

Rochester Institute of Technology

RIT Digital Institutional Repository

Theses

2007

Implications of Reading Assessments: Investigating the Developmental Reading Assessment and Dynamic Indicators of Basic Early Literacy Skills - Oral Reading Fluency as Related to the TerraNova 2nd Edition Vocabulary and Comprehension Tasks for Third-Grade Students

Courtney Wheeler

Follow this and additional works at: <https://repository.rit.edu/theses>

Recommended Citation

Wheeler, Courtney, "Implications of Reading Assessments: Investigating the Developmental Reading Assessment and Dynamic Indicators of Basic Early Literacy Skills - Oral Reading Fluency as Related to the TerraNova 2nd Edition Vocabulary and Comprehension Tasks for Third-Grade Students" (2007). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the RIT Libraries. For more information, please contact repository@rit.edu.

Implications of Reading Assessments: Investigating the Developmental Reading
Assessment and Dynamic Indicators of Basic Early Literacy Skills - Oral Reading
Fluency as Related to the TerraNova 2nd Edition Vocabulary and Comprehension Tasks
for Third-Grade Students

Graduate Thesis

Submitted to the Faculty

Of the School Psychology Program

College of Liberal Arts
ROCHESTER INSTITUTE OF TECHNOLOGY

By

Courtney Wheeler

In Partial Fulfillment of the Requirements
for the Degree of
Master of Science and
Advanced Graduate Certificate

Rochester, New York

June 12, 2007

Approved: Suzanne Graney
Suzanne Graney, Ph.D.

Vincent Pandolfi
Vincent Pandolfi, Ph.D.

R I T
School Psychology Program
Permission to Reproduce Thesis

PERMISSION GRANTED

Title of thesis : Implications of Reading Assessments: Investigating the Developmental Reading Assessment and Dynamic Indicators of Basic Early Literacy Skills - Oral Reading Fluency as Related to the Terra Nova 2nd Edition Vocabulary and Comprehension Tasks for Third Grade
I Courtney Wheeler hereby **grant permission** to the Students
Wallace Memorial Library of the Rochester Institute of Technology to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Date: _____ **Signature of Author:** Courtney Wheeler

PERMISSION FROM AUTHOR REQUIRED

Title of thesis _____

I _____ **prefer to be contacted** each time a request for reproduction is made. I can be reached at the following address:

PHONE: _____

Date: _____ **Signature of Author:** _____

-

PERMISSION DENIED

TITLE OF THESIS _____

I _____ hereby **deny** permission to the Wallace Memorial Library of the Rochester Institute of Technology to reproduce my thesis in whole or in part.

Date: _____ **Signature of Author:** _____

Running head: IMPLICATIONS OF READING ASSESSMENTS

Implications of Reading Assessments: Investigating the Developmental Reading
Assessment and Dynamic Indicators of Basic Early Literacy Skills - Oral Reading
Fluency as Related to the TerraNova 2nd Edition Vocabulary and Comprehension Tasks
for Third-Grade Students

Courtney E. Wheeler

Rochester Institute of Technology

Abstract

The present study investigated the relationship between Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency and the Developmental Reading Assessment (DRA) as related to the TerraNova 2nd Edition Vocabulary and Comprehension. Historical test data gained from 46 third grade participants during 2004-2005 school year from an urban elementary school was utilized. The DRA protocols were coded and correlations were run to determine the relationship between the three tests. Significant correlations were found consistently between the DIBELS scores with DRA Phrasing and Fluency and DRA Story Level. Of particular interest, the overall best predictor for DIBELS scores was the DRA Story Level. Significant correlations were also found between DIBELS scores, DRA scores, and TerraNova 2nd Edition Comprehension and Vocabulary. The best predictor for both the TerraNova 2nd Edition Vocabulary and Comprehension scores was the fall DRA Story Level.

Chapter I

Statement of the Problem

Reading is a fundamental skill, for adults and children alike, to the most basic tasks such as following directions, understanding warning labels, following recipes, and other basic living skills. Yet, if reading skills and strategies are not acquired early in the learning process it can be detrimental and follow that child through the rest of their education and into their adult lives. Children who learn to read are more likely to become productive members of society as compared to those who never learn (Adams, 1990). Those who remain illiterate are more likely to be at increased risk for childhood conduct problems (Bennett, Brown, Boyle, Racine, & Offord, 2003). Adams referred to the statistics on illiterate adults, who make up 85% of juvenile offenders, 60% of prison inmates, and 75% of the unemployed. This is why good instruction, especially in the area of reading, is vital in our current educational system.

In recent times reading instruction in the United States has begun at an earlier age. From pre-school, where students begin to gain pre-literacy skills, until third grade, students are learning to read. They are practicing their literacy skills and much of the educational curriculum focuses on this process. Beginning in fourth grade the curriculum shifts and the focus is no longer on practicing to read. Instead students must utilize what literacy skills they have in order to learn the school related material.

This shift in instruction has the potential to leave behind the students who failed to learn some of the necessary skills to be proficient in reading. Students who acquire their reading skills at a slower rate than other students are considered to be “at risk” for reading failure (Foorman & Torgesen, 2001). It was also found that these students are “at

risk” because of the interaction of the learning environment characteristics with their own. In other words, these students may not be able to learn to read with the method that the curriculum is teaching them. They require different techniques and strategies for acquiring the skills needed to read.

The struggling students who need extra practice with reading have been reported to actively dislike reading and read less both in and out of school (Juel, 1988). In terms of reading this is described by Stanovich (1986) as the “Matthew Effect” phenomenon, “the rich get richer and the poor get poorer.” In other words, the students who acquire literacy skills early in their education have the opportunity to grow with their skills and knowledge base whereas those who fail to develop them continue to fall farther behind (Haaer & Windmueller, 2001). Once those students reach the fourth grade where reading is vital to learning other subject areas they most likely will continue to fall behind in their academic career.

As the students progress through school it becomes increasingly hard to teach the necessary skills to bring them up to grade level expectations. The students who displayed poor reading skills in the first grade had a 88% chance of continuing to have those poor reading skills in fourth grade if no intervention was put into place (Juel, 1988). Those students who had reading problems at the end of third grade were not likely to improve by the end of eighth grade (Felton & Wood, 1992). Early intervention is vital for improvement and accurate assessments must be used in order to detect who is in need of early intervention.

In order to evaluate the level of competency at which the child is currently performing, assessment techniques must be employed. As more assessment techniques

are developed, it is important to investigate what they are testing and their predictive value. Tools used to assess student reading achievement should have certain properties depending on the reason for use (Fuchs and Fuchs, 1999). Assessment tools can be evaluated based on their ability to be used as a screening tool, a progress-monitoring tool, a skill diagnosis tool, or as an outcome measurement tool. When choosing an assessment tool the reasons for testing must be considered among other factors. In some school districts multiple reading assessments are given to students during a one-year period to gauge reading achievement. Having multiple assessments that may measure the same or similar constructs is redundant. This study focuses on three such assessments, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), the Developmental Reading Assessment (DRA), and the TerraNova 2nd Edition Comprehension and Vocabulary.

The DIBELS were derived from Curriculum Based Measurement (CBM), which is a set of procedures that has been empirically validated as a progress monitoring assessment tool. The hallmark of reading CBM is that the measures are reliable and valid, quick and easy to administer, sensitive to small changes in student performance over time, and have multiple forms for repeated measurement (Deno, 1985; Deno, Mirkin, & Chiang, 1982; Martson, 1989). The DIBELS has demonstrated adequate technical characteristics to be used as an indicator of reading achievement.

The DRA is a tool used in school systems to monitor progress and diagnose reading difficulties. It has not had many empirical studies demonstrating the adequacy of its technical properties. The TerraNova 2nd Edition is a group administered assessment instrument used at the end of the school year to measure the academic achievement of

students in several academic domains including reading. This study specifically utilizes the comprehension and vocabulary sections of the exam.

