Relationships among Reading Assessments: How do DIBELS Oral Reading Fluency and DRA Scores relate to TerraNova2 Performance in Second-Grade Students?

Melissa Walker

Follow this and additional works at: https://scholarworks.rit.edu/theses

Recommended Citation

This Thesis is brought to you for free and open access by RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.
Relationships among Reading Assessments: How do DIBELS Oral Reading Fluency and DRA Scores relate to TerraNova2 Performance in Second-Grade Students?

Graduate Thesis

Submitted to the Faculty

Of the School Psychology Program

College of Liberal Arts
ROCHESTER INSTITUTE OF TECHNOLOGY

By

Melissa Walker

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

Rochester, New York

June 13, 2007

Approved: Suzanne B. Graney
Suzanne B. Graney, Ph.D.

Vincent Pandolfi
Vincent Pandolfi, Ph.D.
PERMISSION GRANTED
Title of thesis Relationships among Reading Assessments: How do DIBELS Oral Reading Fluency and DRA Scores Relate to Terra Nova 2 Performance in Second-Grade Students?

I __________________________ hereby grant permission to the 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: Melissa Walker

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: __________________________
Relationships among Reading Assessments: How do DIBELS Oral Reading Fluency and DRA Scores relate to TerraNova2 Performance in Second-Grade Students?

Melissa Kay Walker

Rochester Institute of Technology
Abstract

The present study investigated the relationships between the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency and the Developmental Reading Assessment (DRA) as related to TerraNova2 Vocabulary and Comprehension scores for second grade students. Archival data gained from 60 participants during the 2004-2005 school year were utilized. The participants were from a school receiving Reading First funding in a northeastern city. Correlation and regression analyses were run on the DIBELS Oral Reading Fluency scores, DRA codes, and TN2 scores to determine the relationships among the three tests. Results found significant correlations and predictors for TN2 performance.
CHAPTER I

STATEMENT OF THE PROBLEM

The ability to read is an essential skill in today's complex society. Without this skill, maintaining a productive life is often difficult. Children who learn to read are more likely to become productive members of society compared to those who never learn the skill (Adams, 1990). The effects of illiteracy are widespread and can be long lasting. For example, a striking percentage of prison inmates are illiterate, as well as a number of unemployed adults. Illiteracy is not only detrimental to adults, but to children as well. Those who fail to learn to read are more likely to be at increased risk for childhood conduct problems (Bennett et al., 2003) and suffer the effects of those behaviors (e.g., involvement in the judicial system, social problems). While learning to read can be difficult for any child regardless of background, the effects of economic disadvantages on reading acquisition can be particularly severe (McGill-Franzen, 1987; Snow, Burns, & Griffen, 1998). Specifically, children from an economically disadvantaged background have less familiarity with print, and less exposure to language. This translates into an academic disadvantage prior to beginning school, perpetuating a cycle that can affect them into adulthood. For many children, particularly those from a disadvantaged background, reading is a particularly difficult skill to learn.

Importance of Reading in Elementary School

Children need to learn to read at a young age (Ehri & McCormick, 1998). Children who are not able to comprehend written language by third grade fall behind their peers in almost every academic activity (National Reading Panel, 2000). The Matthew Effect, described by Walberg and Tsai (1983), refers to the idea that the
students who read well early in their schooling improve at a faster rate than their low achieving peers. An example of the Matthew Effect can be seen in the work of Juel (1988) and Shaywitz et al. (1999). Juel (1988) found that children who had poor reading skills in first grade were almost invariably poor readers three years later. Shaywitz et al. followed students between kindergarten and ninth grade and found that poor readers do not typically catch up to the strong readers. The importance of students learning to read early cannot be overstated.

