to the parents of each third grade student (See Appendix E). The form asked permission for their students to participate in the study and for the testing to be tape recorded for inter-rater reliability purposes. If parent consent was obtained, the student was then asked to give their assent to participate in the study (See Appendix F).

After consent and assent were obtained, measures were administered during two sessions. During the first session, students were administered the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes in a counterbalanced order. When administering the probes, the examiner began by presenting the student with a page containing practice items and read the instructions. See Appendix G for an example of the examiner's instructions and practice items. When the student successfully completed the practice items, the examiner started timing the student for 1 minute and recorded any errors the student made while reading. If the student hesitates for more than 3 seconds on a word or word part, the examiner told the student the word or part of the word so that they could continue. At the end of the minute, the examiner said, "Stop" and marked the last sound produced by the student. After administering the probes, the examiner also administered DIBELS NWF and the TOWRE. The second session took place approximately two weeks after the first session. During this session, a second set of Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes were administered to the students in a counterbalanced order. DIBELS benchmarking took place at the school between testing sessions. Because DIBELS 6th Edition passages were used, students were administered DIBELS NEXT ORF during the second session.

Scoring Procedures

After the sessions were completed, the examiners calculated students' scores. Two scoring methods were used on the fluency probes to determine which method was a more accurate representation of phonics abilities and which one correlated higher with the reading measures used to assess students' abilities. In the first method of scoring, students received a point for every word part they said correctly (total sounds) on the Real Word Fluency and Combined Fluency probes. In the case of the Multisyllabic Word Fluency probe, students received credit for each syllable they read correctly. In the second method, students received credit for whole words read (WWR) without sounding out. A Student Assessment Tracking form was filled out for each student to keep track of the assessment dates, as well as the students' scores (See Appendix H).

Interrater Agreement

Interrrater agreement was computed for 25% of the probes. One examiner listened independently to audio-taped recordings of the probes. The examiner scored the probes while listening to the tapes. In order to determine interrater agreement, a ratio of lowest score rated over the higher score rated was calculated. To establish a percentage of agreement between the two scorers, the ratio was then multiplied by 100.

CHAPTER III

RESULTS

Test-retest Reliability

Alternate-form test-retest reliability was established approximately two weeks after the initial fluency probes were administered with a mean of 14 days (*SD* = 1.5) between testing sessions. Test-retest correlations for each of the fluency measures appear below in Tables 1 and 3 for second graders and Tables 2 and 4 for third graders. When comparing the total sounds students read, all correlations were above .82 indicating the scores of all of the probes are stable over time for both second and third graders. For second graders, the Multisyllabic Word Fluency probe had a correlation of .91 which was the highest test-retest correlation of the three fluency probes. For third graders, the Combined Fluency probe read had a correlation of .84 which was the highest for that grade.

Whole words read correlations for both second and third graders were slightly lower than their total sounds correlations. For both second and third graders, the Combined Fluency Probe had the highest correlations on whole words read, .87 and .84 respectively.

Salvia and Ysseldyke (2004) suggests that adequate test-retest reliability for screening measures is .80. Adhering to these standards, all three measures for second graders met these standers. For third graders, all but one measure met the test-retest criterion needed to be considered stable measures of student’s skills. The Multisyllabic Word Fluency probe whole words read was a .75, which was below the criterion. In addition, the other two measures of whole words read barely made the criterion which suggests that using the total sounds may be a more reliable measure for third grade students.

Table 1. *Second Grade Test-retest Reliabilities for Total Sounds (N = 90)*

Fluency Probe	RWF 1	CF1	MWF1
RWF 2	.86**		
CF 2		.87**	
MWF 2			.91**

Note. RWF = Real Word Fluency probe, CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; ***p* < .001.

Table 2. *Third Grade Test-retest Reliabilities for Total Sounds (N = 72)*

Fluency Probe	RWF 1	CF1	MWF1
RWF 2	.82**		
CF 2		.84**	
MWF 2			.82

Note. RWF = Real Word Fluency probe, CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; ***p* < .001.

Table 3. *Second Grade Test-retest Reliabilities for Whole Words Read (N = 90)*

Fluency Probe	RWF 1	CF1	MWF1
RWF 2	.84**		
CF 2		.86**	
MWF 2			.81**

Note. RWF = Real Word Fluency probe, CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; ***p* < .001.

Table 4. *Third Grade Test-retest Reliabilities for Whole Words Read (N = 72)*

Fluency Probe	RWF 1	CF1	MWF1
RWF 2	.80**		
CF 2		.81**	
MWF 2			.75**

Note. RWF = Real Word Fluency probe, CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; ***p* < .001.

Interrater Agreement

Twenty-five percent of the probes were randomly recorded and scored for interrater agreement. Results showed that the mean percentage of agreement for scoring the different fluency measures were all above 82% with an average of 94%. When comparing the agreement when total sounds were scored, the mean interrater agreement was the highest for the Real Word Frequency probe, 97% (*SD* = 4.24) for second grade and 98% (*SD* = 1.28) for third grade. This was followed closely by the Combined Fluency probe at 96% (*SD* = 3.19) for second graders and

97% (*SD* = 2.04) for third graders. See Tables 5 and 6 for the Interrater agreement when total sounds were scored. Results from Table 7 and 8 indicate when whole words read were scored the interrater agreement for second graders was the highest for the Combined Fluency probe (*M* = 95%, *SD* = 5.93). For third graders the Real Word Fluency probe was the highest (*M* = 96.99%, *SD* = 2.75). However, this was only .73% higher than the Combined Fluency probe. For both second and third graders, the Multisyllabic Word Fluency probe had the lowest interrater agreement, especially for whole words read. This may be because the sample for whole words read was small, even as low as 1 word read. Because of this, differences in agreement made a bigger impact. The result of this is a range of interrater agreement from 0%, in which one rated had scored 1 whole words read and the other scored 0 whole words read to 100% agreement.

Table 5. *Second Grade Interrater Agreement for Fluency Probes Total Sounds (N = 90)*

Fluency Probe	Range	*M*	*SD*
RWF	79.17-100	97.21	4.24
CF	87.95-100	96.99	3.19
MWF	62.50-100	93.59	7.58

Note. RWF = Real Word Fluency probe, CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe. The Range and Mean values represent percentages.

Table 6. *Third Grade Interrater Agreement for Fluency Probes Total Sounds (N = 72)*

Fluency Probe	Range	*M*	*SD*
RWF	96.08-100	98.42	1.28
CF	94.38-100	97.39	2.04
MWF	75.86-100	96.13	5.65

Note. RWF = Real Word Fluency probe, CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe. The Range and Mean values represent percentages.

Table 7. *Second Grade Interrater Agreement for Fluency Probes Whole Words Read (N = 90)*

Fluency Probe	Range	*M*	*SD*
RWF	75-100	94.58	7.01
CF	87.95-100	95	5.93
MWF	0-100	82.33	22.70

Note. RWF = Real Word Fluency probe, CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe. The Range and Mean values represent percentages.

Table 8. *Third Grade Interrater Agreement for Fluency Probes Whole Words Read (N = 72)*

Fluency Probe	*Range*	*M*	*SD*
RWF	90.91-100	96.99	2.75
CF	87.5-100	96.28	4.19
MWF	25-100	88.39	17.81

Note. RWF = Real Word Fluency probe, CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe. The Range and Mean values represent percentages.

Means and Standard Deviations

Tables 9 and 10 present the means and standard deviations for the fluency probes. The Real Word Fluency probe appears to have the highest mean score, indicating that it provides the largest sample of behavior. In addition, it also has the highest standard deviation of the fluency probes which indicates that the distribution of student's scores is more spread out for the Real Word Fluency probes than the other fluency measures. Of the three fluency measures, the Multisyllabic Word Fluency probe had the lowest means and standard deviations.

Means and standard deviations for each of the standardized reading measures are presented in Table 11. For DIBELS NEXT NWF, the last benchmark goal for the fall of second grade for correct letter sounds is 54 sounds per minute. Second graders in the study received an average score of 117.09 (*SD* = 49.41) sounds per minute and third graders had an average score of 131.82 (*SD* = 45.52). For whole words read the benchmark is 13 words in one minute. Second graders participating in the study read an average of 36.80 words (*SD* = 19.86) and third graders read 40.85 words (*SD* = 17.93). Mean scores for second grade spring DORF was 108.72 (*SD* = 34.31) words per minute, well above DIBELS benchmark goal of 87 words per minute. Third grade students were also well above their spring benchmark goal of 100 words per minute reading an average of 123 words (*SD* = 13.66).

Average standard scores on the TOWRE fall between the range of 90 to 110. According to the student's scores in the study, the mean for the TOWRE Sight Word Efficiency subtest was 109.96 (*SD* = 23.28) for second graders and 109.21 (*SD* = 9.83) for third graders. On the TOWRE Phoneme Decoding Efficiency subtest, both second and third graders mean standard scores were in the average range. For second graders, their average standard score was a 107.87 (*SD* = 11.81). Third graders received an average standards score of 106.08 (*SD* = 10.61). The average TOWRE Total Word Reading Efficiency scores were on the high end of the average range for both second and third graders. Second graders had a mean of 110.74 (*SD* = 13.84) and third graders had a mean of 109 (*SD* = 11.61).

Table 9. *Second Grade Means and Standard Deviations for Fluency Measures (N = 90)*

	First Set of Probes		*Second Set of Probes*	
Fluency Measure	*M*	*SD*	*M*	*SD*
RWF Total	96.92	44.28	94.49	41.55
RWF WWR	36.51	19.92	33.96	17.88
CF Total	74.42	33.34	71.49	36.82
CF WWR	26.90	14.96	25.17	16.08
MWF Total	46.86	24.49	45.59	26.57
MWF WWR	9.44	7.63	8.79	8.34

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read; *M* = mean; *SD* = standard deviation.

Table 10. *Third Grade Means and Standard Deviations for Fluency Measures (N = 72)*

	First Set of Probes		*Second Set of Probes*	
Fluency Measure	*M*	*SD*	*M*	*SD*
RWF Total	123.89	44.81	118.35	38.25
RWF WWR	48.03	20.36	43.25	17.42
CF Total	91.56	32.44	90.60	36.35
CF WWR	33.74	15.10	32.43	15.85
MWF Total	60.31	23.75	59.61	23.23
MWF WWR	12.93	8.39	12.69	8.00

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read; *M* = mean; *SD* = standard deviation.

Table 11. *Second Grade Means and Standard Deviations for Standardized Reading Measures (N = 90)*

Standardized measure	*M*	*SD*
DIBELS NWF CLS	117.07	49.41
DIBELS NWF WWR	36.80	19.86
Spring DORF	108.72	34.31
TOWRE SWE	109.96	12.18
TOWRE PDE	107.87	11.81
TOWRE Total	110.74	13.84

Note: NWF = Nonsense Word Fluency; CLS = Correct Letter Sounds; WWR = Whole Words Read; DORF = DIBELS Oral Reading Fluency; TOWRE = Test of Word Reading Efficiency; SWE = Sight Word Efficiency; PDE = Phonemic Decoding Efficiency; Total = Total Word Reading Efficiency; *M* = mean; *SD* = standard deviation.

Table 12. *Third Grade Means and Standard Deviations for Standardized Reading Measures (N = 72)*

Standardized measure	*M*	*SD*
DIBELS NWF CLS	131.82	45.52
DIBELS NWF WWR	40.85	17.93
Spring DORF	123.97	31.66
TOWRE SWE	109.21	9.83
TOWRE PDE	106.08	10.61
TOWRE Total	109.07	11.61

Note: NWF = Nonsense Word Fluency; CLS = Correct Letter Sounds; WWR = Whole Words Read; DORF = DIBELS Oral Reading Fluency; TOWRE = Test of Word Reading Efficiency; SWE = Sight Word Efficiency; PDE = Phonemic Decoding Efficiency; Total = Total Word Reading Efficiency; *M* = mean; *SD* = standard deviation.