The administration of all three instruments is a lengthy process, resulting in loss of instructional time for students and teachers. The cost effectiveness, both in the time and money involved in training the evaluators and the administration and scoring of the instruments also must be considered when ultimately deciding which to administer. Although all three instruments were administered to the cohort of students from which the data for this study was obtained, the quality of information gained through their combined use is not known.

Purpose of Study

The purpose of this study was to serve as a starting point for extending the current literature on the relationships among DIBELS Oral Reading Fluency and the DRA as related to the TerraNova 2nd Edition Comprehension and Vocabulary. The object of this study is to determine the relationships between the three reading assessments. The predictive validity was also examined for each test with the TerraNova. The following research questions were addressed in this study:

1. What is the relationship between the DIBELS Oral Reading Fluency and the DRA for third grade students?
2. How does the DIBELS Oral Reading Fluency relate the TerraNova Comprehension and Vocabulary for third grade students?
3. How does the DRA relate to the TerraNova Comprehension and Vocabulary for third grade students?

4. Which classroom assessment, the DIBELS or the DRA, is the strongest predictor of student performance on the TerraNova Comprehension and Vocabulary for third grade students?

Chapter II

Literature Review

Research into literacy development and the emergence of reading assessments to measure that development has a long history. The necessity of gaining literacy skills while in a school setting is obvious, however, many students struggle with acquiring these skills. To ensure a student is progressing at an appropriate rate reading assessments must be employed. As the research on literacy and reading assessments has expanded, instruction must be adjusted to use evidence-based practices. This literature review is focused on (a) literacy development, (b) instruction and assessment, and (c) the assessments utilized in this study.

Literacy Development

All students must master certain skills in reading acquisition in order to become proficient in reading. Three skills essential for literacy include phonological awareness, alphabetic principle, and fluency with connected text (Simmons & Kame'enui, 1998). All three of these “big ideas” must be present for the child to comprehend what they are reading and in turn understand what they are learning. Phonological awareness is when a child understands that there are smaller components to language than the whole word, which lends itself to phonemic awareness (Simmons & Kame'enui, 1998). Phonemic awareness is the ability to distinguish and manipulate the sounds in words. For example, the word “dog” would be heard as /d//o//g/. Alphabetic principle involves learning how letters and sounds correspond with each other, how letters can be used to spell, and how both of these skills lend themselves to reading text. The third “big idea” in literacy is fluency with connected text, which is the most advanced of the three, and this is the

automatic ability to read words in connected text (LaBerge & Samuels, 1974). The National Institute of Child Health and Human Development (NICHD, 2000) reiterated that fluency is one of several crucial factors needed for reading comprehension, however it is the skill that is neglected the most in the classroom. It can be reinforced through guided oral reading and independent silent reading.

The other two “big ideas” that have been outlined in the terms of beginning reading are vocabulary and comprehension (NICHD, 2000). Vocabulary refers to the ability to understand and use words to acquire meaning. Comprehension is the interaction between the reader and the text to convey meaning (Harris & Hodges, 1995; NICHD, 2000). When students are engaged in the text and relating their background knowledge to the story, comprehension will be enhanced (NICHD, 2000).

Three basic conceptualizations of the reading process, bottom-up, top-down, and interactive (Kibby, 1995), have been offered to explain how literacy develops over time. The bottom-up theory emphasizes that higher-level processes await the completion of lower level ones (LaBerge & Samuels, 1974). LaBerge and Samuels labeled this the automaticity model of reading and it is the most widely used rationale for measuring oral reading fluency as an indicator general reading competence (Fuchs, Fuchs, Hosp, & Jenkins, 2001). With this model success is dependent on certain components being executed automatically. If not executed automatically, each individual skill required for reading would exceed the attentional capacity of an individual (Fuchs et. al., 2001; LaBerge & Samuels, 1974). The top-down theory emphasizes that reading is merely an extension of the language acquisition process and comprehension is gained from the text by having meaning already and applying it to the current text (Kibby, 1995; Goodman,

1967). For example, the bottom-up theory is print driven whereas the top-down theory is meaning driven (Stanovich, 1980).

The third conception of reading is the interactive model, which is a combination of the bottom-up and top-down models. In this view, reading comprehension results from aspects of the reader, aspects of the text, and the situational context in which the reading occurs (Stanovich, 2000; Kibby, 1995; Harste, Woodward, & Burke, 1984; Lipson, 1983). The interactive process involves prior contextual knowledge that helps to identify words rather than high level processes being dependent on the completion of the lower level processes (Fuchs et al., 2001). All three perspectives of reading hold that efficient low-level word recognition is necessary for the high level comprehension of the text to occur (Fuchs et. al, 2001).

Instruction and Assessment

Every child benefits from effective classroom instruction. With effective instruction the children “at risk” for reading failure are reduced to a small percentage as compared to non-effective reading instruction (Foorman & Toregesen, 2001). Assessment serves important purposes in providing effective instruction to all students. Assessment can also drive intervention programs and give teachers clearer information about individual students’ performances. It gives quantitative data to qualitative observations teachers make in the classroom about the students’ proficiency in reading. Reading assessment serves to identify those in need of supplemental instruction, guiding the instructional planning, and monitoring the students’ progress toward reading more proficiently (Haager & Windmueller, 2001). Assessment that is frequent and ongoing helps teachers provide effective interventions for the students that are struggling with

reading (Deno, 1997; Fuchs, 1989). It also can assist in educational planning for students who are not struggling with reading by identifying the students whose level is instructionally correct and the students that may need to be challenged more through supplemental reading material. Four uses of assessment that are frequently considered in schools are assessment for screening purposes, diagnosis, progress monitoring, and outcome (Salvia & Ysseldyke, 2004). The information needed from an assessment will determine which type will be used.

Regardless of the purpose, reading assessment tools must be reliable and valid if they are to be used to make decisions about children. Reliability is a major consideration in evaluating a test. It ensures that the test is consistent in measuring the construct. Ways in which a test would show reliability include: the test would be able to use similar questions and obtain the same results, the behavior would be the same if testing occurred during a different time of day or on a different day, and that if another qualified examiner were to give the test the results would be the same (Sylvia & Ysseldyke, 2004). Validity is the most fundamental consideration for an assessment test. It is the degree that the test is measuring what it sets out to measure, the degree that the evidence and theory behind a specific test supports the interpretation and use of those test scores (Sylvia & Ysseldyke, 2004). Validity cannot exist without reliability.

Curriculum based measurement (CBM) is the most widely studied form of classroom assessment (Fuchs, Fuchs, & Compton, 2004). This type of assessment allows educators to assess students with an instrument that is reliable, valid, and efficient to administer (Deno, 1985). It also allows educators to view the student both at a certain point in time and the progress that student is making across a period of time. It can also

be used to compare a particular student to a classroom of students. CBM can be distinguished from other classroom assessments in many ways. It is standardized with documented reliability and validity. The focus of CBM is long term so assessment can be done frequently over a long period of time. It also has alternate forms of each test, which is essential for progress monitoring. It serves to create optimal learning conditions by providing corrective feedback for both teachers and students (Gravois & Gickling, 2002). Ultimately, CBM is used to assess the skills as a student enters the classroom and provides insight to interventions around which instruction can be designed. It is also used to evaluate the effectiveness of the instruction over time.

Adams (1990) referred to the speed in which text is spoken out loud as the most important characteristic of reading. This is oral reading fluency, that is the oral reading of text with speed and accuracy (Fuchs et. al., 2001). One of the most widely researched and implemented CBM tasks examines students' rates of oral reading (R-CBM). It involves counting the number of words read correctly from meaningful connected text in a 1 minute time period. Because it looks at words read correctly in a given time frame, it is sensitive enough to detect small changes in reading proficiency (Fuchs & Fuchs, 1986). It can be used to generate time-series sets of data to be displayed graphically and can result in better instruction through higher expectations and more adaptations to instructional programs (Fuchs et. al, 1989). The CBM also is related to the reader's skill at making meaningful connections between sentences, relating text meaning to a reader's existing knowledge, and making inferences to complete the story (Fuchs et. al, 2004).