Reading Assessment

Because students are expected to make large academic gains in the first four years of school, it is necessary to assess their achievement frequently to ensure adequate progress. For those students who are not making expected gains, intense scientifically based reading instruction is given with the goal of elevating progress to an appropriate level. In order for this process to work, reliable and valid reading assessment measures are needed to ensure scores represent accurate estimates of reading achievement (Fuchs & Fuchs, 1999).

Measurement tools used to assess student reading achievement should have certain properties that depend on the reason for use (Fuchs & Fuchs, 1999). Typically, an assessment tool is evaluated based on its utility for screening, progress monitoring, skill diagnosis, or as an outcome measure (Kame’enui et al., 2002). These four uses for tests must be considered prior to choosing an assessment tool due to differences in the desired characteristics of the tool. These characteristics include, but are not limited to the following attributes: the financial and personnel cost of administration, the type of data obtained from the assessment, and the usefulness of the data.
**Instruments of Study**

When a school district makes a decision regarding how to assess student reading achievement, many factors such as financial resources, time, and the utility of acquired information need to be considered. Although a district's decision makers can easily review issues regarding finances and effects on instructional time, they also need to be aware of the quality and utility of information that will be gained with these investments. In many school districts, multiple tests are given to all students to gauge reading achievement. Three reading assessments, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency, Developmental Reading Assessment (DRA), and TerraNova2 (TN2), which are used in combination in one large urban district, are examined within this study. Under specific investigation is the extent to which student performance on the DRA and DIBELS ORF relates to performance on the TN2 Comprehension and Vocabulary measures.

The DIBELS were derived from Curriculum Based Measurement, a set of procedures, which have been empirically validated as an assessment tool that can be used to monitor progress (Deno, Mirkin, & Chiang, 1982; Martson, 1989). Although the DIBELS measures are relatively new to the field of reading assessment, they have demonstrated adequate technical characteristics to be used as an indicator of reading achievement.

The second instrument for examination is Developmental Reading Assessment (DRA). Although this tool is used in school systems to monitor progress and diagnose reading difficulties, empirical support to evaluate the technical properties is limited.
The TN2 (also referred to in the literature at the California Achievement Test 6) is a popular, 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. Its long history of use and technical characteristics make it useful as standardized criterion measure.

Administering all three assessment can be financially expensive, including the purchase of training programs, test materials and scoring procedures; and educationally expensive with the loss of instructional time. Although these three instruments are all administered to cohorts of students, the quality of information gained through their combined use is not known.

Purpose of Study

The focus of this research is on understanding the information that is gained after administering the three previously introduced assessments, which are designed to assess and monitor students’ reading achievement. The outcomes of this study address how the results of DIBELS ORF and DRA are related, and the extent to which they relate to TN2 scores. This study addresses the following research questions:

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

2) How does the DIBELS Oral Reading Fluency relate to TerraNova2 Comprehension and Vocabulary for second grade students?

3) How does the DRA relate to TerraNova2 Comprehension and Vocabulary for second grade students?
4) Do results from the DRA and DIBELS Oral Reading Fluency offer differential information in predicting student performance on the TerraNova2 Comprehension and Vocabulary measures?
CHAPTER II
LITERATURE REVIEW

Research into the importance of proficient reading abilities has a long history. Although the necessity of this skill is widely acknowledged in school settings, many students continue to advance through the educational system without the requisite literacy skills to engage in the complexities of modern society. As the literature base regarding the importance of early acquisition of literacy skills has expanded, teachers have had to adjust their instruction to meet the advances of scientific knowledge. Assessment of literacy skills is needed to ensure instruction facilitates student achievement.