Concurrent Validity

The correlations between the fluency measures and the standardized reading measures are presented in Tables 15 through 18.

TOWRE Sight Word Efficiency Subtest

It would be expected that the Real Word Fluency probe would be correlated the highest with the Sight Word Efficiency subtest. This prediction was true at both grade levels for both total sounds and whole words read. In addition, both the total sound and whole words read for the Real Word Fluency probes had the highest correlations. Both second and third grade had correlations of .77 for the Real Word Fluency total sounds. For Real Word Fluency whole words read, second graders had a correlation of .77 and third graders had a correlation of .78. Not only did the Real Word Fluency probe have the highest correlations of the fluency probes, but the correlations were similar for both grade levels, indicating the concurrent validity of the measure is stable across both grades.

TOWRE Phoneme Decoding Efficiency Subtest (PDE)

Since the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes were designed to measure students' phonics skills and fluency they would be expected to correlate significantly with the TOWRE Phoneme Decoding subtest which is a phonics measure. Furthermore, it would be anticipated that the Multisyllabic Word Fluency probe, a measure of the ability to read longer words with multiple syllables, would have a lower correlation with the subtest. This is predicted because the more difficult words with multiple syllables are later in the subtest and fewer students would get to that section of the test in order to use their phonics skills to sound out the longer words on the test. The correlations of the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes with the Phonemic Decoding Efficiency subtest supported the prediction and were significant ($p < .001$). For both second and third graders, the Combined Fluency probes total sounds and whole words read had the highest correlations. In second grade, the Phonemic Decoding Efficiency subtest had a correlation of .88 with the Combined Fluency probes whole words read .87 for total sounds. Third grade had a correlation of .77 for both total sounds and whole words read.

Although all the correlations were significant, correlations with the subtest were higher on all the fluency measures for second graders than for third graders. This indicates that the concurrent validity the fluency probes has with other phonics measures may lose its strength in older grade levels.

It would be predicted that the Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes would correlate lower with the TOWRE Sight Word Efficiency subtest than the Phonemic Decoding Efficiency subtest. In order to test this hypothesis a series of one-tailed t-tests were conducted to test whether there were significant differences between the TOWRE Sight Word Reading Efficiency and Phonemic Decoding Efficiency correlations for each fluency measure. The results from these *t*-tests are presented in Tables 13 and 14. None of the fluency probes showed significant differences between their correlations with the TOWRE Sight Word Efficiency and Phonemic Decoding Efficiency subtests.

Table 13. *Second Grade Testing Significant Differences Between TOWRE Subtests Correlations (N = 90)*

Fluency Probe	*t*	*p*	*Significance*
RWF Total	-2.47	.99	*ns*
RWF WWR	-2.85	1.00	*ns*
CF Total	-4.18	1.00	*ns*
CF WWR	-5.00	1.00	*ns*
MWF Total	-3.15	1.00	*ns*
MWF WWR	-2.80	1.00	*ns*

Note. RWF = Real Word Fluency probe, CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe, Total = Total Sounds; WWR = Whole Words Read; *ns* = not significant.

Table 14. *Third Grade Testing Significant Differences Between TOWRE Subtests Correlations (N = 72)*

Fluency Probe	*t*	*p*	*Significance*
RWF Total	.56	.29	*ns*
RWF WWR	.20	.42	*ns*
CF Total	-1.15	.87	*ns*
CF WWR	-1.33	.91	*ns*
MWF Total	.17	.57	*ns*
MWF WWR	-.67	.75	*ns*

Note. RWF = Real Word Fluency probe, CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe, Total = Total Sounds; WWR = Whole Words Read; *ns* = not significant.

TOWRE Total Word Reading Efficiency

The TOWRE measures a student's ability to pronounce words accurately and fluently. Because the test begins with easy words and the difficulty increases the further students read, it would be predicted that the Real Word Fluency and Combined Fluency probes would have highest correlation. The Multisyllabic Word Fluency probe, which contains more difficult words, would be predicted to have the lowest correlations of the three fluency probes. This prediction was true for both second and third graders. Furthermore, the correlations for the Real Word Fluency and Combined Fluency probes were almost identical with no more than a .01 difference in the correlations at each grade level.

DIBELS Nonsense Word Fluency

Given that DIBELS NWF measures students' phonic skills it was hypothesized that the both the Real Word Fluency and Combined Fluency probes should be significantly correlated. In addition, it would be expected that the Multisyllabic Word Fluency probes would correlate lower than the other probes. The prediction was demonstrated and all the fluency probes were significantly correlated at the .001 level with DIBELS NWF. When comparing the probes to NWF Correct Letter Sounds (CLS), the Real Word Fluency and Combined Fluency probes correlations were the same and higher than the correlations with the Multisyllabic Word Fluency probes. For third graders, the Real Word Fluency probes correlations were the highest.

When looking at the correlations with DIBELS NWF Whole Word Read (WWR), the Real Word Fluency probe (both total sounds and whole words read) had the highest correlations for second grade and the Combined Fluency probe total sounds had the highest correlation for third grade. In addition, even though the Multisyllabic Word Fluency probe was significantly correlated with DIBELS NWF CLS and WWR it had the lowest correlations of the fluency probes which was expected.

DIBELS Spring Oral Reading Fluency (DORF)

When looking at the correlations of the fluency probes from the DIBELS NEXT spring DORF benchmarks, all probes for second and third graders had significant correlations. For second graders, the Multisyllabic Word Fluency probe total sounds had the highest correlations ($r = .78$, $p < .001$). The Combined Fluency probe had the lowest correlations. However, all correlations were still significant at the .001 level. When looking at the third graders correlations, the Real Word probe both total and whole words read had the highest correlations ($r = .82$, $< .001$). The Multisyllabic Word Fluency probe whole words read had the lowest correlation of .76. Overall, the third graders' DORF scores were correlated slightly higher with the fluency probes than the second graders' scores.

Whether the scores from any of the fluency probes could explain a significant amount of variance in spring DORF scores can be assessed by squaring the correlations between the fluency probes and DIBELS ORF. This indicates that for second grade students, the fluency probes individually explained 52% to 61% of the variance in ORF scores. For third graders, the fluency probes individually explained 58% to 67% of the variance in ORF scores. These percentages indicate the fluency measures explained a significant amount of variance in spring DORF scores and are significant at the .001 level.

Correlations Between DIBELS and the TOWRE

In addition to looking at the correlations of the fluency measures, the correlations between DIBELS NWF, ORF and the TOWRE were also calculated. It would be predicted that

DIBELS NWF would correlate the highest with the TOWRE Phonemic Decoding Efficiency (PDE) subtest. In addition, it would be predicted that DIBELS NWF scores would correlate significantly with DIBELS ORF scores.

For both second and third graders, DIBELS NWF was significantly correlated with DIBELS ORF as well as the TOWRE. Additionally, DIBELS ORF was also significantly correlated with the TOWRE for both second and third graders. The Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes had higher correlations with DIBELS ORF than DIBELS NWF had with DIBELS ORF for both second and third graders.

When looking at the correlations between DIBELS NWF and the TOWRE, the prediction was true for second graders, but not third graders. For second graders, NWF correlated the highest with the TOWRE PDE subtest. For third graders, NWF correlated the highest with the TOWRE Total. However, all correlations were significant at the .05 level.

DIBELS ORF was also significantly correlated with the TOWRE for both second and third graders. Specifically, for second graders ORF had the highest correlations with the TOWER Total. For third graders, ORF correlated higher with the TOWRE Total and SWE subtest than the PDE subtest.

Table 15. *Second Grade Correlations Between Fluency Probes Total Sounds and Standardized Measures (N = 90)*

	RWF	CF	MWF	NWF CLS	NWF WWR	Spring DORF	TOWRE PDE	TOWRE SWE	TOWRE Total
RWF	1								
CF	.89**	1							
MWF	.87**	.88**	1						
NWF CLS	.83**	.83**	.82**	1					
NWF WWR	.83**	.80**	.80**	.93**	1				
Spring DORF	.74**	.72**	.78**	.60**	.66**	1			
TOWRE PDE	.85**	.87**	.83**	.81**	.83**	.79**	1		
TOWRE SWE	.77**	.74**	.72**	.66**	.67**	.75**	.83**	1	
TOWRE Total	.85**	.84**	.81**	.76**	.78**	.80**	.95**	.96**	1

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; NWF = DIBELS NEXT Nonsense Word Fluency; CLS = Correct Letter Sounds; DORF = DIBELS NEXT Spring Oral Reading Fluency; TOWRE = Test of Word Reading Efficiency; PDE = Phonemic Decoding Efficiency; SWE = Sight Word Efficiency; Total = TOWRE Total Word Reading Efficiency; $^{**}p < .001$, $^{*}p < .05$.

Table 16. *Third Grade Correlations Between Fluency Probes Total Sounds and Standardized Measures (N = 72)*

	RWF	CF	MWF	NWF CLS	NWF WWR	Spring DORF	TOWRE PDE	TOWRE SWE	TOWRE Total
RWF	1								
CF	.87**	1							
MWF	.81**	.80**	1						
NWF CLS	.75**	.75**	.70**	1					
NWF WWR	.68**	.69**	.64**	.67**	1				
Spring DORF	.82**	.80**	.80**	.59**	.53**	1			
TOWRE PDE	.74**	.79**	.71**	.76**	.74**	.68**	1		
TOWRE SWE	.77**	.73**	.70**	.69**	.64**	.77**	.73**	1	
TOWRE Total	.82**	.82**	.76**	.78**	.74**	.77**	.94**	.91**	1

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; NWF = DIBELS NEXT Nonsense Word Fluency; CLS = Correct Letter Sounds; DORF = DIBELS NEXT Spring Oral Reading Fluency; TOWRE = Test of Word Reading Efficiency; PDE = Phonemic Decoding Efficiency; SWE = Sight Word Efficiency; Total = TOWRE Total Word Reading Efficiency; **$p < .001$, *$p < .05$.

Table 17. *Second Grade Correlations Between Fluency Probes Whole Word Reads and Standardized Measures (N = 90)*

	RWF	CF	MWF	NWF CLS	NWF WWR	Spring DORF	TOWRE PDE	TOWRE SWE	TOWRE Total
RWF	1								
CF	.90**	1							
MWF	.83**	.84**	1						
NWF CLS	.82**	.82**	.76**	1					
NWF WWR	.83**	.80**	.75**	.93**	1				
Spring DORF	.75**	.73**	.74**	.60**	.66**	1			
TOWRE PDE	.86**	.90**	.78**	.81**	.83**	.79**	1		
TOWRE SWE	.77**	.73**	.66**	.66**	.67**	.75**	.83**	1	
TOWRE Total	.85**	.84**	.75**	.76**	.78**	.80**	.95**	.96**	1

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; NWF = DIBELS NEXT Nonsense Word Fluency; CLS = Correct Letter Sounds; DORF = DIBELS NEXT Spring Oral Reading Fluency; TOWRE = Test of Word Reading Efficiency; PDE = Phonemic Decoding Efficiency; SWE = Sight Word Efficiency; Total = TOWRE Total Word Reading Efficiency; **$p < .001$, *$p < .05$.