Oral reading fluency, as a measure of comprehension, was assessed by Fuchs, Fuchs, and Maxwell (1998) along with three other measures, including question

answering, cloze, and passage recall, to determine how each procedure related to reading comprehension. Results suggest that the correlation between ORF and reading comprehension (.91) was at least as strong as the other three measures (question answering .82, recall .70, and cloze .72). The other measurement techniques had additional drawbacks including difficulties creating questions about the passage in which the answer was not directly in the reading material, and maintaining objective scoring criteria for story recall.

Other studies have analyzed oral reading fluency and its link to reading comprehension. Jenkins, Fuchs, Espin, van den Broek, and Deno (2003) examined words read in text and words read in isolation to see which approach predicted comprehension better. Overall, they found that reading words in text has more in common with reading comprehension than reading words in lists. Fuchs et al. (2001) expanded on this finding by suggesting that oral reading fluency goes beyond the word level, tapping different reading subcomponents. Silent reading fluency versus oral reading fluency was assessed in the realm of reading comprehension by Fuchs et al. (2001). The correlation between oral reading fluency scores and the Reading Comprehension portion of the Iowa Test of Basic Skills (ITBS-RC) (Hoover, Dunbar, & Frisbie, 2001) was significantly higher than the correlation between the ITBS-RC and the silent reading fluency scores. Because oral reading fluency is vital to reading comprehension, it is an important skill to measure in making decisions about students' reading instruction.

Dynamic Indicators of Basic Early Literacy Skills

A specific type of CBM for reading is the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) developed by Kaminski and Good (1996). DIBELS is a series of tasks that assess a student's fluency in fundamental reading skills. Its measures start in kindergarten and progress through sixth grade. It takes into account the "big ideas" in early reading, i.e., phonological awareness, alphabetic principle, and fluency with connected text, through several brief measures of phonemic awareness, letter names, word attack, and reading fluency. Initial Sound Fluency (ISF) assesses the student's ability to recognize and produce the initial sound in an orally presented word (Kaminski & Good, 1996, 1998). Phoneme Segmentation Fluency (PSF) assesses the student's ability to segment three and four phoneme words into their individual phonemes fluently and is considered a good predictor of future reading achievement (Kaminski & Good, 1996). The Nonsense Word Fluency (NWF) tests alphabetic principle, which includes letter-sound correspondence and blending letters into nonsense words. In Letter Naming Fluency students are asked to name both upper and lowercase letters to see if a student is at risk for a reading problem.

The Oral Reading Fluency (ORF) assesses the amount of words that students can read correctly in a 1 minute time period. In that one minute the students are asked to read as quickly and correctly as they can. The examiner counts words omitted, substituted, and hesitated on for more than three seconds as errors and at the end of one minute records the total number of words read. The oral reading fluency rate is then calculated by looking at the number of words read correctly in one minute (Kaminski & Good, 1996). This measure is based on the R-CBM procedures previously discussed.

Martson (1989) reviewed the early reliability and validity of ORF in Curriculum-Based Measurement. It was found that the overall test-retest reliability ranged from .82 to .97. Marston also found that alternate-form and inter-rater reliability had correlation coefficients of above .90. This indicates that ORF is a reliable assessment tool. The passages on the DIBELS-ORF were designed to be consistent with the Test of Reading Fluency (TORF) (Children's Educational Services, 1987) in readability, reliability and validity (Kaminski & Good, 1996; Good, Wallin, Simmons, Kame'enui, Kaminski, 2002). The TORF consists of standardized passages and administration procedures that can be used a screening device for children and measure growth in reading skills (Children's Educational Services, 1987). Concurrent validity with the TORF passages ranged from .92 to .96 (Good et al., 2002).

DIBELS Oral Reading Fluency becomes especially relevant when looking at the relationship between ORF and high stakes testing. Since the emergence of No Child Left Behind (NCLB), curriculum has often focused on preparing students for high stakes state-mandated tests. NCLB utilizes high stakes testing to help raise student achievement. The goal of NCLB is to increase reading proficiency among all students. NCLB was created to strengthen accountability in the United States' public schools by requiring states to have challenging standards in reading and mathematics, annual testing for all students in grades 3-8, and progress objectives (NCLB, 2001). From this act, Reading First grants were formed to help eliminate the reading deficit by establishing high-quality and comprehensive reading instruction in the early grades (NCLB, 2001).

Numerous studies have been conducted comparing DORF and the relationship to various state high stakes tests that satisfy NCLB legislation. For example, Wilson (2005)

found that third-grade students who achieved the benchmark on the DORF were highly likely to pass the Arizona Instrument to Measure Standards. Students who were considered to be at risk based on their DORF performance were unlikely to meet the proficiency standard on the Arizona Instrument to Measure Standards. The DORF measure could be used about as equally well for various demographic subgroups based on variables such as gender, ethnicity, and income level. Shaw and Shaw (2002) used DORF for third-grade students to examine the relationship between those measures and performance on the Colorado State Assessment Program (CSAP). The third-graders who read 90 or more words correctly in one minute on the DORF were 91% likely to score “proficient” or “advanced” on the CSAP and those that read less than 90 on the DORF were likely to receive a score of “unsatisfactory” or “partially proficient” (Shaw & Shaw, 2002). Buck and Toregesen (2003) investigated the relationship between the CBM and the Florida Comprehensive Assessment Test (FCAT) and found that there was a significant correlation between CBM scores and reading FCAT scores. Another study found DORF was an accurate predictor of whether a student would achieve a proficient score on the North Carolina End of Grade Reading Assessment (Barger, 2003).

Developmental Reading Assessment

While a multitude of evidence exists for the validity of the DIBELS measure of Oral Reading Fluency and the instructional implications for those who are assessed with CBM, some school districts opt to use different assessments. One such assessment is the Developmental Reading Assessment (DRA). School districts use the DRA in different fashions. Some use it in conjunction with CBM, while others use it exclusively. The DRA was designed to inform instruction by assessing a student’s independent reading

level and diagnosing the strengths and weaknesses in various areas including oral reading fluency (Pearson Learning Group, 2005). Specifically, the DRA was designed to measure how well students read literature, to monitor student growth and development on a variety of skills and strategies, to help teachers pinpoint students' needs, to prepare the student to meet school expectations, and to support school personnel and parents by informing them of the level of achievement (Pearson Learning Group, 2005).

The DRA K-3 was developed using a consensus of what educators thought were good characteristics for reading. It purports to measure three components of reading including engagement, fluency, and comprehension (Pearson Learning Group, 2005). The DRA is based on 12 criteria that readers must meet in order to be considered a "good reader". According to the DRA publishers, a good reader is able to select appropriately leveled texts to meet their reading needs, spend time reading, monitor their reading, and know authors and books.

Each of the stories that the DRA utilizes has been assigned a level based on the story's difficulty. For example, as the difficulty increases so does the level of the story. The DRA K-3 test kit consists of two assessment texts for levels A through 44. This measure takes 10-20 minutes to administer per student and is typically administered twice a year. However, it can be administered more frequently. Instead of starting at a grade level, the story is selected by the student being tested from a group of three to four texts. The level of text, Story Level, is selected by the teacher to attempt to match the student's independent reading level. The student then reads the book orally to the teacher. The student is assessed through the teacher making observations of the student's oral reading and citing the strategies used. It is broken up into different sections, five of which were

utilized for this study. The five include Story Level, Phrasing and Fluency, Intonation, Accuracy, and Comprehension.

First, the examiner and the student choose the Story Level. The Story Level is determined based on the teacher perception of how the student is reading and at what instructional level. The teacher selects three or four texts of varying different difficulty that likely match the student's independent reading level. The student then chooses one story from these options, which will be used for the assessment. Essentially the Story Level is the teacher's judgment on where a child is reading. There is not a standardized set of rules or algorithms for judging which story level should be chosen for the child. Instead, The DRA Technical Manual (2005) stipulates that the Story Level should be adjusted to find the instructional level based on the level of accuracy, comprehension, and fluency.