_Literacy Development_

Literacy development begins in infancy with acquisition of language and continues through early childhood as children learn about the symbolic representation of language. During the first three years of schooling, students "learn to read." That is, they develop the capacity to connect written symbols to a language they have heard since birth. To keep pace with curricular expectations, children need to make great strides during the elementary school years. If, by the end of kindergarten, all children had developed phonemic awareness (meaning they could identify and manipulate individual sounds in spoken words) and were proficient with the alphabetic principle, (linking speech sounds to letter symbols), they would be on an appropriate track towards being able to decode words in first grade (Ehri & McCormick, 1998). Along this line of progress, children should be able to decode words, recognize basic sight words, and read some connected text by the end of first grade. By second and third grade, children should be able to read connected text with increasing fluency and accuracy to allow for
comprehension of written material (Ehri & McCormick). Although the initial goal is to learn to read, children are expected to become capable of fulfilling more requirements as they progress through their schooling.

As children progress through elementary school the academic focus turns from learning to read towards using reading as a tool to learn. A major shift in the focus of education usually occurs around the third grade. After this shift occurs, functional reading skills are necessary in order to benefit from instruction about any academic topic. If efficient reading skills are not developed by this time, acquisition of higher-level academic concepts will be difficult (National Reading Panel, 2000). Therefore, the utmost priority is to have all students in primary grades learn how to read fluently and efficiently. The importance of this goal is highlighted with a look at the differences between students who read well and those who do not.

The Matthew Effect, described by Walberg and Tsai (1983), refers to the idea that the performance of students who read well early in their schooling improves at a faster rate than their low achieving peers. Walberg and Tsai attribute this observation to an explanation that students who initially achieve well are able to use new educational experiences more advantageously than their lower achieving counterparts. Stanovich (1986) explains this phenomenon as a result of differential exposure to written text. According to Stanovich’s theory, students who read well in early grades read more, and therefore become more adept readers. However, students who do not read well are not exposed to the same quantity of text, and therefore do not gain reading proficiency. This creates an achievement gap between these students and their peers, which only increases over time. An example of the Matthew Effect can be seen in the work of Juel (1988) and
Shaywitz et al. (1999). Juel found that students who were poor readers at the end of first grade had an 88% probability of being poor readers at the end of fourth grade. Students who were average readers at the end of first grade had an 87% probability of being an average reader in fourth grade. Shaywitz et al. followed students between kindergarten and ninth grade and found that students who were poor readers in primary school were unlikely to catch up to their proficient reading peers. The persistently poor readers were less likely to have plans to complete high school, less likely to spend time reading, and more likely to be expelled from school compared to their proficient reading peers. These findings bring to the surface the importance of identifying students in need of targeted academic interventions early to promote reading success.

Legislation

In recent years, the federal government has recognized the need to hold schools accountable for their students’ progress. The No Child Left Behind Act (2001) is the government’s latest attempt to ensure that all students make “adequate yearly progress” (NCLB, 2001). One program funded under NCLB, Reading First, is aimed at enhancing reading development in high-risk elementary schools. Between the years 2002 and 2005 the United States government spent 3.96 billion dollars towards the goal of establishing literacy in all students in kindergarten through third grade (US Department of Education, 2005). Much of this money was allocated to districts serving large numbers of low-income or minority students who read below grade level. Some of the financial resources from the Reading First Initiative are allocated toward developing scientifically based instructional programs and valid assessment tools to measure reading skills. In order for schools to receive financing, instruction must be provided that leads to successful
outcomes. As a result of the Reading First Initiative (NCLB, 2001), districts are required to demonstrate academic gains in students' reading achievement. Although all schools need to demonstrate yearly progress under NCLB, those schools receiving Reading First funding need to adhere to even higher achievement standards. Districts receiving this grant money need assessment procedures to measure the reading progress of their students.

_Assessment_

Student assessment within schools is typically done for one of four purposes: screening, diagnosis, progress monitoring, and outcome measurement (Kame‘enui et al., 2002). Necessary proprieties of assessment tools vary depending in the way the information gained from assessment will be used. For example, although screening instruments should be easy and efficient to administer and provide information relevant to an outcome skill, they are typically administered only one time, and therefore do not need to have multiple alternative forms. In contrast to screening tools, diagnostic assessments do not have to be as easy or efficient to administer because they are administered to fewer students, and are often more time intensive due to the need to assess a broader range of skills. The importance of technical characteristics, ease and efficiency of use, and properties of the obtained data are broad areas of consideration when selecting an assessment tool for a predetermined function.