Table 18. *Third Grade Correlations Between Fluency Probes Whole Words Read and Standardized Measures (N = 72)*

	RWF	CF	MWF	NWF CLS	NWF WWR	Spring DORF	TOWRE PDE	TOWRE SWE	TOWRE Total
RWF	1								
CF	.88**	1							
MWF	.78**	.76**	1						
NWF CLS	.76**	.73**	.64**	1					
NWF WWR	.70**	.68**	.62**	.67**	1				
Spring DORF	.82**	.80**	.79**	.59**	.53**	1			
TOWRE PDE	.77**	.79**	.72**	.76**	.74**	.68**	1		
TOWRE SWE	.78**	.72**	.68**	.69**	.64**	.77**	.73**	1	
TOWRE Total	.83**	.82**	.76**	.78**	.74**	.77**	.94**	.91**	1

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; NWF = DIBELS NEXT Nonsense Word Fluency; CLS = Correct Letter Sounds; DORF = DIBELS NEXT Spring Oral Reading Fluency; TOWRE = Test of Word Reading Efficiency; PDE = Phonemic Decoding Efficiency; SWE = Sight Word Efficiency; Total = TOWRE Total Word Reading Efficiency; ***p* < .001, **p* < .05.

Concurrent Validity for Students Below Benchmark

Overall, the goal is to determine if one of the fluency measures may be useful for progress monitoring students who are receiving interventions and monitor their growth in reading skills. Because of this, it is important to examine the concurrent validity of the fluency probes for students who are below DIBELS ORF benchmark. Twenty-six of the ninety second grade students (25.56%) and 16 of the seventy-two (22.22%) third grade students who participated in the study were below the DIBELS ORF benchmark. The correlations between the fluency measures and the standardized reading measures for students who scored below the DIBELS Spring ORF benchmark are presented in Tables 19 through 22.

For students below the DIBELS ORF Spring Benchmark, the fluency measures continued to have high correlations with the standardized reading measures. In general, for students below benchmark, the fluency probes had higher correlations with DIBELS NWF CLS and WWR than with DIBELS ORF. This indicates that for students below benchmark, the fluency probes have stronger concurrent validity with other phonics measures than measures of overall reading.

For second graders, all of the correlations were significant at the .001 level except for the correlations between 2nd Grade Spring ORF and the Combined Fluency Probe total sounds and the Multisyllabic Word Fluency Whole Words Read. However, these correlations were significant at the .05 level. Correlations with DIBELS NWF for second grade students below benchmark were similar to the correlations found with all students. However for second grade students below benchmarks, the correlations with DIBELS ORF and the TOWER were slightly lower than the correlations for all second grade students. These correlations indicate that concurrent validity for the fluency probes is strong for second grade students who are below the DIBELS Spring ORF benchmark.

The data for third grade students is more variable. For third graders, below the DIBELS Spring ORF benchmark, the fluency probes correlated significantly at the .001 level with DIBELS NWF CLS, WWR, TOWRE Phonemic Decoding Efficiency subtest, and the TOWRE Total. However, the correlations were lower with the 3rd grade DIBELS Spring ORF and the TOWRE Sight Word Efficiency subtest. In fact these correlations were only significant at the .05 level. For third grade students below benchmark, the Combined Fluency probe total sounds and the Multisyllabic Word Fluency probe Whole Words Read did not correlate significantly with the TOWRE Sight Word Efficiency subtest. This data indicates that especially for third grade students below benchmark, the concurrent validity is the strongest with other phonics measures. When looking at the correlations, the Real Word Fluency probe and the Combined Fluency probe had the highest correlations with the standardized reading measures. The Multisyllabic Word Fluency probe had the lowest correlations with the standardized reading measures. In addition, the correlations between the Multisyllabic Word Fluency probe and the TOWRE were lower than the other fluency probes. This indicates that the Multisyllabic Word Fluency probe may not be the best measure for students below benchmark.

For both second and third graders below benchmark, DIBELS NWF was significantly correlated with DIBELS ORF as well as the TOWRE. Additionally, DIBELS ORF was also significantly correlated with the TOWRE for both second and third graders. The Real Word Fluency, Combined Fluency, and Multisyllabic Word Fluency probes had higher correlations with DIBELS ORF than DIBELS NWF had with DIBELS ORF for second graders. For third graders, the correlations between DIBELS NWF Correct Letter Sounds were better with DIBELS ORF than the Combined Fluency probe Whole Words Read and the Multisyllabic Word Fluency probe Whole Words Read.

For students below benchmark, NWF correlated the highest with the TOWRE PDE subtest than the SWE or TOWRE Total. For second graders, ORF correlated highest with the TOWRE Total and PDE subtests. For third graders, ORF correlated the highest with the TOWRE SWE subtest.

Again, whether the scores from any of the fluency probes could explain a significant amount of variance in spring DORF scores of students who were below the DORF benchmark can be assessed by squaring the correlations between the fluency probes and DIBELS ORF. This indicates that for second grade students, most of the fluency probes individually explained 37% to 41% of the variance in ORF scores. However, the Multisyllabic Word Fluency probe only explained 25% of the variance in ORF scores. For third graders, the fluency probes individually explained 33% to 46% of the variance in ORF scores. These percentages indicate that for students below the ORF benchmark the fluency measures explained a significant amount of variance in spring DORF scores. For students below benchmark, the amount of variance explained was less than the amount of variance in DIBELS ORF scores that was explained by all students. However, all measures explained a significant amount of variance in spring DORF scores at the .05 level.

Table 19. *Correlations Between Fluency Probes Total Sounds and Standardized Measures for Second Graders Below Benchmark (n = 23)*

	RWF	CF	MWF	NWF CLS	NWF WWR	Spring DORF	TOWRE PDE	TOWRE SWE	TOWRE Total
RWF	1								
CF	.85**	1							
MWF	.84**	.87**	1						
NWF CLS	.86**	.81**	.82**	1					
NWF WWR	.84**	.78**	.80**	.95**	1				
Spring DORF	.64**	.61**	.65**	.50*	.57**	1			
TOWRE PDE	.69**	.82**	.75**	.80**	.82**	.48*	1		
TOWRE SWE	.66**	.70**	.72**	.64**	.64**	.44*	.76**	1	
TOWRE Total	.72**	.81**	.79**	.76**	.77**	.48*	.93**	.95**	1

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; NWF = DIBELS NEXT Nonsense Word Fluency; CLS = Correct Letter Sounds; DORF = DIBELS NEXT Spring Oral Reading Fluency; TOWRE = Test of Word Reading Efficiency; PDE = Phonemic Decoding Efficiency; SWE = Sight Word Efficiency; Total = TOWRE Total Word Reading Efficiency; ***p* < .001, **p* < .05.

Table 20. *Correlations Between Fluency Probes Total Sounds and Standardized Measures for Third Graders Below Benchmark (n = 16)*

	RWF	CF	MWF	NWF CLS	NWF WWR	Spring DORF	TOWRE PDE	TOWRE SWE	TOWRE Total
RWF	1								
CF	.86**	1							
MWF	.83**	.84**	1						
NWF CLS	.88**	.90**	.78**	1					
NWF WWR	.82**	.89**	.78**	.97**	1				
DORF	.68**	.64**	.65**	.62**	.56*	1			
TOWRE PDE	.81**	.83**	.71**	.89**	.85**	.54*	1		
TOWRE SWE	.68**	.49*	.47	.65**	.58*	.60*	.61*	1	
TOWRE Total	.82**	.74**	.68*	.78**	.78**	.59*	.92**	.84**	1

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; NWF = DIBELS NEXT Nonsense Word Fluency; CLS = Correct Letter Sounds; DORF = DIBELS NEXT Spring Oral Reading Fluency; TOWRE = Test of Word Reading Efficiency; PDE = Phonemic Decoding Efficiency; SWE = Sight Word Efficiency; Total = TOWRE Total Word Reading Efficiency; ** $p < .001$, * $p < .05$.

Table 21. *Correlations Between Fluency Probes Whole Words Read and Standardized Measures for Second Graders Below Benchmark (n = 23)*

	RWG	CF	MWF	NWF CLS	NWF WWR	Spring DORF	TOWRE PDE	TOWRE SWE	TOWRE Total
RWF	1								
CF	.87**	1							
MWF	.77**	.86**	1						
NWF CLS	.85**	.81**	.74**	1					
NWF WWR	.84**	.80**	.69**	.95**	1				
Spring DORF	.64**	.65**	.50*	.50*	.57**	1			
TOWRE PDE	.70**	.82**	.67**	.80**	.82**	.48*	1		
TOWRE SWE	.67**	.68**	.67**	.64**	.64**	.44*	.76**	1	
TOWRE Total	.73**	.80**	.72**	.76**	.77**	.48*	.93**	.95**	1

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; NWF = DIBELS NEXT Nonsense Word Fluency; CLS = Correct Letter Sounds; DORF = DIBELS NEXT Spring Oral Reading Fluency; TOWRE = Test of Word Reading Efficiency; PDE = Phonemic Decoding Efficiency; SWE = Sight Word Efficiency; Total = TOWRE Total Word Reading Efficiency; $^{**}p < .001$, $^{*}p < .05$.

Table 22. *Correlations Between Fluency Probes Whole Words Read and Standardized Measures for Third Graders Below Benchmark (n = 16)*

	RWF	CF	MWF	NWF CLS	NWF WWR	Spring DORF	TOWRE PDE	TOWRE SWE	TOWRE Total
RWF	1								
CF	.88**	1							
MWF	.76**	.74**	1						
NWF CLS	.90**	.93**	.72**	1					
NWF WWR	.85**	.93**	.77**	.97**	1				
DORF	.64**	.57*	.60*	.62**	.56*	1			
TOWRE PDE	.83**	.84**	.65**	.89**	.85**	.54*	1		
TOWRE SWE	.68**	.53*	.56*	.65**	.58*	.60*	.61*	1	
TOWRE Total	.84**	.77**	.69**	.84**	.78**	.59*	.92**	.84**	1

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; NWF = DIBELS NEXT Nonsense Word Fluency; CLS = Correct Letter Sounds; DORF = DIBELS NEXT Spring Oral Reading Fluency; TOWRE = Test of Word Reading Efficiency; PDE = Phonemic Decoding Efficiency; SWE = Sight Word Efficiency; Total = TOWRE Total Word Reading Efficiency; ***p* < .001, **p* < .05.

Correlations between Fluency Probes

Since the Real Word, Combined Fluency, and Multisyllabic Word Fluency probes are designed to be different methods of measuring students' phonics and fluency skills the probes should be highly correlated with each other. In addition, it would be expected that the Multisyllabic Word Fluency probe would have a lower correlation with the probes because it is testing a slightly different reading skill involving reading longer words with multiple syllables. Correlations from Tables 13 through 16 indicate that the fluency probes are all highly correlated with each other. The correlations with the Multisyllabic Word Fluency probe were only slightly lower than the correlations between the Real Word Fluency and Combined Fluency probes. This indicates that the Multisyllabic Word Fluency probes are measuring skills similar to the other fluency probes. Furthermore, the correlations between the fluency probes are similar to the correlations between the fluency probes and the standardized measures. This indicates the fluency probes, DIBELS, and the TOWRE are all testing fluency skills.

Past DIBELS Data

Students' past and present reading performance should be correlated with the fluency probes. Tables 23 and 24 display the correlations between the students' fluency probes and *DIBELS* data from their previous grade and their current grade. It would be expected that the Real Word and Combined Fluency probes would be correlated the highest with the DIBELS NWF. The Multisyllabic Word Fluency probe would have the lowest correlations since it requires reading longer words. In addition, all of the fluency probes should be correlated with DIBELS ORF. It would be expected that the whole words read scores would be higher correlated since it is measuring different types of word reading. Second graders' scores were also correlated with their DIBELS Phoneme Segmentation Fluency (PSF) scores from first grade. The fluency probes would be expected to have lower correlations with PSF since it is a measure of phonological awareness where the fluency probes are a measure of phonic skills.