Phrasing and Fluency allows the examiner to describe the length of phrases the student uses while reading aloud. The examiner determines which of six statements most accurately describes the student's phrasing and fluency. These include "word by word", "in short phrases at times", "in short phrases most of the time", "in longer phrases at times, but with an inconsistent rate", "in longer phrases most of the time with adequate rate", and "in longer phrases with rate adjusted appropriately". This is an indication of the reader's fluency with the selected text.

Intonation is a description of how the student sounds when he or she reads the text. The examiner determines which of six statements most accurately describes the student's intonation. The statements include, "no intonation", "little intonation", "some intonation with some attention to punctuation", "adjusts intonation to convey meaning at

times”, “adjusts intonation to convey meaning attending to punctuation”, and “begins to explore subtle intonation that reflects mood, pace, and tension”. These aspects of reading are important because it indicates a student’s degree of engagement with the text.

Accuracy is the percentage of words read correctly. In order for a student to gather meaning from the text, he or she must read it accurately. Finally, The Comprehension section of the DRA K-3 measures the examiner’s perception of how much the student comprehends after asking them questions regarding the story. The student’s ability to retell the text in a way that indicates understanding of the main ideas, key facts, and characters is examined. Comprehension is rated with a rubric that is divided into the following categories: “Very Little Comprehension”, “Some Comprehension”, “Adequate Comprehension”, and “Very Good Comprehension”.

Williams conducted an initial large-scale field-testing of the DRA K-3 in 1996. It involved 346 students, 10 of whom were in the 3rd grade, and the majority of the sample was from a suburban community. At the conclusion of testing the teachers rated six aspects of the texts and it was found that they did not “positively” agree with the statements “the illustrations fit the text” and “the books were leveled appropriately”. As a result the texts for the DRA K-3 were reviewed, and edited, rewritten, or replaced to make them grade appropriate. Another field test was conducted on the DRA K-3 alternative texts in 2000. It involved 208 students, 19 of which were in 3rd grade. From this slight changes were made to the texts (Williams, 2000).

Weber (2000) conducted a study that examined the test-retest reliability and the criterion validity of the DRA K-3 using the Story Level for the analysis. The test-retest reliability portion involved 306 students from four different elementary schools in which

there were approximately three weeks between the two test administrations. The obtained correlation coefficients ranged from .92 to .99, supporting its reliability. The criterion validity portion of the study involved 300 students ranging from first to third grade. The DRA data was compared to reading comprehension data from the Iowa Tests of Basic Skills and the obtained correlation coefficients were .54 for third grade, .84 for second grade, and .65 for first grade. Weber concluded that performance on the DRA K-3 is predictive of the performance on the reading comprehension sections of the Iowa Tests of Basic Skills. This suggests a moderate level of criterion validity for third-grade students.

Williams (1999) examined the inter-rater reliability of the scoring and the internal consistency of the DRA K-3. The original teacher and two additional blind raters scored 306 students' taped testing sessions. In all there were 127 originating teachers and raters whom had prior experience administering the DRA. The study utilized the story level, accuracy, comprehension, reading stage, phrasing, and reading rate. The analysis determined that inter-rater agreement was .80 between the original teacher and the second rater, however it was .74 across all three raters. The internal consistency was assessed by correlating individual scores from 2,470 participants in the second grade on the DRA with their scores during the fall of third grade on the Iowa Tests of Basic Skills Subscales: Vocabulary, Reading Comprehension, and Total Reading (Williams, 1999). It was found that all correlations were significant at the .01 level, however the most meaningful and strongest correlation was with Total Reading at .71.

Summary

The acquisition of literacy skills is vital for children to be successful in school and in life. A critical component in ensuring that a child is acquiring literacy skills involves

accurately assessing their progress. Reading assessments vary in their use from screening measures, to progress monitoring devices, to diagnostic instruments, to outcome measures. Three reading assessments were utilized in the current study, the DIBELS, the DRA, and the TerraNova 2nd Edition. The DIBELS and the DRA are designed for classroom use and both purport to measure certain aspects of reading. However, the research base behind each assessment differs ranging from extensive research background with the DIBELS to limited research background with the DRA. The present study investigated the relationships among the DIBELS Oral Reading Fluency and DRA as related to the TerraNova 2nd Edition Vocabulary and Comprehension for third-grade students.

Chapter III

Method

Participants

Participants were 42 third grade students from an urban elementary school in western New York. This elementary school has pre-kindergarten through sixth grade students and has a student body of 566 individuals, of whom 61% were African American, 23% were Hispanic, 14% were White, and less than 1% were Asian/Pacific Islander. In 2004, 39% of the 4th graders in this school earned a passing score on the New York State's English Language Arts exam, compared to a 70% state average. Of the students attending, 89% were eligible for reduced or free lunch and 21% had an Individualized Education Plan (IEP).

The particular school was selected because it was receiving Reading First funding and was conducting frequent assessments of reading progress. The school was using DIBELS and DRA to monitor reading progress and utilized the TerraNova 2nd Edition at the end of the school year. Teachers and a school based assessment team collected the data during the 2004-2005 school year. The collection teams were trained in a one-day workshop in DIBELS in the spring of 2004, and follow up sessions occurred periodically to ensure sufficient training and understanding of the test. Both the DIBELS and the DRA were administered on an individual basis and the TerraNova 2nd Edition was administered in a group setting.

These participants had met the set exclusion criteria of having at least two sets of data for the DIBELS Oral Reading Fluency, the DRA, and one set of data from the TerraNova 2nd Edition Vocabulary and Comprehension. The DRA also must have a date

on the protocol, the student identification number, date the assessment was given, and relevant scoring sections completed: phrasing and fluency, intonation, accuracy, and comprehension.

Measures

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

DIBELS Oral Reading Fluency is administered to a student over a 1-minute time period. In that one minute the student is asked to read as quickly and correctly as they can. The examiner marks words omitted, substituted, and hesitated on for more than three seconds as errors and at the end of one minute records the number of words read. The oral reading fluency rate is then calculated by subtracting the errors from the total words attempted, resulting in a score that represents the number of words read correctly in one minute (Kaminski & Good, 1996).

Developmental Reading Assessment (DRA)

Five sections of the DRA were utilized for this study. The five include Phrasing and Fluency, Intonation, Accuracy, Comprehension, and Story Level. The story is selected by the student being tested from a group of three to four texts. The level of text, Story Level, is selected by the teacher to attempt to match the student's independent reading level. The teacher then rated the student's phrasing and fluency, intonation, comprehension level, and accuracy as the student read the text and responded to teacher questions.

A coding sheet was created in order to consolidate the data from the DRA protocols (see Appendix A). It consisted of six sections. The identifying information included a place for the student's identification number, date the assessment was given,

and a place to note the other information on the coding sheet for easy access. The phrasing and fluency section was coded on a six-point scale with 1 being “word by word” and 6 being “in longer phrases with rate adjusted appropriately”. The intonation section was also coded on a six-point scale with 1 being “no intonation; monotone” and 6 being “begins to explore subtle intonation that reflects mood, pace, and tension”. The accuracy section listed the percentage of words read correctly. The comprehension section included four descriptions that corresponded with the different comprehension numbers, 6-24. These were “Very Little Comprehension”, “Some Comprehension”, “Adequate Comprehension”, and “Very Good Comprehension”. Lastly, the story level section contained three descriptions that the story level numbers fell under for third graders. These included “Below Grade Level”, “On Grade Level”, and “Above Grade Level”. For both comprehension level and story level the number, not the descriptor, was recorded.