Regardless of the reason for assessment, the user must be informed of the technical adequacy of the assessment in order to interpret and use the data appropriately (Fuchs & Fuchs, 1999). Specifically, the user should be informed of the reliability and validity of an assessment he or she is considering for use. Without reliability, the user
does not know if the results are stable from one measurement to the next (test retest), if two evaluators would have found the same results (inter-rater), or if the items assessed are related to one another (internal consistency) (Sylvia & Ysseldyke, 2004). Adequate demonstration of reliability is necessary, but not sufficient to demonstrate validity, the overarching measure of the technical quality of a given test. Content validity addresses the extent to which a test’s content matches the domain it is intended to assess. Criterion-related validity addresses the relationship between a test score and a desired outcome in the future (predictive) or present (concurrent). Construct Validity considers the degree to which a test is reflective of the underlying theory from which it is based (Sylvia & Ysseldyke, 2004). The reason of use for a given test dictates the extent to which a test’s technical characteristics must adhere to stringent requirements. Using assessment instruments with sufficient levels of reliability and validity is a start toward having meaningful results.

Assessment tools used for progress monitoring require characteristics beyond adequate technical properties. For example, because progress monitoring is usually completed on a weekly or biweekly basis, it is important for such a tool to be quick and easy to administer and score (Fuchs & Fuchs 1999). This will allow the user to complete the whole assessment procedure in a limited amount of time. Additionally, when monitoring progress, the instrument should be sensitive to small academic gains and provide equal scaling of progress (provide equal scaling of data). This will enable the user to document the amount of progress a student has made without skewing the appearance of progress with mathematical influences (Fuchs & Fuchs). For example, it is easy to recognize the amount of growth when considering a child who previously read 25
words correctly per minute now reads 50 words correctly per minute. It is more difficult to determine the amount of progress when the child previously read at the 25th percentile, but now reads at the 50th percentile. An instrument used for progress monitoring should monitor the achievement of the component skills needed to master a domain. For reading, this means the instrument should reflect skills such as phonemic awareness, alphabetic principle, oral reading fluency, vocabulary, and/or comprehension (National Reading Panel, 2000). Although current research on the topic of teacher preferences of progress monitoring tools is limited, one might consider the benefits of using a tool in which teachers value the information gained from the progress monitoring. All of these features coalesce into an appropriate progress monitoring tool.

School systems need an approach to reading assessment for progress monitoring that involves the aforementioned criteria. Brown-Chidsey (2005) and Elliott and Fuchs (1997) provide overviews of some of the disadvantages of the traditional published norm referenced tests used by school psychologists. Traditional, norm referenced tests are not feasible to administer to every student at risk for failure because they are time consuming to administer and interpret. The information provided by those tests is in the form of a score that only means as much as the comparison students with whom it is normed (Elliott & Fuchs, 1997). Because of the need for traditional tests to assess a variety of domains across a wide span of ages, only a few questions are devoted to each level, and they may or may not correspond to the curriculum at one’s school (Brown-Chidsey, 2005). As a result, these tests do not provide adequate instructional information relevant to specific academic skills and do not adequately measure the curricular content taught in many schools (Elliott & Fuchs, 1997).
Teachers use a variety of assessment methods within their classrooms to measure the achievement of their students. Many of these tests are teacher created, and therefore have unknown technical properties, and an unknown relationship to a desired outcome. Some teachers rely on one or more published tests called informal reading inventories (IRIs). Although construction of these tests typically includes some form of review process, the reliability might not be adequate to confidently use the results to make decisions about students (Spector, 2005). Spector reviewed nine recently revised IRIs to determine the overall reliability of the assessment tools. Fewer than half of the reviewed IRIs had published information regarding reliability. Of the tests that did offer reliability information, the information indicates they can be used for lower stakes decisions such as selecting classroom reading material, but not high stakes decisions such as identifying reading difficulties.