Overall, for both second and third graders, all three fluency probes were highly correlated with their DIBELS NWF and ORF data. For the Real Word and Combined Fluency probes, the whole words read score had slightly higher correlations with both their DIBELS NWF and ORF scores than the total sounds score. This was expected for ORF scores, but not for NWF scores. However, neither the whole words read nor total sounds scores for the Multisyllabic Word

Fluency probe showed consistently higher correlations than the other on students' NWF and ORF scores.

For second graders, the fluency probes had low correlations with DIBELS PSF. In fact after the Fall PSF, negative correlations were found. This indicates that the students' phonics skills being assessed on the fluency probes tested are not related to their past phonological awareness skills. As expected, all three fluency probes were correlated with students' NWF data. Correlations with NWF were high across 1st grade, but the correlations dropped at the beginning of second grade. However, the correlations between the fluency probes and their second grade fall NWF score were still highly correlated. Even though it was expected that the Multisyllabic Word Fluency probe would have the lowest correlations, that prediction was only partly true. The Multisyllabic Word Fluency probes' correlations were the lowest of the three probes in the winter and spring of first grade, but the highest for the Fall of 1st grade and similar to the Real Word Fluency correlations for Fall of 2nd grade. When looking at the correlations with ORF, correlations with the fluency probes were high and remained high throughout their first and second grade years. Of the three probes, the Multisyllabic Word Fluency probe had the highest correlations with ORF scores than the other two probes but only by a small difference.

For third grade students, again all correlations with both DIBELS NWF and ORF were highly correlated. The fluency probes had higher correlations with the ORF scores than with NWF scores. When looking at the correlations with their 2nd grade Fall NWF scores, correlations for the third graders were similar to the correlations of the second graders in the study. For the third graders, the Real Word Fluency probe had the highest correlations with both DIBELS NWF and ORF scores. The Multisyllabic Word Fluency Probe had the lowest correlations, which was unexpected.

Table 23. *Second Grade Correlations Between Fluency Probes (Total and Whole Words Read) and Past DIBELS Data (N = 90)*

Fluency Probe	1st Grade								2nd Grade		
	F PSF	F NWF	W PSF	W NWF	W ORF	S PSF	S NWF	S ORF	F NWF	F ORF	W ORF+
RWF Total	.22*	.65**	-.79	.77**	.72**	-.49	.71**	.80**	.53**	.77**	.76**
RWF WWR	.23*	.67**	-.07	.79**	.75**	-.05	.72**	.82**	.54**	.80**	.78**
CF Total	.20	.65**	-.01	.74**	.72**	-.04	.67**	.78**	.51**	.74**	.77**
CF WWR	.20	.69**	-.01	.77**	.76**	-.04	.70**	.82**	.51**	.77**	.79**
MWF Total	.23*	.70**	-.02	.73**	.77**	.01	.65**	.82**	.54**	.79**	.79**
MWF WWR	.20	.69**	-.05	.73**	.78**	-.06	.64**	.82**	.47**	.77**	.77**

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read; F = Fall; W = Winter; S = Spring; PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; + = DIBELS NEXT; ***p* < .001, **p* < .05.

Table 24. *Third Grade Correlations Between Fluency Probes (Total and Whole Words Read) and Past DIBELS Data (N = 72)*

Fluency Probe	2nd Grade				3rd Grade	
	F NWF	F ORF	W ORF	S DORF	F ORF	W ORF
RWF Total	.55**	.70**	.80**	.80**	.73**	.82**
RWF WWR	.56**	.72**	.82**	.81**	.74**	.83**
CF Total	.51**	.69**	.80**	.78**	.70**	.78**
CF WWR	.51**	.70**	.80**	.79**	.70**	.80**
MWF Total	.49**	.67**	.74**	.71**	.66**	.78**
MWF WWR	.48**	.69**	.73**	.69**	.64**	.76**

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read; F = Fall; W = Winter; S = Spring; PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; + = DIBELS NEXT; ***p* < .001, **p* < .05.

Multiple Regressions

TOWRE Phonemic Decoding Efficiency Subtest

Because the fluency probes were designed to be a measure of phonics skills, they should be able to explain a unique amount of variance in scores from other phonics measures beyond what could be explained by an overall reading measure alone. To test this, a series of hierarchical multiple regressions were run to determine whether scores from any of the fluency probes could explain significantly more of the variance in the TOWRE Phonemic Decoding Efficiency subtest standard scores over and above what the students' DIBELS NEXT Spring ORF scores explain. Table 25 displays the results of the regressions for second graders and Table 26 displays the results of the regressions for third graders.

According to the results, about 62% of the variance in second graders' Phonemic Decoding Efficiency subtest scores was explained by ORF, $R^2 = .62$, $F(1,88) = 143.41$, $p < .00$. All three fluency probes explained a significant amount of unique variance in the Phonemic Decoding Efficiency scores above what was explained by ORF alone. For second graders, the Combined Fluency probe explained the most significant amount of unique variance. Adding the Combined Fluency probe total sounds explained an additional 19% of the variance, $\Delta R^2 = .19$, $\Delta F(1,87) = 86.46$, $p = .00$. When looking at the whole words read on the Combined Fluency probe it also explained an additional 19%, $\Delta R^2 = .1p$, $\Delta F(1,87) = 93.45$, $p = .00$. In addition, DIBELS NWF also explained a significant amount of variance in Phonemic Decoding Efficiency subtest scores above what ORF explained alone. DIBELS NWF explained a similar amount of variance in PDE scores as the Real Word Fluency and Combined Fluency probes.

The regressions for third graders showed only 46% of the variance in the Phonemic Decoding Efficiency subtest scores was explained by ORF, $R^2 = .46$, $F(1,70) = 59.91$, $p < .00$. Similar to the second graders, all of the fluency probes for third grade explained a significant amount of unique variance in the Phonemic Decoding Efficiency subtest scores above what was explained by ORF alone. Again, the Combined Fluency probe explained the highest amount of unique variance of three fluency probes. For third graders, the Combined Fluency probe total sounds explained an additional 16% of the variance, $\Delta R^2 = .16$, $F(1,69) = 30.04$, $p < .00$. The Combined Fluency probe whole words read explained an additional 17% of the variance, $\Delta R^2 = .17$, $F(1,69) = 31.87$, $p < .00$.

DIBELS NWF also explained a significant amount of variance in Phonemic Decoding Efficiency subtest scores. For third graders, DIBELS NWF explained slightly more variances then the Real Word, Combined Fluency and Multisyllabic Word Fluency probes. Specifically, DIBELS NWF CLS explained an additional 20% of the variance in Phonemic Decoding Efficiency scores above what ORF was able to explain alone, $\Delta R^2 = .20$, $F(1,69) = 40.54$, $p < .00$. DIBELS NWF WWR also explained an additional 20%, $\Delta R^2 = .17$, $F(1,69) = 40.12$, $p < .00$.

Table 25. *Second Grade Regressions for Predicting TOWRE Phonemic Decoding Efficiency Subtest from DIBELS NEXT 2nd Grade Spring ORF (DORF) and Fluency Probes*

Variable	R^2	ΔR^2	*p*
Step 1	.62	--	.00**
DORF			
Step 2			
RWF Total + DORF	.78	.16	.00**
RWF WWR + DORF	.78	.16	.00**
CF Total + DORF	.81	.19	.00**
CF WWR + DORF	.81	.19	.00**
MWF Total + DORF	.74	.12	.00**
MWF WWR + DORF	.71	.08	.00**
NWF CLS + DORF	.80	.18	.00**
NWF WWR + DORF	.79	.17	.00**

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read; ***p* < .001.

Table 26. *Third Grade Regressions for Predicting TOWRE Phonemic Efficiency Decoding Subtest from DIBELS NEXT 3rd Grade Spring ORF (DORF) and Fluency Probes*

Variable	R^2	ΔR^2	*p*
Step 1	.46	--	.00**
DORF			
Step 2			
RWF Total + DORF	.56	.11	.00**
RWF WWR + DORF	.60	.14	.00**
CF Total + DORF	.62	.16	.00**
CF WWR + DORF	.63	.17	.00**
MWF Total + DORF	.54	.08	.00**
MWF WWR + DORF	.55	.09	.00**
NWF CLS + DORF	.66	.20	.00**
NWF WWR + DORF	.66	.20	.00**

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read; ***p* < .001.

TOWRE Total Word Reading Efficiency

The fluency measures should be highly related to other measures of reading fluency. Hierarchical multiple regressions were run to determine whether scores from any of the fluency probes could explain significantly more of the variance in TOWRE Total Word Reading

Efficiency scores above what the students' DIBELS NEXT Spring ORF scores explain. The results from the regressions are presented in Tables 27 and 28.

The regressions show that DORF was able to explain 65% ($R^2 = .65$, $F(1,87) = 157.88$, $p < .00$) of the variance in second grade scores and 60% ($R^2 = .60$, $F(1,70) = 104.73$, $p < .00$) of the variance in third grade scores. All three fluency probes were able to explain a significant amount of unique variance in TOWRE Total Word Reading Efficiency scores above what DORF explained alone. Second graders' data revealed that the Real Word Fluency and Combined Fluency probes total sounds and whole words read all explained an additional 14% of the variance in scores which was significant at the .001 level. Similarly, for third graders the Real Word Fluency and Combined Fluency probes explained the most additional variance (ranging from 10-12%) in TOWRE scores above what DORF explained alone. For both second and third graders, even though significant, the Multisyllabic Word Fluency probe explained the smallest amount of unique variance.

DIBELS NWF also explained a significant amount of additional variance in TOWRE Total scores above what DIBELS ORF was able to explain alone. For second graders, NWF explained more variance than the Multisyllabic Word Fluency probe, but not as much variance as the Real Word and Combined Fluency probes. For third graders, DIBELS NWF explained slightly more variance in TOWRE Total scores above what ORF explained alone than either the Real Wore, Combined Fluency, or Multisyllabic Word Fluency Probe. For third graders, DIBELS NWF CLS and WWR explained an additional 16% of unique variance in TOWRE Total scores above DIBELS ORF.

Table 27. *Second Grade Regressions for predicting TOWRE Total Word Reading Efficiency from DIBELS NEXT 2nd Grade Spring ORF (DORF) and Fluency Probes (N = 90)*

Variable	R^2	ΔR^2	*p*
Step 1	.65	--	.00**
DORF			
Step 2			
RWF Total + DORF	.79	.14	.00**
RWF WWR + DORF	.79	.14	.00**
CF Total + DORF	.79	.14	.00**
CF WWR + DORF	.79	.14	.00**
MWF Total + DORF	.73	.08	.00**
MWF WWR + DORF	.70	.05	.00**
NWF CLS + DORF	.77	.12	.00**
NWF WWR + DORF	.76	.11	.00**

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read; ** $p < .001$.

Table 28. *Third Grade Regressions for predicting TOWRE Total Word Reading Efficiency from DIBELS NEXT 3rd Grade Spring ORF (DORF) and Fluency Probes (N = 72)*

Variable	R^2	ΔR^2	*p*
Step 1	.60	--	.00**
DORF			
Step 2			
RWF Total + DORF	.70	.10	.00**
RWF WWR + DORF	.72	.12	.00**
CF Total + DORF	.71	.11	.00**
CF WWR + DORF	.71	.11	.00**
MWF Total + DORF	.66	.06	.00**
MWF WWR + DORF	.67	.07	.00**
NWF CLS + DORF	.76	.16	.00**
NWF WWR + DORF	.76	.16	.00**

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read; ** $p < .001$.