TerraNova2nd Edition

The TerraNova 2nd Edition, also known as the California Achievement Test 6th Edition (CAT/6), is a norm referenced test that is designed to measure school based concepts and student achievement. It was standardized with over 275,000 students between 1999-2000 and contained a nationally representative, stratified random sample based on geographic region, school size, socioeconomic status, race/ethnicity, and type of school system (CTB/ McGraw-Hill, 2001). The TerraNova 2nd Edition was designed to assess how well students are acquiring the basic skills of a typical curriculum (CTB/ McGraw-Hill, 2001). The test content was derived from curriculum guides, teacher groups, and state frameworks.

The TerraNova 2nd Edition utilizes a standard setting method entitled the Bookmark Procedure (Lewis, Mitzel, & Green, 1996; Lewis, Mitzel, Baum, & Patz, 1998). This procedure studies the test items in order of difficulty, determines what a student needs to do in order to answer the items, and writes descriptions of the expected student performance at each level (CTB/ McGraw-Hill, 2001). The performance levels are described as “Advanced”, “Proficient”, “Nearing Proficiency”, “Progressing”, and “Step 1/Starting Out.” It is noted that the Bookmark Procedure has been used to set cut scores for 18 different states’ assessment programs (CTB/ McGraw-Hill, 2001).

The TerraNova 2nd Edition Reading/Language Arts portion of the test examines the student’s basic and higher-order thinking skills as it applies to reading comprehension, language expression, vocabulary, and reference skills. The content is reflective of the goals from districts and states, the standards for English/Language Arts, and the frameworks of the National Assessment of Educational Progress (CTB/ McGraw-Hill, 2001). The exam utilizes passages of authentic traditional and contemporary literature, newspaper, and magazine articles. The student results on the test include both norm-referenced scores, such as national percentiles, national stanines, grade equivalents, and normal curve equivalents, and a criterion-referenced score called an objectives performance index. The publishers recommend using the objectives performance index to inform instruction because it breaks down the test into components that the student needs to master. It indicates on a scale of 0-100 the student’s level of mastery of the particular component for the subject (CTB/ McGraw-Hill, 2001). Based on this information the teacher can target specific skills within a content area. For the purposes of this study the comprehension and vocabulary sections were utilized as a criterion measure.

Design and Procedure

This study was conducted using archival data from an urban elementary school. Protocols for the DRA were obtained from the district office, copied, a number was assigned to each student, and all identifying information was blacked out. A decision was made to determine if the protocol met the set exclusionary criteria including having at least two sets of data for the DIBELS Oral Reading Fluency, the DRA, and one set of data from the TerraNova 2nd Edition Vocabulary and Comprehension. The DRA also must have a date on the protocol, the student identification number, and relevant scoring sections completed: phrasing and fluency, intonation, accuracy, and comprehension. The protocols that were not excluded were then coded (see Appendix A). The codes were then entered into SPSS 13.0 (SPSS Inc., 2004) along with the obtained DIBELS scores and TerraNova 2nd Edition Vocabulary and Comprehension scores.

Of the 42 students who met the aforementioned exclusionary criteria, a number had been assessed with the DRA more than one time during one or more of the three assessment periods. For example, student 'A' may have been administered four different story levels within a three-day period in the fall. For this investigation, only one assessment from each student was considered for analysis from each assessment period. The assessment chosen for analysis was selected randomly from the SPSS database.

Data Analysis

Pearson bivariate correlations were used to determine if relationships existed between DIBELS, the DRA, and the TerraNova 2nd Edition. Regression analyses were completed to determine the best predictor of the winter and spring DIBELS Oral Reading Fluency scores and the TerraNova 2nd Edition Vocabulary and Comprehension.

Chapter IV

Results

Descriptive statistics for the DRA, TerraNova 2nd Edition, and DIBELS Oral Reading Fluency are presented in Tables 1, 2, and 3, respectively. Correlations and regressions were then run to address the four research questions, as described below.

1. What is the relationship between the DIBELS Oral Reading Fluency and the DRA for third grade students?

Table 4 provides the correlations between the components of the DRA and the DIBELS Oral Reading Fluency for fall, winter, and spring. In the fall, DIBELS scores were significantly correlated at the .01 level with DRA Phrasing and Fluency, DRA Intonation, DRA Accuracy, and DRA Story Level. In the winter, DIBELS scores were significantly correlated at the .01 level with DRA Phrasing and Fluency, DRA Intonation, and DRA Story Level. In the spring, DIBELS scores were significantly correlated at the .01 level with DRA Phrasing and Fluency and DRA Story Level. They were also correlated at the .05 level with DRA Accuracy.

Table 1

Means and Standard Deviations for Third Grade Fall, Winter, and Spring DRA Results

Subscale	Mean	Standard Deviation
DRA – Fall (n=42)		
DRA-PF	3.76	1.01
DRA- I	2.74	.80
DRA – A	96.10	1.32
DRA – CL	16.88	2.64
DRA – SL	22.48	8.51
DRA – Winter (n= 40)		
DRA-PF	3.60	1.17
DRA- I	3.10	1.03
DRA – A	96.28	2.76
DRA – CL	16.18	3.26
DRA – SL	28.45	8.13
DRA – Spring (n= 30)		
DRA-PF	3.70	1.19
DRA- I	3.10	.99
DRA – A	95.30	2.26
DRA – CL	15.70	3.03
DRA – SL	32.60	7.37

Table 2

Means and Standard Deviations for Third Grade DIBELS Oral Reading Fluency

DIBELS ORF	Mean	Standard Deviation	Sample Size
Fall	55.21	21.55	43
Winter	68.59	42.96	44
Spring	90.41	30.57	44

Table 3

Mean and Standard Deviation for Third Grade TerraNova 2nd Edition Results

TerraNova 2nd Edition Mean		Standard Deviation	Sample Size
Comprehension	611.66	48.32	44
Vocabulary	611.14	42.96	44

Table 4

Correlations and Significance for DIBELS ORF with the DRA for Third Graders

DRA	DIBELS ORF Fall	DIBELS ORF Winter	DIBELS ORF Spring
Fall	(n= 42)	(n=42)	(n= 42)
DRA-PF	.650 (p = .000)	.674 (p = .000)	.633 (p = .000)
DRA- I	.526 (p = .000)	.570 (p = .000)	.452 (p = .003)
DRA – A	.427 (p = .005)	.441 (p = .003)	.447 (p = .003)
DRA – CL	.211 (p = .181)	.159 (p = .314)	.185 (p = .242)
DRA – SL	.778 (p = .000)	.832 (p = .000)	.827 (p = .000)
Winter	(n=39)	(n=40)	(n=40)
DRA-PF	.610 (p = .000)	.698 (p = .000)	.593 (p = .000)
DRA- I	.478 (p = .002)	.531 (p = .000)	.458 (p = .003)
DRA – A	.328 (p = .042)	.276 (p = .084)	.257 (p = .109)
DRA – CL	-.232 (p = .156)	-.161 (p = .321)	-.180 (p = .266)
DRA – SL	.774 (p = .000)	.830 (p = .000)	.809 (p = .000)
Spring	(n=29)	(n=30)	(n=30)
DRA-PF	.772 (p = .000)	.717 (p = .000)	.615 (p = .000)
DRA- I	.326 (p = .085)	.316 (p = .089)	.253 (p = .177)
DRA – A	.390 (p = .036)	.292 (p = .117)	.463 (p = .010)
DRA – CL	-.064 (p = .743)	-.037 (p = .844)	-.152 (p = .423)
DRA – SL	.730 (p = .000)	.769 (p = .000)	.718 (p = .000)

To determine the extent to which specific variables predicted DIBELS Oral Reading Fluency and TerraNova 2nd Edition Vocabulary and Comprehension performance multiple regression analyses was performed using a stepwise procedure. With the stepwise procedure, specific variables are entered into the analysis if they meet set criteria. As new variables enter into the model, each variable is reanalyzed to determine the extent to which it continues to meet the qualifications. If a given variable loses its individual predictive value, after other variables enter the equation, it is removed. For the purpose of these analyses, the “F to enter” the equation had to be greater than or equal to .05. The criterion for “F to remove” was greater than or equal to .1.