The combined limitations of published norm referenced tests and teacher made tests demonstrate the need for an assessment measure that is efficient and easy to administer, reflects classroom content, provides instructionally relevant information, and is sensitive to students’ progress in American elementary schools.

Oral Reading Fluency

The effortless decoding of connected text is a necessary skill for developing reading competence. Assessment of this skill, referred to as Oral Reading Fluency, is typically accomplished by having the child read a passage aloud while the evaluator records the number of words read correctly within one minute. Beyond acknowledgement that ORF is a component skill of reading, measures of ORF have been gaining acceptance as a valid indication of overall reading achievement throughout the past twenty-five
years. A variety of theories have been proposed to describe how ORF accurately measures comprehension. LaBerge and Samuels (1974) propose an automaticity model for the effects of ORF on comprehension. The automaticity model posits that the execution of a complex skill requires the effortless coordination of several lower-level skills. When each lower-level skill can be executed without undue consumption of attentional resources, one’s attention can be utilized in higher-level reasoning (in this case, reading comprehension). Oral reading fluency is a strong component of skillful reading. Stanovich (2000) proposed a model of ORF and comprehension in which each process is interdependent on the other. In other words, higher leveled reasoning facilitates word reading at the same time that word reading facilitates comprehension. Regardless of the model, accurate, fluent reading is an integral component of overall reading competence.

Deno, Mirkin, and Chiang (1982) initially demonstrated the utility of ORF when they validated Curriculum Based Measures (CBM) for use with students with and without learning disabilities. Deno et al. identified five measures of reading (oral reading from passages, isolated word reading, reading in context, cloze passages, and expressive vocabulary) that could be used to monitor students' progress frequently. They found that oral reading fluency was a good predictor of reading performance on popular criterion measures of the time such as the Reading Comprehension task from the Peabody Individual Achievement Test (Dunn & Markwardt, 1970). Martson (1989) provided an overview of early reliability and validity studies of ORF in CBM (referred to as R-CBM) and pointed out that overall test-retest reliability ranged from .82 to .97. Alternate forms
and inter-rater reliability were also high, with most correlation coefficients above .90. This indicates that R-CBM measures are acceptable as an assessment tool.

Oral reading fluency as an indirect measure of comprehension has been compared to other more direct measures of reading comprehension often used by teachers (Fuchs, Fuchs, & Maxwell, 1998). These authors looked at students’ ORF, question answering, story recall, and performance on cloze as alternative ways to assess comprehension, and compared these to student performance on the Reading Comprehension portion of the Stanford Achievement Test (Gardner, Rudman, Karlsen & Merwin, 1982). Results suggest that the correlation between ORF and reading comprehension (.91) was higher than the other three measures (question answering .82, recall .70, and cloze .72). In addition to not having as strong a correlation with reading comprehension, these other measurement techniques had additional drawbacks. Specifically, creating questions about the passage in which the answer was not directly in the reading material, and maintaining objective scoring criterion for story recall proved difficult. Oral reading fluency has been demonstrated to be a reliable and valid measure of overall reading skills, including comprehension (Fuchs, Fuchs, & Maxwell).

Aside from the strong technical properties of R-CBM, this form of assessment has other advantages (Elliott & Fuchs, 1997). The availability of alternate test forms allows the user to assess repeatedly without concern of the students learning the test material through practice. Results of R-CBM are easily graphed, which pictorially displays academic progress for team members. For more in-depth information about a student’s reading skills, qualitative information can be gained from error analysis completed during R-CBM assessment. By doing this, the assessor can determine which basic reading skills
the student possesses, or more importantly, has not learned. For example, while the student is reading aloud, the assessor might determine that the student has mastered alphabetic principle, but has not learned common sound blends. R-CBM procedures are also standardized, which means that while gathering data, teachers subscribe to specific administration rules, which yield quantitative results of reading rate and allows for objective comparisons between students. According to Madelaine and Wheldall (2005), R-CBM is more accurate at identifying low readers than teacher judgment alone. In their study, they found that only 15% of the teachers could accurately identify the lowest three readers in their classroom. Information from the study warns against the sole use of teacher judgments to determine which students are low readers, and suggests that R-CBM may be more objective yet still time efficient. Overall, R-CBM has many characteristics that make it a useful measurement tool in the classroom.