Overall, the goal is to determine if one of the fluency measures may be useful for progress monitoring students who are receiving interventions and monitor their growth in reading skills. Since this is done for students who are below the DIBELS benchmarks, one cannot look only at whether the fluency probes completed on all students can explain additional variance in reading scores on the TOWRE Total Word Reading Efficiency above what ORF explains. It is also critical to examine if additional amount of variance in reading scores on the

TOWRE Total Word Reading Efficiency can be explained above what ORF explains only for those students who are below benchmark, for whom the probes were made. To examine this further, a series of hierarchical multiple regressions were run using only students who tested below the DIBELS NEXT Spring ORF benchmark for their grade level. Results of the regressions appear in Tables 29 and 30 below.

Regressions indicate that DIBELS ORF alone explained more variability in TOWRE scores for all students than it did for only the students below benchmark. For second graders below the DIBELS benchmark, DORF was able to explain 23% ($R^2 = .23$, $F(1,21) = 6.22$, $p = .02$) of the variance in second graders' scores and 34% ($R^2 = .34$, $F(1,14) = 7.29$, $p = .02$) of the variance in third graders' TOWRE scores. Adding the fluency probes explained a significant amount of unique variance in TOWRE Total Word Reading Efficiency scores beyond what DORF explained alone for both second and third graders. The only measure that did not explain significantly more variance was the third grade students' Multisyllabic Word Fluency probe total sounds. When adding DIBELS NWF, it also explained a significant amount of unique variance in TOWRE Total Word Reading Efficiency scores beyond what DORF explained alone for both second and third graders.

For second graders, the Combined Fluency probe explained the highest amount of unique variance. The Combined Fluency probe total sounds explained an additional 42% ($\Delta R^2 = .65$, $F(1,20) = 24.03$, $p < .001$) of the variance in TOWRE Total Word Reading Efficiency scores above what DORF explained alone. The Combined Fluency probe whole words read score explained an additional 41% ($\Delta R^2 = .64$, $F(1,20) = 23.11$, $p < .001$) of the variance in TOWRE Total Word Reading Efficiency scores above what DORF explained alone. The Combined Fluency probe also explained more additional variance than DIBELS NWF. This was not the case for third graders. For third graders below benchmark of the three fluency probes, the Real Word Fluency probe explained the largest amount of additional variance in TOWRE Total Word Reading Efficiency scores above what DORF explained alone. The Real Word Fluency probe whole words read explained an additional, 38% ($\Delta R^2 = .72$, $F(1,13) = 23.11$, $p < .001$) of the variance in TOWRE Total Word Reading Efficiency scores above what DORF explained alone. For total sounds, the Real Word Fluency probe explained an additional 33% ($\Delta R^2 = .72$, $F(1,13) = 17.18$, $p < .001$) of the variance in TOWRE scores above what DORF was able to explain. However, for third graders, DIBELS NWF CLS was able to explain an equal amount of variance in scores as the Real Word Fluency probe Whole Words Read and more variance than the other fluency probes. DIBEL NWF CLS was able to explain an additional 38% of the variance in TOWRE Total Word Reading Efficiency scores above what DIBELS ORF was able to explain alone, $\Delta R^2 = .72$, $F(1,13) = 17.28$, $p < .001$.

Table 29. *Regressions for predicting TOWRE Total Word Reading Efficiency from DIBELS NEXT 2nd Grade Spring ORF (DORF) and Fluency Measures for Second Graders Below Benchmark (N = 90)*

Variable	R^2	ΔR^2	*p*
Step 1	.23	--	.02*
DORF			
Step 2			
RWF Total + DORF	.52	.29	.003*
RWF WWR + DORF	.53	.30	.002*
CF Total + DORF	.65	.42	.00**
CF WWR + DORF	.64	.41	.00**
MWF Total + DORF	.62	.39	.00**
MWF WWR + DORF	.54	.31	.002*
NWF CLS + DORF	.59	.36	.00**
NWF WWR + DORF	.60	.37	.00**

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read; ***p* < .001; **p* < .05.

Table 30. *Regressions for predicting TOWRE Total Word Reading Efficiency from DIBELS NEXT 3rd Grade Spring ORF (DORF) and Fluency Measures for Third Graders Below Benchmark (N = 72)*

Variable	R^2	ΔR^2	*p*
Step 1 DORF	.34	--	.02*
Step 2			
RWF Total + DORF	.68	.33	.00**
RWF WWR + DORF	.72	.38	.00**
CF Total + DORF	.57	.23	.02*
CF WWR + DORF	.62	.28	.01*
MWF Total + DORF	.45	.11	.14
MWF WWR + DORF	.52	.18	.05*
NWF CLS + DORF	.72	.38	.00**
NWF WWR + DORF	.64	.30	.01*

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read; $^{**}p < .001$; $^{*}p < .05$.

Factor Analysis

A Principal Component Analysis (PCA) was used because the primary purpose was to identify the factors underlying the phonics fluency probes, DIBELS NWF, and the TOWRE as variables. For indicators of the factorability of R, a Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were used. The second grade sample had a KMO of .85 and the third grade had a KMO of .81. Both are above the .6 value required for a good factor analysis. Bartlett's test of sphericity is a hypothesis test with a null hypothesis stating that correlation coefficients are not zero. The second graders' sample had a significant Bartlett test, χ^2 (55, *N* = 90) = 2230.17, *p*=.000. The third graders' sample also had a significant Bartlett test, χ^2 (55, *N* = 72) = 1673.92, *p*=.000. Results from the KMO and Bartlett tests indicate the current data set was appropriate for factor analysis.

Based on the eigen values only one factor was extracted from the data set for both second and third graders. For second graders this component explained 84.8% of the variance. The one component explained 79.9% of the variance for third graders. Variable loadings on the factor are presented in Tables 31 and 32. Overall, the component factor for this sample suggests that all of the measures are highly related and are measuring the same skills. This suggests that one factor can account for the variance better than two or more factors.

Table 31. *Second Grade Principal Components Analysis Rotated Factor Loadings (N = 90)*

	Factor 1
RWF Total	.95
RWF WWR	.95
CF Total	.95
CF WWR	.95
MWF Total	.93
MWF WWR	.89
DIBELS NWF CLS	.90
DIBELS NWF WWR	.90
TOWRE Phonemic Decoding Efficiency	.94
TOWRE Sight Word Efficiency	.84
TOWRE Total Word Reading Efficiency	.93

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read.

Table 32. *Third Grade Principal Components Analysis Rotated Factor Loadings (N = 72)*

	Factor 1
RWF Total	.92
RWF WWR	.94
CF Total	.93
CF WWR	.92
MWF Total	.89
MWF WWR	.86
DIBELS NWF CLS	.87
DIBELS NWF WWR	.82
TOWRE Phonemic Decoding Efficiency	.88
TOWRE Sight Word Efficiency	.85
TOWRE Total Word Reading Efficiency	.94

Note. RWF = Real Word Fluency probe; CF = Combined Fluency probe; MWF = Multisyllabic Word Fluency probe; Total = Total Sounds; WWR = Whole Words Read.

CHAPTER IV

DISCUSSION

The National Reading Panel (2000) identified phonics as an important reading skill and there is no current progress monitoring specifically targeted to measure advanced phonics skills. The purpose of this study was to evaluate the reliability and validity of three fluency measures assessing complex phonics skills that are typically taught before fourth grade, and students' ability to read words with multiple syllables. This is the first step in reaching the goal of finding a measure that may be more sensitive for monitoring the students' growth in reading than oral reading fluency alone.

The current study sought to answer several questions. First, is the reliability of the fluency probes adequate? When looking at the alternate form test-retest reliability of the different fluency probes, a two-week test-retest interval was used. Overall, the reliability for scoring total sounds was slightly higher than scoring for whole words read for both second and third grade. This indicates that scoring for total sounds may be a more stable measure of students' skills over time. For third graders, the Multisyllabic Word Fluency probe's whole words read was below the .80 cut-off for being considered a reliable measure. Therefore, this would not be a reliable measure for third graders. All of the other probes were stable over the two week period indicating that even though these measures are brief they still are reliable measures of students' skills.

The interrater agreements for all of the fluency probes were very high. This is especially promising given that the examiners only received one short training session on how to administer and score the probes. Since these measures would not require a lot of time or training in order to use, they would be easier to administer to several students than a standardized reading measure and still provide reliable information. Again, scoring for total sounds had higher reliability than scoring for whole words read. Furthermore, the Real Word Fluency and Combined Fluency probes' reliability scores were very comparable. Similar to previous research, findings in this study indicate that the uses of a phonics or word list measure are all reliable fluency measures.

Next the study sought to determine if the fluency probes correlate with other measures of phonics skills. Overall, all probes correlated significantly with DIBELS NWF and the TOWRE Phonemic Decoding Efficiency subtest. The Combined Fluency probe had the highest correlations with the Phonemic Decoding Efficiency subtest. However, when comparing the probes to DIBELS NWF results were mixed between the Real Word Fluency and Combined Fluency probe for having the highest correlations. Neither probe showed an advantage over the other when compared to DIBELS NWF.

Not only did the study try to determine if the fluency probes correlated with other phonics skills measure, but also whether they correlated with other reading skills and overall measures of reading ability. All of the probes correlated significantly with DIBELS ORF and the Test of Word Reading Efficiency. The Multisyllabic Word Fluency whole words read had the highest concurrent validity with spring DORF scores for second graders. For third graders, the Real Word Fluency probe had higher concurrent validity correlations with spring DORF scores by a slim margin.

Not only did the fluency probes correlate significantly with ORF, but students' DIBELS NWF scores also correlated significantly with spring ORF scores. These correlations were approximately .60. In the *DIBELS NEXT Technical* Manual, Good et al. (2011) reviewed the technical adequacy research that has been conducted on DIBELS 6th Edition. This research showed that predictive validity between NWF and end of first grade ORF scores had correlations in the .70s. Correlations for second grade ORF were in the low .60s, which was similar to the correlations of NWF and DORF found in the present study.

The Real Word Fluency probe had the highest correlations with the TOWRE Sight Word Efficiency subtest for both second and third graders. Concurrent validity with the TOWRE Total Word Reading Efficiency showed that the Multisyllabic Word Fluency total sounds had the highest correlation for second graders. For third graders the Real Word Fluency and Combined Fluency probes had the highest correlations with the TOWRE Total Word Reading Efficiency scores. Furthermore, these correlations were equivalent.

The concurrent validity of the fluency probes was also examined for only the students below the DIBELS ORF benchmark. In general, for students below benchmark, the fluency probes correlated higher with phonics measures then the other reading measures. For second graders, the fluency probes continued to correlate significantly with standardized reading measures, with almost all correlations significant at the .001 level. For third graders, fluency probes maintained high correlations with DIBELS NWF, the TOWRE Phonemic Decoding Efficiency subtest, and the TOWRE total score. The fluency probes for third graders below benchmark were correlated lower with DIBELS ORF and the TOWRE Sight Word Efficiency subtest and most correlations were only significant at the .05 level. In addition, the Multisyllabic Word Fluency probe had the lowest correlations with the TOWRE indicating its concurrent validity is not as strong as that of the Real Word Fluency and Combined Fluency probes.

The fluency probes correlated significantly with other reading measures. Because of this, they were also able to explain a significant amount of variance in students' spring DIBELS ORF scores, which was another question the study sought to answer. This was true for both second and third graders. In addition, not only were the fluency probes able to explain a significant amount of variance for all students, but when looking only at students below the DIBELS ORF benchmark, who would use the probes for progress monitoring, the fluency probes were still able to explain a significant amount a variance in their scores. For both second and third graders, the Real Word Fluency probes explained the most variance since it had the highest correlations with other reading measures.