A series of stepwise multiple regression analyses were performed to determine the extent to which the different components of the DRA predicted the DIBELS Oral Reading Fluency for winter and spring. Table 5 provides the unstandardized regression coefficients (B), the standard error of the unstandardized regression coefficient, the standardized regression coefficients (β), and the squared semi-partial correlation for DIBELS Oral Reading Fluency for winter. It was found in that the DRA Story Level uniquely accounted for 53.9% of the variance and the DRA Accuracy uniquely accounted for 4.0% of the variance in DIBELS Oral Reading Fluency for winter. As seen in Table 6, it was found in that the DRA Story Level uniquely accounted for 69.7% of the variance in DIBELS Oral Reading Fluency for spring.

Table 5

Summary of Stepwise Regression Analysis for Variables Predicting DIBELS ORF Winter Scores (N=42)

Variable	<i>B</i>	<i>SE B</i>	β	Squared Semi-Partial
Step 1				
DRA-SL Fall Score	2.270	.236	.832**	.692
Step 2				
DRA-SL Fall Score	2.098	.236	.769**	.539
DRA-A Fall Score	3.700	1.521	.211*	.04

* $p < .05$

** $p < .01$

Table 6

Summary of Stepwise Regression Analysis for Variables Predicting DIBELS ORF Spring Scores (N=38)

Variable	<i>B</i>	<i>SE B</i>	β	Squared Semi-Partial
Step 1				
DRA-SL Fall Score	3.063	.337	.835**	.697

** $p < .01$

2. How does the DIBELS Oral Reading Fluency relate to the TerraNova Comprehension and Vocabulary for third grade students?

Table 7 provides the correlations between DIBELS and the TerraNova 2nd Edition. DIBELS scores for the fall, winter, and spring were all significantly correlated at the .01 level for the TerraNova 2nd Edition Vocabulary and Comprehension, with the exception of the DIBELS winter score and the Terra Nova 2nd Edition Comprehension score being correlated at the .05 level.

Table 7

Correlations and Significance for DIBELS ORF with the TerraNova 2nd Edition for Third Graders

	TerraNova 2nd Edition DIBELS ORF Fall (n=43)	DIBELS ORF Winter (n=44)	DIBELS ORF Spring (n=44)
Comprehension	.458 (p = .002)	.379 (p = .011)	.576 (p = .000)
Vocabulary	.569 (p = .000)	.549 (p = .000)	.639 (p = .000)

3. How does the DRA relate to the TerraNova Comprehension and Vocabulary for third grade students?

Table 8 provides the correlations between the DRA and the TerraNova 2nd Edition. The DRA Phrasing and Fluency was significantly correlated at the .01 level for the fall and at the .05 level for winter and spring with the Terra Nova 2nd Edition Vocabulary. It was also significantly correlated at the .05 level for the winter with TerraNova 2nd Edition Comprehension. The DRA Intonation for the winter was significantly correlated at the .05 level for both the TerraNova 2nd Edition Comprehension and Vocabulary. DRA Accuracy for the spring was significantly correlated at the .01 level with the TerraNova 2nd Edition Comprehension. For both the TerraNova 2nd Edition Vocabulary and Comprehension, the DRA Story Level was significantly correlated at the .01 level for fall and winter and at the .05 level for spring.

Table 8

Correlations and Significance for TerraNova 2nd Edition with the DRA for Third Graders

DRA	TerraNova 2nd Edition Comprehension	TerraNova 2nd Edition Vocabulary
Fall	(n=42)	(n=42)
DRA-PF	.237 (p = .130)	.481 (p = .006)
DRA- I	.285 (p = .068)	.282 (p = .070)
DRA – A	.215 (p = .172)	.203 (p = .197)
DRA – CL	.219 (p = .164)	.219 (p = .164)
DRA – SL	.582 (p = .000)	.746 (p = .000)
Winter	(n=40)	(n=40)
DRA-PF	.379 (p = .016)	.374 (p = .017)
DRA- I	.354 (p = .025)	.388 (p = .013)
DRA – A	.301 (p = .059)	.233 (p = .148)
DRA – CL	.094 (p = .565)	-.122 (p = .452)
DRA – SL	.553 (p = .000)	.712 (p = .000)
Spring	(n=30)	(n=30)
DRA-PF	.320 (p = .085)	.387 (p = .034)
DRA- I	.031 (p = .872)	.069 (p = .717)
DRA – A	.579 (p = .001)	.310 (p = .096)
DRA – CL	-.155 (p = .413)	.038 (p = .843)
DRA – SL	.428 (p = .018)	.386 (p = .035)

4. Which classroom assessment, the DIBELS or the DRA, is the strongest predictor of student performance on the TerraNova Comprehension and Vocabulary for third grade students?

A stepwise multiple regression analysis was performed to determine the extent to which specific variables, the DRA and DIBELS Oral Reading Fluency, predicted TerraNova 2nd Edition Vocabulary and Comprehension performance, as seen in Table 9 and Table 10. It was found that the DRA Story Level uniquely accounted for 60.1% of the variance in TerraNova 2nd Edition Comprehension and 37.3% of the variance in TerraNova 2nd Edition Vocabulary.

Table 9

Summary of Stepwise Regression Analysis for Variables Predicting TerraNova2nd Edition Comprehension Scores (N=38)

Variable	<i>B</i>	<i>SE B</i>	β	Squared Semi-Partial
Step 1				
DRA-SL Fall	4.152	.564	.775**	.601

** $p < .01$

Table 10

Summary of Stepwise Regression Analysis for Variables Predicting TerraNova2nd Edition Vocabulary Scores (N=38)

Variable	<i>B</i>	<i>SE B</i>	β	Squared Semi-Partial
Step 1				
DRA-SL Fall	3.658	.790	.611**	.373

** $p < .01$

Chapter V

Discussion

This research was done to investigate the relationship among three reading assessments, DIBELS Oral Reading Fluency, the DRA, and the TerraNova 2nd Edition Comprehension and Vocabulary. The previous research in the area of reading assessment had not addressed these measures, but it is important to investigate this area to inform school districts utilizing these reading assessments about the utility of the three tests.

It was hypothesized that DRA Story Level, DRA Comprehension and DRA Accuracy would be strongly correlated with DIBELS Oral Reading Fluency for fall, winter, and spring. This is because (a) the DRA Story Level appears to be an overall predictor of reading achievement, (b) previous research suggests that DIBELS are an indirect measure of comprehension, and (c) the DRA Accuracy is similar to DIBELS Oral Reading Fluency.

When looking at the fall data, the DRA Phrasing and Fluency, Intonation, Accuracy, and Story Level measures were all strongly correlated with DIBELS Oral Reading Fluency fall scores. This result was reasonable considering that a student's intonation, ability to phrase written language, and read more accurately is directly related to fluency. Surprisingly, the only aspect of the measure that was not significantly correlated was DRA Comprehension Level. When examining the raw data closely, individual results from the DRA Comprehension Level appear to remain relatively consistent across various text difficulties. It is not known if range restriction played a role in the lack of a significant correlation.

Analyses of the winter data yielded slightly different results. DRA Accuracy and Comprehension were the only measures not correlated with DIBELS Oral Reading Fluency for the winter. With the spring data, it was found that DRA Phrasing and Fluency and DRA Story Level were strongly correlated and DRA Accuracy was moderately correlated with DIBELS. DRA Phrasing and Fluency and Story Level were the only measures that had consistently strong correlations for fall, winter, and spring with DIBELS Oral Reading Fluency. Again, a child's oral reading fluency is tied to the difficulty level of the text and their phrasing and fluency.

The results of the stepwise regression analyses show that the fall DRA Story Level was the best predictor of DIBELS Oral Reading Fluency scores, uniquely accounting for half of the variance in the winter and almost three quarters of the variance in the spring. DRA Accuracy uniquely accounted for a small percentage of the variance in the winter DIBELS Oral Reading Fluency scores. Not surprisingly, the DRA Story Level accounted for the majority of the variance with DIBELS Oral Reading Fluency because it represents a difficulty level of text that is geared toward the individual student's reading ability. Although DRA Accuracy significantly accounted for only a small portion of variance, it is a logical relationship because DIBELS Oral Reading Fluency is partially a measure of accuracy. However, this small portion of the variance is not practically meaningful.

Also hypothesized was that the DIBELS Oral Reading Fluency and DRA Story Level would be strongly correlated with TerraNova Comprehension, and DRA Story Level would be strongly correlated with Vocabulary. Again, this is because the DRA Story Level appears to be an overall predictor of reading achievement and on the surface

appears to encompass more skills than the other DRA measures. The DIBELS are a measure of comprehension, which is why a correlation with the TerraNova Comprehension was hypothesized. It was found that DIBELS Oral Reading Fluency for fall, winter, and spring were strongly correlated to TerraNova Vocabulary. This was surprising because DIBELS Oral Reading Fluency is not a measure of vocabulary. However, with the TerraNova Vocabulary, reading is involved and relates to success on the test. The TerraNova Vocabulary was also strongly correlated to the fall and winter DRA Story Level. However, it was moderately correlated to the fall, winter and spring DRA Phrasing and Fluency, winter DRA Intonation, and spring DRA Story Level.

When looking at TerraNova Comprehension, fall and winter DIBELS Oral Reading Fluency scores were moderately correlated and spring DIBELS Oral Reading Fluency scores were strongly correlated. This is commensurate with prior research, which suggests DIBELS Oral Reading Fluency correlates with other well-established tests of comprehension. This provides further support that DIBELS is an adequate measure of comprehension. Also for the TerraNova Comprehension, fall and winter DRA Story Level, and spring DRA Accuracy were strongly correlated. Several other DRA measures including winter Phrasing and Fluency, winter Intonation, and spring Story Level were moderately correlated. The DRA Comprehension Level was not correlated with the TerraNova Comprehension, providing further evidence either for range restriction or that this portion of the DRA does not truly measure comprehension.

The results of the stepwise regression analyses show that the best predictor for TerraNova Comprehension and Vocabulary was the fall DRA Story Level, which uniquely accounted for 60% of the variance in comprehension and 37% in vocabulary. It

remains an open question as to why the fall DRA Story Level accounted for so much variance and winter DRA Story Level accounted for none of the variance.

Implications for Theory and Practice

Knowing the best predictor for a student's performance on standardized testing is especially relevant for educators since the advent of and emphasis on high stakes standardized testing. This study found that the best predictor for the TerraNova Comprehension and Vocabulary for third-grade students was the DRA Story Level. Because the DRA Story Level is a teacher derived number, in other words the teacher selects the level of text the student will be reading, this research demonstrates tentatively that one of the best predictors for this particular standardized test is teacher judgment on how well a student is reading. However, this is not likely a pure measure of teacher judgment because the Story Level is adjusted after the student completes the assessment. Also, numerous variables could be related to the teacher judgment such as training, years of experience, and interaction between the teacher and the student. The DRA Story Level was also the best predictor for DIBELS Oral Reading Fluency scores.

Prior research has identified that R-CBM is more accurate at identifying low readers than teacher judgment alone (Madelaine & Wheldall, 2005). Madelaine and Wheldall (2005) warned against solely using teacher judgments for low readers because they found that only 15% could accurately identify the three lowest readers in their classroom. They offered that R-CBM might be a more objective measure that is still time efficient. However, in the current study the sample included poor and proficient readers suggesting that teacher judgment might be sufficient in an overall class. This being said, since the current study did not differentiate between the two types of readers implications

can not be given as to which assessment technique would be best for either. Instead it can only offer that the DRA Story Level, as a tentative product of teacher judgment, as a whole was the best predictor for the TerraNova 2nd Edition Comprehension and Vocabulary and DIBELS Oral Reading Fluency.

This study offers some points of consideration for those who need an understanding of the reading assessments within school district. When making decisions based solely on the predictive power for the TerraNova Vocabulary and Comprehension, the DRA Story Level may be utilized. Other measures should also be included since approximately 40% of the variance for the TerraNova Comprehension and 63% of the variance for the TerraNova Vocabulary remains unaccounted for. However, when taking into consideration other factors such as efficiency and ease of administration, it is noted that DIBELS also was highly correlated with both the TerraNova Vocabulary and Comprehension. Ultimately, the assessments that a school district utilizes will draw from a variety of factors including the current research in the field and the legal mandates for multi-method nondiscriminatory assessments.

Limitations and Directions for Future Research

Several limitations existed in this study. First, an existing data set was utilized and the researcher had no control over the data collection. The extent to which the examiners were trained in administering the measures was not known. Members of the school district entered the data into a database; however, it was not known what, if any, steps were taken to ensure accuracy. Regarding the data collection, it is not known whether the students were first administered the DIBELS or the DRA. It is possible that some students took the DIBELS test before the DRA, whereas others took the DRA first. The

order of administration could affect teachers regarding the DRA text selection and could possibly influence the teachers' scoring of the DRA. For example, a teacher who had previously administered the DIBELS might select a more appropriate text level for the student than one who had no DIBELS information prior to administering the DRA.

While these points do pose significant limitations to the internal validity of this study, this is likely a reflection of how information is obtained and used within a large school district.

Numerous DRA protocols were not completed in their entirety and were therefore discarded from analysis. Additionally, scores of particular students were excluded to allow consideration of only one data point from each assessment period. Additional exclusions were made to allow for consideration of only one student data point from each assessment period. For example, many students took the DRA two or more times on a single day, yet only one test protocol, selected randomly, was analyzed. Inner rater reliability was not calculated when the data was coded from the DRA protocols.

The final limitation considers the analysis of data. Pearson correlations were run to assess the strength of correlations between the assessment tools. As a result, the DRA data, although it is likely ordinal data, was treated as though it was interval data.

Furthermore, the extent to which each variable was normally distributed was not investigated. This is an inherent assumption when conducting correlational analyses. As a result, the extent to which the results of correlational analyses would be replicated in future studies is not known. Another consideration related to the analysis of data relates to the use of the Stepwise multiple regression procedure. The extent to which predictor variables that were entered into the stepwise equation are related is not known. As a

result, some variables included in the output may have been due to chance. This is a particular consideration for those variables that contributed a small portion to prediction, and that are not likely to be related to the criterion based on other evidence.

Many of these limitations could be controlled for in future research. To increase the internal validity to make more sound judgments from the results the examiners, who are blind to the purpose of the study, could be trained to reliably conduct the test administration and score the protocols. This study could also be replicated to include a larger sample size. This could be limited to one school district with more schools included in the sample or utilize a variety of school districts and educational settings. Another facet that could be explored is determining the predictive validity of the DRA and DIBELS to the English Language Arts exam or other state mandated tests that measure reading. The current research did not differentiate between the poor and proficient readers in the sample. Another study could differentiate between these two types of readers to better determine if different assessment techniques are necessary.

References

- Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Barger, J. (2003). *Comparing the DIBELS oral reading fluency indicator and the North Carolina end of grade reading assessment* (Technical Report). Ashville, NC: North Carolina Teacher Academy.
- Bennett, K.J., Brown, K.S., Boyle, M., Racine, Y., Offord, D. (2003). Does low reading achievement at school entry cause conduct problems? *Social Science and Medicine*, 56, 2443-2448.
- Buck, J., & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* (Technical Report). Tallahassee, FL: Florida Center for Reading Research.
- Chard, D., & Kame'enui, E. (2000). Struggling first-grade readers: The frequency and progress of their reading. *The Journal of Special Education*, 34, 28-38.
- Children's Educational Services, I. (1987). Test of Reading Fluency. Minneapolis, MN: Author.
- Deno, S. (1997). Whether thou goest... Perspectives on progress monitoring. In J Lloyd, E. Kame'enui, & D. Chard (Eds.), *Issues in educating students with disabilities* (pp. 77-99), Mahwah, NJ: Lawrence Erlbaum Associates.
- Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219-232.