Another question the study sought to answer was whether the fluency probes were more of a measure of phonics or other reading skills. The t-test comparing the probes' correlations with the TOWRE subtests as well as the factor analysis were conducted to analyze this question. Overall, the t-tests were not significant and the factor analysis only found one factor. This suggests all of the reading measures given, DIBELS NWF, DIBELS ORF, the TOWRE, and the fluency probes, are all measuring the same skill (reading). This suggests that the scores of phonics measures, including the fluency probes, were not significantly different than that of other reading measures.

The fluency probes correlated significantly with other reading measures. Because of this, they were also able to explain a significant amount of variance in students' spring DIBELS ORF scores, which was another question the study sought to answer. This was true for both second and third graders. In addition, not only were the fluency probes able to explain a significant amount of variance for all students, but when looking only at students below the DIBELS ORF benchmark, who would use the probes for progress monitoring, the fluency probes were still able to explain a significant amount a variance in their scores. For both second and third graders, the Real Word Fluency probes explained the most variance.

Furthermore the study sought to find whether the fluency probes could provide additional information about students' overall reading scores above DIBELS ORF. Regressions predicting the TOWRE Total Word Reading Efficiency score indicated that DIBELS ORF explained a significant amount of variance in students' scores. However, the fluency probes were still able to explain significantly more variance above what DIBELS ORF explained alone for all students as well as only students below DIBELS ORF benchmark indicating they do provide additional information. This question is especially important when considering whether a measure of advanced phonics skills may be needed. The fact that the fluency probes explain a significant amount of variance in reading scores above what DIBELS ORF can explain alone suggests that having an advanced phonics measure can give us additional information in understanding students' reading abilities. It can give us a better understanding of how fluently students are at using advanced phonics skills to read individual words in isolation. Because of this, having an advanced phonics measure is needed to help explain the variances in students' reading abilities.

When Fien et al. (2010) studied the predictive validity of fall first grade NWF with end of the year ORF scores they found that the predictive validity was lower for the highest performing students. They noted it is unknown if further growth on more advanced phonics elements such as multisyllabic words or more complex phoneme-grapheme relations would be important in predicting overall reading proficiency for students above the NWF threshold. The present study suggests that in terms of reliability and validity, it may be more beneficial to measure the growth of advanced phonics skills than multisyllabic word reading skills. Overall, results from the current study indicated that the Multisyllabic Word Fluency probe had the fewest advantages. First of all, the Multisyllabic Word Fluency probe had the lowest reliability data of the three probes. Second, because reading longer words with multiple syllables is a more difficult skill than reading words focused specifically on testing more advanced phonics skills, a smaller sample size of students' reading was obtained giving the measure. Meaning students read fewer words on the Multisyllabic Word Fluency probe than the other probes. Because of this, there is less variability in students' scores. This would make it more difficult to distinguish between students who are reading on grade level and those who are struggling. Furthermore, because of the smaller behavioral sample, more variance was seen in the interrater reliability for the Multisyllabic Word Fluency probe than the other probes. Third, the Multisyllabic Word Fluency probes' had the lowest correlations with standardized reading measures for third graders. However, for second graders, the Multisyllabic Word Fluency probe's correlations with standardized reading measures were slightly higher or equivalent to those of the other fluency probes. Lastly, when the Multisyllabic Word Fluency probe was entered into linear or multiple regressions, it often explained the least amount of variance in students' standardized test scores. These findings indicate that the Multisyllabic Word Fluency probe may be the least beneficial as a progress monitoring measure and that more than just the ability to read longer words with multiple syllables needs to be measured. Alternatively, perhaps different or less complex multisyllabic words would have better reliability and validity. Further research would be needed in this area.

Researchers have recognized the limitations of DIBELS NEXT and the need to research the usefulness of measures requiring students to read a variety of more complex phonetic patterns (Fuchs, Fuchs & Compton, 2004; Fein et al., 2010). In this study as well as Doty's (2008) thesis, the advanced phonics probes had adequate reliability and concurrent validity. In the current study, when comparing the different measures in terms of their reliability and validity, it appears that the Real Word Fluency and Combined Fluency probes were the best. However, neither the Real Word Fluency nor Combined Fluency probes were clearly more advantageous to use than the other. For the most part, the Real Word Fluency and Combined

Fluency probes performed comparably on most measures, with very little difference between them. Both of the measures had good test-retest and interrater reliability. In addition, both correlated highly with other measures of reading and were able to explain a significant amount of variance in students' scores on other reading measures. However, the Combined Fluency probes often had slightly better validity with subtests that were strictly phonics measures such as DIBELS NWF and the TOWRE Phonemic Decoding Efficiency subtest. On the other hand, the Real Word Fluency probe had slightly better validity with overall reading measures. One advantage the Real Word Fluency probe does have is that the students read more words on the probe than the other fluency probes, which may give a person additional information to analyze in determining what specific skills students are struggling with. Even though both measures did very well, future research would need to be conducted to determine if one of these measures is more sensitive to monitoring students' progress than the other.

Similar to Doty's (2008) thesis, neither the Real Word Fluency nor Combined Fluency probes appear to have a greater advantage over the other. From these studies, it appears that fluency measures of advanced phonics skills, rather than measuring skills such as high frequency word or multisyllabic words, have better concurrent validity with overall reading measures. In Doty's (2008) thesis a high frequency word probe did not do as well as the advanced phonics measure. In the present study, the Multisyllabic Word Fluency probe had the least advantages. This suggests that measures of advanced phonics skills are more advantageous. In addition, similar to Doty's thesis, scoring for total sounds was the most advantageous. Even though scoring only tested phonics skills read correctly or scoring for whole words read correctly were adequate in terms of reliability and validity, scoring for total sounds did slightly better. This may be because it gives a larger sample of student's reading ability.

The present study gives additional information beyond what was found in Doty's thesis. For instance, this study has found that the Real Word and Combined Fluency probes are not only reliable and valid for third graders, but also for second graders. In addition, scoring the probes for whole words read was found to be a valid and reliable scoring option. In Doty's thesis, the fluency probes had lower correlations with the Test of Silent Contextual Reading Fluency (TOSCRF) making it difficult to determine whether the probes have good concurrent validity with other measures of reading fluency. Results from the present study had high correlations with the Test of Word Reading Efficiency (TOWRE). This indicates that the fluency probes do have good concurrent validity with other standardized fluency measures. Correlations with the TOSCRF may have been lower because of it being a silent reading measure or that the way reading fluency was measured by drawing lines to separate words is too different than reading words aloud.

Limitations

Limitations of the methodology from the current study should be noted. First of all, the study was conducted at the end of the year. Since these measures were made for students in second and third grade, the study and results should be replicated with students at the beginning of the year. This needs to be done in order to determine, especially for second graders, if the fluency measures are a valid and reliable measure for students just coming into second grade. In addition, this study would need to be replicated with fourth grade students to determine if the fluency measures are valid and reliable measures for students at that grade level.

An additional limitation of the study is the small sample size of the students below DIBELS benchmark. Because of this, the correlations and regressions may be lower because of the restricted range. Additional studies would need to be conducted with a larger sample of students below benchmark to gain a more accurate depiction of the fluency measures' correlations for these students.

Another limitation has to do with the Multisyllabic Word Fluency probe. The Multisyllabic Word Fluency Probe had the lowest reliability of the fluency probes. In addition, the reliability for whole words read for third graders was below the requirement for a screening measure to be considered reliable. In addition, it had the lowest interrater reliability and the most variance in interrater reliability. Because it had the smallest sample of student reading behavior, differences in interrater scoring made more of an impact. Furthermore, the sample size for whole words read was much smaller and more difficult to score. It was much more difficult to determine if students were sounding out words on the Multisyllabic Word Fluency probe or reading the word as a whole word. It took longer for examiners to become reliable at scoring the Multisyllabic Word Fluency whole words read than scoring the other fluency probes during training. It may also be that the Multisyllabic Word Fluency probe was too difficult. The probe contained words that were mostly four or more syllables. Adding more words that had fewer syllables may increase the reliability and validity of the probe. This would also help increase the variability in scores between students who are good readers and those who are struggling to better identify them. In addition, it would allow the measure to be more sensitive to students' growth in reading skills over time. However, whether the Multisyllabic Word Fluency probe could be changed to be a more reliable and valid measure would have to be researched further.

Future Research

The present study is an initial step for identifying a valid and reliable brief measure of advanced phonics skills. Although the results for some of the measures are promising, further research needs to be conducted to determine if these measures are valid and reliable over time. Since the ultimate goal is to develop a progress monitoring measure, studies need to be conducted using the same type of fluency probe over a longer period of time. In order to do this, future research would need to be conducted to develop probes that are consistent in difficulty. Then research would need to be conducted looking at whether the probes are sensitive to measuring changes in students' reading skills. Future studies could examine whether student's scores on the fluency probes increase as they are learning more advanced phonics skills. If they are not sensitive enough to measure the growth in students' reading skills, they would not be a useful progress monitoring tool to measure the effectiveness of reading interventions. In addition, research would need to be conducted to identify norms for determining readers who are or are not at-risk for having difficulties, and at what point the measures is no longer sensitive to students' gains in reading skills.

Conclusion

The current study is an initial step towards identifying a brief measure of advanced phonics skills that is valid and reliable. Because DIBELS Nonsense Word Fluency only measures letter-sound correspondence and does not assess more complex phonics skills, a brief measure of advanced phonics skills may be needed. Results from this study identified two possible measures that may be useful to progress monitor students' reading growth when receiving interventions. Further research would help to identify which measure may be the most sensitive for progress monitoring. Since it is important to have measures that are reliable and valid, this study is an important step towards in determining which seems to be the most promising fluency measure.

APPENDICES

APPENDIX A

PHONICS SKILLS ASSESSED

Final e rule

ace	ade	ame	ate	ave	ice
ide	ike	ile	ine	ipe	ite
ive	oke	ole	ome	one	ove
ube	ude	uke	une	ute	

Initial consonant blend

bl	br	cl	dr	fl	fr
gl	gr	pl	pr	sw	tr
tw					

Final consonant blend

ft	ld	lt	mb	mp	nd
ng	nk	nt	rd	rk	rm
rn	rt				

Prefix

anti	im/in	inter	non	over	pre
re	sub				

Suffix

ence	ful	ize	ment	ness	ion/tion
ous					

Consonant blend

ch	ck	cr	ph	sc	sh
sk	sl	sn	sp	st	th
wh	wr				

Three letter consonant blend

scr	spl	spr	str	tch

R controlled vowel

ar	er	ir	or	ur

Diagraphs and dipthongs

ai (ā)	ay (ā)	aw/au	ea (ē)*	ea (ĕ)*
ee (ē)	ew (ŏŏ)	ie (ē)*	ie (ī)*	oa (ō)
oi (oi)	oo (ōō)*	oo (ŏŏ)*	ou (ou)	ow (ō)
ow (ou)*	oy (oi)			

VCC

all/ill/ell/oll/ull	and	ant	ark
art	igh/ight	ind*	old

Y as long e

ity	ly

Soft g

age	dge

Other combinations

air	ex	qu

* can be pronounced in two ways

APPENDIX B

SAMPLE REAL WORD FLUENCY PROBE

s/ize	k/ind	p/ai/n/ful	th/orn
j/ea/n	b/ai/t	gr/oa/n	tr/ium/ph
sl/ur/p	c/oa/t	ex/a/lt	dr/ip
obvi/ous	fl/ew	qu/ip	wh/ite
str/ipe	t/art	cr/ew	tr/e/nd
ph/one	non/pr/ofit	th/ence	ma/th
sc/ene	po/nd	hi/nt	to/mb
gl/oa/t	fl/aw	ch/ar	w/ade
inter/mix	ri/ng	h/and	h/ole
over/fl/ow	v/oi/ce	fl/ed	st/em
f/ork	y/arn	m/ine	non/st/op
de/str/oy	gr/ou/p	hi/lt	f/ile
h/ark	r/oy/al	scr/oll	m/ill
p/oke	e/dge	s/igh	m/ute
st/oo/l	n/ou/n	sc/old	ma/lt
p/erk	sm/all	cl/ose	m/ou/se
dec/oy	r/ee/f	qu/ee/n	r/ea/l/ize
no/tch	r/ide	l/one	b/arn
str/ive	st/u/ck	re/st/ful	j/ou/st
l/urk	c/ube	str/ee/t	f/ine
br/ea/d	c/orn	ce/nt	re/cr/uit
m/ai/d	ki/lt	gr/ey	m/oi/st
pr/od	ann/ex	di/sc	kn/oll
dr/ea/m	fl/ow/n	w/arm	bo/nd
final/ize	bo/th	spr/ay	pr/oo/f
Total Sounds: **/59**	**/60**	**/57**	**/60**
WWR **/25**	**/25**	**/25**	**/25**

Total Sounds Correct: _______________ Whole Words Read (WWR): ______________

APPENDIX C

SAMPLE COMBINED FLUENCY PROBE

di/tch	pr/ip	l/ie/d	g/all
va/ck	tr/iv	gr/ell	sw/a/mp
re/ck	r/age	g/ice	j/all
gl/ind	qu/i/ck	pl/ay	j/ate
j/ile	anti/th/e/ft	sw/on	gr/at
fr/i/sion	scr/ub	wu/th	pl/ur
tr/u/st	r/oo/m	spl/a/tch	ch/im
tr/ence	pl/ate	help/ful	r/ile
anti/l/ind	kn/ight	lu/ng	wal/ity
l/old	j/ark	s/old	b/irt
d/ew/n	use/ful	bl/up	w/old
bl/ind	bl/ay	pr/une	qu/ine
m/ai/p	r/ai/l	m/int	la/mb
k/ai/r	ge/sc	sh/ud	br/ute
b/ent	spr/aw/l	be/l/ie/ve	t/ile
spl/at	dr/o/ng	final/ize	ela/tion
b/ai/t	tw/and	t/ar/g	kn/ew
s/ade	t/one	non/v/ank	da/st/ful
t/and	zu/nd	over/age	sk/u/nt
br/ie/f	g/ir/l	j/oi/n	m/ell
im/pr/ess	s/ark	v/ude	non/sc/i/tch
cr/ate	j/oke	sp/ark	c/ame
gr/ey	pum/ity	sc/one	l/ark
dr/aw	d/ame	pl/ea	l/ence
b/ome	om/ity	th/igh	fl/ar

Total Sounds:	**/59**	**/58**	**/57**	**/55**
WWR	**/25**	**/25**	**/25**	**/25**

Total Sounds Correct: _______________ Whole Words Read (WWR): ______________

APPENDIX D

SAMPLE MULTISYLLABIC WORD FLUENCY PROBE

sub/daly	knight/ly	em/pink/ing	re/fest/on/er
pol/li/nate	de/cum/a/tion	drick/le	dis/a/gree
re/cline	re/dunt/ing	shay/ment/ed	es/ti/mate
ov/er/flark	gov/ern/ment	fright/ful	win/den/tion
na/tion/al/ize	re/pour/ian	com/ment/il/ate	dom/sell/er/op
in/vite	tras/mot	yin/ous	an/ti/trust
gro/cer/y	ov/er/size	re/mit/tance	strong/hold
re/sind/ment	ma/dge/ment	pay/roll	in/ter/jift
in/band/in/ate	pre/vail	ov/er/flow	pre/man/a/ble
un/speak/a/ble	par/take	hur/went/ed	sub/tright
quick/sand	ov/er/teep	pre/lum/a/tion	de/bate
ed/u/ca/tion	tay/lence	sub/nelt	tel/men/al/ness
an/vers/i/ble	in/ex/pen/sive	a/dor/a/ble	hil/ab/lin
twup/ment	wh/ame/ful	le/gal/ize	im/ple/ment
re/parn	an/ti/cent/al	un/sin/mon/ary	prep/a/ra/tion
sub/group	sat/is/fac/tion	trair/dune	joy/ful
un/dull/a/ble	mar/ket/place	re/vers/i/ble	com/par/i/son
new/sloke	pre/sart	pro/duc/tive/ness	pro/duce
ex/pan/sive	dim/il/i/ness	an/ti/splart	non/fant
un/der/stand	re/phike	air/sick/ness	non/ten/cert/ed
ber/la/tion/al	mel/ful	re/con/sid/er	con/ver/sa/tion
sim/i/lar/ly	ex/plan/a/tion	chal/leng/ing	fresh/ness
com/ban/ment	re/spade	im/plant	im/print
fork/lift	in/ter/tork	knip/blar	ap/pear/ance
po/lit/i/cal	in/gleep/end	mol/ec/tion/ess	pop/u/la/tion

Syllables:	**/75**	**/72**	**/75**	**/76**
WWR:	**/25**	**/25**	**/25**	**/25**

Total Sounds Correct: Whole Words Read (WWR):

APPENDIX E

PARENT CONSENT FORM

CMU
CENTRAL MICHIGAN UNIVERSITY

Parent/Guardian Consent Form

Title of Project: Advanced Phonics and Multisyllabic Word Fluency Measures: A Reliability and Validity Study

Name of Investigator: Sara Doty **Phone:** (989) 600-8242

Institutional Affiliation: Graduate Student, School Psychology Program, Central Michigan University

Advisor: Michael Hixson **Phone:** (989) 774-3001 **Email:** hixso1md@cmich.edu

Invitation to Participate:
Your child/ward is invited to participate in this research study. The following information is provided to help you make an informed decision whether or not to let your child participate. If you have any questions, please do not hesitate to ask.

What is the purpose of this study?
The purpose of the study is to determine what type of word reading measures are the best to use to monitor the growth of students reading skills when they are receiving reading interventions. This research is being conducted in fulfillment of requirements for a degree from Central Michigan University.

What will my child do in this study?
If you decide to let your child participate in this research project we will ask your child for their assent to participate in the study. If your child chooses to participate, they will participate in the study over two sessions. During the first session your child will be asked to do several things. First, they will be asked to read as many real and nonsense (made-up) words as they can in three one-minute sections. Then your child will complete two other standardized reading measures. In the second session, your child will again be asked to read as many real and nonsense words as they can in three one-minute sections. Some of the sessions will be audiotape recorded in order to ensure the measures are being scored properly.

Initials: ________

How long will it take my child to do this?
This research study will take place in two sessions. The first session will last approximately 20 minutes. A second session will last approximately 10 minutes and will take place two weeks after the first session.

Are there any risks of participating in the study?
There are minimal risks to your child's well-being in participating in this study. Potential discomforts are those normally associated with academic tasks, such as frustration or fatigue. Your child will miss approximately 30 total minutes of classroom time, which may result in missing important classroom instruction. To minimize this possibility, the researcher will schedule the sessions during periods the teacher believes will not adversely affect your child's education.

What are the benefits of participating in the study?
After the study is completed, your child's fluency results will be shared with you. This will give you some brief information about your child's reading skills.

Will anyone know what my child does or says in this study (Confidentiality)?
Any information obtained during this study which could identify your child will be kept strictly confidential. Teachers and other school personnel will not have access to your child's information. All child identifying materials and audiotapes will be stored in a locked cabinet in the faculty advisors' research lab. After the study, all identifying information will be shredded and audiotapes will be erased and destroyed. The information may be published in scientific journals or presented at scientific meetings but your child's identity will be kept strictly confidential.

Will my child receive any compensation for participation?
Your child will not be receiving compensation for participating in the study.

Is there a different way for my child to receive this compensation or the benefits of this study?
There is no alternative compensation that is offered for not participating in the study.

Who can I contact for information about this study?
If you have any questions Sara Doty will be happy to answer them at: 3612 Cree Drive, West Valley City, UT 84120; (989) 600-8242; or Michael Hixson (989) 774-3001.

You are free to refuse to allow your child to participate in this research project or to withdraw your consent and discontinue your child's participation in the project at any time without penalty or loss of benefits to which you are otherwise entitled. Your participation will not affect your child's or your relationship with the institution(s) involved in this research project.

Initials:_______

If you are not satisfied with the manner in which this study is being conducted, you may report (anonymously if you so choose) any complaints to the Institutional Review Board by calling 989-774-6777, or addressing a letter to the Institutional Review Board, 251 Foust Hall Central Michigan University, Mt. Pleasant, MI 48859.

My signature below indicates that all my questions have been answered. I agree to allow my child participate in the project as described above.

______________________________ __________________

Signature of Parent/Guardian Date Signed

____

Name of Child

A copy of this form has been given to me. _________ Parent/Guardian Initials

For the Research Investigator—I have discussed with this subject the procedure(s) described above and the risks involved; I believe he/she understands the contents of the consent document and is competent to give legally effective and informed consent.

______________________________ __________________

Signature of Responsible Investigator Date Signed

APPENDIX F

CHILD ASSENT FORM

CENTRAL MICHIGAN UNIVERSITY

Child Assent Form

Study Title: Advanced Phonics and Multisyllabic Word Fluency Measures: A Reliability and Validity Study

My Name: Sara Doty **Phone:** (989) 600-8242

Who am I? Graduate Student, School Psychology Program, Central Michigan University

Advisor: Michael Hixson **Phone:** (989) 774-3001 **Email:** hixso1md@cmich.edu

What is this research about?
We would like you to join in a research study about reading. You can ask a question at any time and you can say no anytime you want to. Your parents or legal guardian said that it is OK for you to be in this study, but we want to let you choose if you want to do this.

What will happen to me in this research?
We are doing this study to try a new way of testing student's reading skills. If you decide to do this, then you will get to do a few reading activities. You will get to read some real words, made-up words, and some stories. Some of your reading may be audiotaped.

How long will it take me to be in your research?
We will meet with you two times. The first time will last about 20 minutes. The second time will last about 10 minutes and will take place two weeks after the first time.

Can anything bad happen to me?
In this study, you may get frustrated if the reading is difficult or may feel tired from the reading activities. You will also miss about 30 minutes of class time, so you may have to make up the work you miss.

Can anything good happen to me?
You will be helping us to find a short reading test of letter sounds to use with students in the future.

Initials: ________

Do I have other choices?
You may choose not to do this.

Will anyone know I am in the research?
Only those people conducting the study will know that you are in it.

Will I be paid?
No

Who can I talk to about the research?
Sara Doty Phone: (989) 600-8242 Email: doty1sj@cmich.edu
Michael Hixson Phone: (989) 774-3001 Email: hixso1md@cmich.edu

What if I do not want to do this?
You do not have to be in this research study. You can say no at any time. No one will be upset with you if you stop.

SIGNATURE CLAUSE

Do you want to be in the study?

☐ *Yes, I want to be in the study* ☐ *No, I do not want to be in the study*

____
Name of Child (Print)

____________________________________ __________________
Signature of Child Date

____________________________________ __________________
Signature of Person Explaining Assent Date

A copy of this form has been given to me ________ Subject's Initials

APPENDIX G

EXAMPLE PROBE INSTRUCTIONS AND PRACTICE ITEMS

Combined Fluency Probe Directions

I want you to read some make-believe and real words. This word is 'chold' (pointing to the first practice word). **It is a make-believe word. I can read the word 'chold', or I can sound out the parts of the word, /ch/ /old/. Now I want you to read a make-believe word. Try your best to read this word or sound out the parts of the word** (point to the second practice item)."