- Deno, S., Mirkin, P., & Chiang, B. (1982). Identifying valid measures of reading, *Exceptional Children*, 49, 36-45.
- Felton, R. & Wood, F. (1992). A reading level match study of nonword reading skills in poor readers with varying IQ. *Journal of Learning Disabilities*, 25, 318 – 326.
- Foorman, B., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research and Practice*, 16 (4), 203-212.
- Fuchs, L., & Deno, S. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, 57, 488-501.
- Fuchs, L., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice*, 13, 204-219.
- Fuchs, L., & Fuchs, D. (1986). Curriculum-based assessment of progress toward long term and short term goals. *Journal of Special Education*, 20, 69-82.
- Fuchs, L., Fuchs, D., & Compton, D. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children*, 71 (1), 7-21.
- Fuchs, L., Fuchs, D., & Hamlett, C. (1989). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children*, 55, 429-438.
- Fuchs, L., Fuchs D., Hosp, M., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5, 239-256.
- Fuchs, L.S., Fuchs, D. & Maxwell (1988). The validity of informal reading

- comprehension measures. *Remedial and Special Education*, 9, 20-28.
- Good, R., Wallin, J., Simmons, D., Kame'enui, J., & Kaminski, R. (2002) *System-wide percentile ranks for DIBELS benchmark assessment* (Technical Report). Eugene, OR: University of Oregon.
- Goodman, K. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, 6, 126-135.
- Gravois, T., & Gickling, E. (2002). Best practices in curriculum-based assessment. In A Thomas, & J Grimes (Eds.), *Best Practices in School Psychology IV* (pp. 885-898), Bethesda, MD: The National Association of School Psychologists.
- Haager, D., & Windmueller, M. (2004). Early reading intervention for English language learners at risk for learning disabilities: Student and teacher outcomes in an urban school. *Learning Disability Quarterly*, 24, 235-250.
- Harris, T., & Hodges, R. (Eds.) The literacy dictionary (p. 207). Newark, DE: International Reading Association.
- Harste, A., Woodward, V., & Murke, C. (1984). *Language stories and literacy lessons*. Portsmouth, NH: Heinemann.
- Hoover, H., Dunbar, S., & Frisbie, D. *Iowa test of basic skills*. Rolling Meadows, IL, Riverside Publishing; 2001.
- Jenkins, J. Fuchs, L., van den Broek, P., Espin, C, & Deno, S. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, 95, 719-729.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80, 437 – 447.

- Kaminski, R., & Good, R. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, 25 (2), 215-227.
- Kaminski, R., & Good, R. (1998). Assessing early literacy skills in a problem-solving model: Dynamic indicators of basic early literacy skills. In M. Shinn (Ed.), *Advanced applications of Curriculum-Based Measurement* (pp.113-142). New York: Guilford.
- Kibby, M. (1995). *Practical steps for informing literacy instruction: A diagnostic decision making model*. International Reading Association: Delaware: Newark.
- LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.
- Lewis, D., Mitzel, H., & Green, D. (1996). Standard setting: A bookmark approach. In D. R. Green (Chair), *IRT-Based Standard-Setting Procedures Utilizing Behavioral Anchoring*. Symposium conducted at the 1996 Council of Chief State School Officers 1996 National Conference on Large Scale Assessment, Phoenix, AZ.
- Lewis, D., Green, D., Mitzel, H., Baum, K., & Patz, R. (1998, April). *The bookmark standard setting procedure: Methodology and recent implications*. Paper presented at the 1998 National Council for Measurement in Education annual meeting, San Diego, CA.
- Lipson, M. (1983). The influence of religious affiliation on children's memory for text information. *Reading Research Quarterly*, 18, 488-457.
- Madelaine, A., & Wheldall, K. (2005). Identifying low progress readers: Comparing teacher judgment with a curriculum-based measurement procedure. *International Journal of Disability, Development, and Education*, 52 (1), 33-42.

- Martson, D.B. (1989). *A curriculum-based measurement approach to assessing academic performance: What is it and why do it?* In M.R. Shinn (Ed.), Curriculum-based measurement: Assessing special children (18-78). New York: The Guilford Press.
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.
- No Child Left Behind Act of 2001*, 107th Congress 107-110 (2002).
- Pearson Learning Group: Developmental Reading Assessment (DRA) K-8 Technical Manual. Retrieved October 10, 2005, from http://www.pearsonlearning.com/correlation/rsp/DRA_Technical_Manual.pdf
- Shaw, D., & Shaw, R. (2002). *DIBELS Oral Reading Fluency – based indicators of third grade reading skills for Colorado State Assessment Program (CSAP)* (Technical Report). Eugene, OR: University of Oregon.
- Shin, M. (2002). Best practices in using curriculum based measurement in a problem

solving model. In A Thomas, & J Grimes (Eds.), *Best Practices in School Psychology IV* (pp. 671-697), Bethesda, MD: The National Association of School Psychologists.

Simmons, D. & Kame'enui, E. (Eds.). (1998). *What reading research tells us about children with diverse learning needs: Bases and basics*. Mahwah, NJ: Erlbaum.

Stanovich, K. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford.

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-406.

Stanovich, K. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16, 32-71.

Salvia, J., & Ysseldyke, J. (2004) *Assessment in special and inclusive education: Ninth edition*. New York: Houghton Mifflin Company.

TerraNova, The Second Edition: Technical Quality. (2001). Monterey, California: CTB/McGraw-Hill.

Weber, W. (2000). Developmental reading assessment and evaluacion del desarrollo de la lectura: A validation study. Retrieved October 10, 2005 from http://www.pearsonlearning.com/correlation/rsp/ResearchPaper_DRA_Weber.pdf

Williams, E. (no date). *Preliminary validation of the Developmental Reading Assessment*. Parsippany, NJ: Celebration Press.

Williams, E. (1999). Developmental reading assessment: Reliability study 1999.

Retrieved October 10, 2005 from

www.pearsonlearning.com/correlation/rsp/DRA/doc.

Williams, E. (2000). *Alternative Developmental Reading Assessment: Development and field-testing results*. Research paper published by Pearson Learning Group, Parsippany, NJ.

Wilson, J. (2005). *The relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) oral reading fluency to performance on Arizona Instrument to Measure Standards (AIMS)* (Technical Report). Tempe, AZ.

Appendix A Coding Sheet for 3rd Grade

Student's ID #:

Date:

Phrasing and Fluency: _____

Intonation: _____

Accuracy: _____

Comprehension Level: _____

Story Level: _____

Phrasing and Fluency:

- 1 – word by word
- 2 – in short phrases at times
- 3 – in short phrases most of the time
- 4 – in longer phrases at times; inconsistent rate
- 5 – in longer phrases most of the time; adequate rate
- 6 – in longer phrases; rate adjusted appropriately

Intonation:

- 1- no intonation; monotone
- 2 – little intonation; rather monotone
- 3- some intonation; some attention to punctuation; monotone at times
- 4 – adjusts intonation to convey meaning at times; attends to punctuation most of the time
- 5 – adjusts intonation to convey meaning; attends to punctuation
- 6 – begins to explore subtle intonation that reflects mood, pace, and tension

Accuracy:

100; 99; 98; 97; 96; 95; 94; 93; 92; 91; 90; 89; 88

Comprehension:

Very Little Comprehension

6; 7; 8; 9

Some Comprehension

10; 11; 12; 13; 14; 15

Adequate Comprehension

16; 17; 18; 19; 20; 21

Very Good Comprehension

22; 23; 24

Story Level:

Below Grade Level:

12; 14; 16; 18; 20

On Grade Level:

28; 30; 34

Above Grade Level:

34; 38; 40; 44