Correct Response: Child responds "braish" with some or all of the of the parts, say	Incorrect or No Response: If the child does not respond within 3 seconds or responds incorrectly, say
Good job! The sound of the word parts are /br/ /ai/ /sh/ or "braish"	**"Listen, I can sound out the parts of the word /br/ /ai/ /sh/** (point to each letter combination) **or read the whole word "braish"** (run your finger fast under the word**). Try again to read this word** (pointing to the word "braish").

Here is a real word (point to the third practice word). **Try your best to read this word**.

Correct Response: Child responds "inflate" with some or all of the of the parts, say	Incorrect or No Response: If the child does not respond within 3 seconds or responds incorrectly, say
Good job! The sound of the word parts are /in/ /fl/ /ate/ or "inflate"	**"Listen, I can sound out the parts of the word /in/ /fl/ /ate/** (point to each letter combination) **or read the whole word "inflate"** (run your finger fast under the word**). Try again to read this word** (pointing to the word "braish").

Here are some more make-believe and real words. When I say 'begin' start here (pointing to the first word) **and read down the page. Try to read the words the best you can or you can sound out the parts of the word. Be sure to point to the word that you are reading. Put your finger on the first word. Ready, begin."**

chold braish inflate

APPENDIX H

STUDENT ASSESSMENT TRACKING SHEET

Student ID Number: ___________________

Test	**Date Given**	**Score**
Real Word Fluency Probe # 1		Total Score: ________ WWR: _______
Real Word Fluency Probe # 2		Total Score: ________ WWR: _______
Combined Fluency Probe # 1		Total Score: ________ WWR:_______
Combined Fluency Probe # 2		Total Score: ________ WWR: _______
Multisyllabic Word Fluency Probe #1		Total Score: ________ WWR:_______
Multisyllabic Word Fluency Probe #2		Total Score: ________ WWR:_______
TOWRE Word Reading Efficiency Subtest		Raw Score: _______ Standard Score: ____
TOWRE Phonetic Decoding Efficiency Subtest		Raw Score: _______ Standard Score: _____
TOWRE Total Word Reading Efficiency		Total Raw Score: ______ Standard Score: _____
DIBELS: Oral Reading Fluency (DORF)		Median Words Read Correctly: ______
DIBELS: Nonsense Word Fluency (NWF) (2nd and 3rd graders)		CLS: __________ WWR: ________

REFERENCES

Aaron, P. G., Joshi, R. M., Ayotollah, M., Ellsberry, A., Henderson, J., & Lindsey, K. (1999). Decoding and sight-word naming: Are they independent components of word recognition skill? *Reading and Writing: An Interdisciplinary Journal, 11*, 89-127.

Assessment Committee. (2002). *Analysis of reading assessment measures: Stanford Achievement Test, 9th Edition*. Retrieved April 5, 2005, from http://idea.uoregon.edu/assessment/analysis_results.

Beaver, J. (1997). *Developmental Reading Assessment, K-3*. Glenview, IL: Celebration Press.

Brigance, A., & Glascoe, F. P. (1999). *Brigance Diagnostic Comprehensive Inventory of Basic Skills-Revised*. North Billerica, MA: Curriculum Associates.

California Department of Education. (2002). *Technical report for the California English Language Development Test (CELDT)*. Retrieved March 20, 2005, from www.cde.ca.gov.

California Department of Education. (2004). No Child Left Behind, 2001. Retrieved March 20, 2005, from www.cde.ca.gov.

Carlson, C. D., Chen, D., Mouzaki, A., et al. (1998). *Texas Primary* R*eading Inventory*. Houston: University of Texas-Houston Health Sciences Center, Center for Academic and Reading Skills, and University of Houston.

Carnine, D. W., Silbert, J., Kame'enui, E. J., & Tarver, S. (2004). *Direct instruction reading-fourth edition*. Upper Saddle River, NJ: Prentice Hall.

Cartwright, K. B. (2002). Cognitive development and reading: The relation of reading-specific multiple classification skill to reading comprehension in elementary school children. *Journal of Educational Psychology, 94*, 56-63.

CBT/McGraw-Hill. (1997). *TerraNova-Second Edition*. Monterey, CA: Author.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*, 98(2), 294-409.

Compton, D.L. (2000). Modeling the growth of decoding skills in first grade children. *Scientific Studies of Reading*, 4(3), 219-259.

Deno, S. L., Mirkin, P. K., & Chaing, B. (1982). Identifying valid measures of reading. *Exceptional Children*, 49, 36-45.

DiPerna, J. C., & Elliott, S. N. (2000). *Academic Competence Evaluation Scales*. San Antonio,TX: The Psychological Corporation.

Doty, S. J. (2008). *Evaluating the reliability and validity of three advanced phonics skills measures and a high frequency word measure*. (Unpublished thesis). Central Michigan University, Mt. Pleasant, MI.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test-Third Edition*. Circle Pines, MN: American Guidance Service.

Ehri, L.C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2), 167-188.

Fantuzzo, J.W., & Hampton, V. R. (2000). Penn interactive peer play scale: A parent and teacher rating system for young children. In K. Gitlin-Weiner, A. Sandgrund, & C. Schaefer (Eds.), *Play diagnosis and assessment* (2^{nd} ed., pp. 599-620). NewYork: Wiley.

Fien, H., Parker, Y., Baker, S. K., Smith, J. L., Stoolmiller, M., & Kane'enui, E. J. (2010). An examination of the relation of nonsense word fluency initial status and gains to reading outcomes for beginning readers. *School Psychology Review,* 39 (4), 631-653.

Fredrickson, N., Frith, U., & Reason, R. (1997). *Phonological Assessment Battery*. Windsor, UK: NFER Nelson Publishing Company.

Fry, E. (2004). Phonics: A large phoneme-grapheme frequency count revised. *Journal of Literacy Research*, 36(1), 85-98.

Fry, E.B., & Kress, J. E. (2006). *The reading teacher's book of lists* (5^{th}. ed). San Francisco, CA: Jossey-Bass.

Fuchs, L. S., Fuchs, D., & Comptom, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children*, 71(1), 7-21.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Monitoring reading growth using student recalls: Effects of two teacher feedback systems. *Journal of Educational Research, 83*, 103-111.

Ginsburg, H. P., & Baroody, A.J. (1990). *Test of Early Mathematics Ability-Second Edition.* Austin, TX: PRO-ED.

Good, R. H., Baker, S. K., & Peyton, J. A. (2008). Making sense of nonsense word fluency: Determining adequate progress in early first-grade reading. *Reading and Writing Quarterly*, 25(1), 33-56.

Good, R. H., Gruba, J., & Kaminski, R. A. (2002). Best practices in using Dynamic indicators of Basic Early Literacy Skills (DIBELS) in an outcome driven model. In Thomas, A., & Grimes, J., *Best Practices in School Psychology IV*. Bethesda, MD: NASP Publications.

Good, R. H., & Kaminski, R. A. (2002). Nonsense Word Fluency, In R. H. Good & R. A. Kaminski (Eds.) *Dynamic Indicators of Basic Early Literacy Skills* (6^{th} ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K.A., & Latimer, R. A. (2011). *DIBELS Next Technical Manual Draft*. Eugene, OR.

Good, R. H., Kaminski, R. A., Smith, S., Simmons, D., Kame'enui, E., & Wallin, J. (2003). Reviewing outcomes: Using DIBELS to evaluate a school's core curriculum and system of additional intervention in kindergarten. In S.R. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom, Systems for observing teaching and learning*. Baltimore: Paul H. Brookes.

Graves, A. W., Plasencia-Peinado, J., Deno, S. L., & Johnson, J. R. (2005). Formatively evaluating the reading progress of first-grade English learners in multiple-language classrooms. *Remedial and Special Education*, 26(4), 215-225.

Hammill, D. D., Wiederhold, J. L., & Allen, E. A. (2006) *Test of Silent Contextual Reading Fluency*. Austin, TX: PRO-ED.

Hanna, P. R., Hanna, J. S., Hodges, R. E., & Rudorf, E. H. (1966). *Phoneme-grapheme correspondences as cues to spelling improvement.* Washington, DC: U.S. Department of Health, Education, and Welfare.

Harcourt Brace Educational Measurement. (1996). *Stanford Achievement Test Series-Ninth Edition*. San Antonio, TX: PRO-ED.

Harn, B. A., Stoolmiller, M., & Chard, D. J. (2008). Measuring the dimensions of alphabetic principle on the reading development of first graders the role of automaticity and unitization. *Journal of Learning Disabilities*, 41(2), 143-157.

Jordan, N. (1994). Developmental perspectives on reading disabilities. *Reading & Writing Research Quarterly*, *10*, 297-311.

Karlsen, B., Madden, R., & Gardner, E. F. (1975). *Stanford Diagnostic Reading Test*. New York: Harcourt Brace Jovanovich.

Lindamood, C., & Lindamood, P. (1979). *The LAC Test: Lindamood Auditory Conceptualization Test- Revised.* Austin, TX: PRO-ED.

Mather, N., Hammill, D. D., Allen, E. A., & Roberts, R. (2004). *Test of Silent Word Reading Fluency*. Austin, TX: PRO-ED.

McCoach, B. D., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing Readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, *98*, 14-28.

National Reading Panel. (2000). *Report of the National Reading Panel; Report of the subgroups*. Washington, DC: National Institute of Child Health and Human Development Clearinghouse.

Perry, W. (1970). *Forms of intellectual and ethical development in the college years: A scheme.* New York: Holt, Rinehart & Winston.

Psychological Corporation. (2002). *Wechsler Individual Achievement Test-Second Edition*. San Antonia, TX: The Psychological Corporation.

Reid, D. K., Hresko, W. P., & Hammill, D. D. (2001). *Test of Early Reading Ability-Third Edition*, Austin, TX: PRO-ED.

Rouse, H. L., & Fantuzzo, J. W. (2006). Validity of the dynamic indicators for basic early literacy skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review*, 35(3), 341-355.

Salvia, J., Ysseldyke, J. & Bolt, S. (2012). *Assessment: In special and inclusive education-twelfth edition*. Boston, MA: Houghton Mifflin Company.

Samuels, S. J. (1981). Some essentials of decoding. *Exceptional Education Quarterly*, 1, 11-25.

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, 21(2), 459-47.

Speece, D. L., Mills, C., Ritchey, K. D., & Hillman E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *The Journal of Special Education*, 36(4), 223-233.

Thomson, B., Crewther, D. P., & Crewther, S. G. (2006). Wots that werd? Pseudowords (non-words) may be a misleading measure of phonological skills in young readers. *Dyslexia*, 12, 289-299.

Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement* (Research Rep. 109). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.

Torgesen, J. K., Wagner, R., & Rashotte, C.A. (1997). *Test of Word Reading Efficiency*. Austin, TX: PRO-ED.

Vanderwood, V. L., Linklater, D., & Healy, K. (2008). Predictive accuracy of nonsense word fluency for English language learners. *School Psychology Review*, 37(1), 5-17.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: PRO-ED.

Wiederholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Tests-Fourth Edition*. Austin, TX: PRO-ED.

Wilkinson, G. S. (1993). *Wide Range Achievement Test-Third Edition*. Wilmington, DE: Wide Range, Inc.

Woodcock, R. W. (1987). *Woodcock Reading Mastery Test*. Circle Pines, MN: American Guidance Service.

Woodcock, R.W. (1991). *Woodcock Language Proficiency Battery-Revised*. USA: Riverside Publishing Co.

Woodcock, R. W. (1998). *Woodcock Reading Mastery Test-Revised/Normative Update.* Circle Pines, MN: American Guidance Service.

Woodcock, R. W. (1973). *Woodcock Reading Mastery Test.* Circle Pines, MN: American Guidance Service.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psychoeducational Battery-Revised*. Allen, TX: DLM.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.