Despite the advantages of R-CBM, there are drawbacks that prevent it from becoming a commonly used method for assessing students' academic progress in schools. R-CBM does not look like it measures anything aside from rapid word calling. In other words, it lacks face validity. Foegen, Espin, Allinder, and Markell (2001) found that teachers generally accepted R-CBM as a measure of reading ability, but did not believe it could assess reading comprehension. Even after being provided with evidence of the validity of oral reading fluency as a measure of overall reading performance, their acceptance of this property was low. Teachers are not as likely to use and value an assessment tool if they do not believe it to measure what they want to measure. Another disadvantage of R-CBM is that the child needs to be able to read connected text in order for their progress to be assessed. Many students entering kindergarten and first grade are
unable to read connected text, yet are not necessarily delayed in reading. However, these are years in which students are expected to make significant reading progress, and in which it is necessary to identify students not able to make these gains without help. While R-CBM has been demonstrated to be reliable and valid indicators of overall reading performance, teachers' acceptability of them is low. Also, their utility for assessing children at school entry is low. To combat the latter drawback, a new set of instruments, which are based on the utility of oral reading fluency, was designed to address the needs of children in kindergarten and first grade.

*Dynamic Indicators of Basic Early Literacy Skills*

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are assessment tools that have developed from Curriculum Based Measures of reading. They are standardized, individually administered, efficient measures of pre-reading and early reading skills (Kaminski & Good, 1996). Probes measuring awareness of initial sounds, and rapid letter naming are given to children in kindergarten to assess prereading skills. This makes the DIBELS measures usable in the assessment of early literacy prior to requiring students to read connected text. Phoneme segmentation and nonsense word fluency can be administered to children acquiring reading skills. DIBELS Oral Reading Fluency was designed to be implemented and used in a manner similar to R-CBM and is administered to children able to read connected text (Kaminski & Good).

DIBELS Oral Reading Fluency tasks were first developed by creating short passages with leveled difficulty. Readability of the passages was estimated primarily using the Spache (1953) formula (Good & Kaminski, 2002). DIBELS Oral Reading Fluency passages were refined until they matched either the end of a given grade or
beginning of the next grade. Passages were then determined to be either a benchmark passage or progress monitoring passage based on readability. The end results yielded one easier, one moderate, and one difficult passage for each of three benchmarks and progress monitoring probes with some variability in passage difficulty. As a result of this process, a student’s increase in benchmark scores should be due to an increase in skill, not a variation of passage difficulty.

Nationally derived DIBELS cut scores were developed based on longitudinal data (Good, Simmons, Kame'enui, Kaminski, & Wallin, 2002). DIBELS subtests were administered to preliterate students multiple times per year, and then to each student one or more years later with a well-established outcome measure. The students’ scores from early DIBELS administration were then analyzed along with the outcome data to determine the predictive ability of DIBELS scores. The DIBELS results from students in third grade have been shown to correlate with proficiency on statewide tests of achievement in Arizona (Wilson, 2005), Colorado (Shaw & Shaw 2002) and North Carolina (Barger, 2003).

Wilson (2005) studied third graders in schools receiving Reading First funding and found a moderately large correlation (r = .74) between DIBELS ORF and the third grade Arizona Instrument to Measure Standards (AIMS). These findings were largely consistent across stratified subgroups (based on English language acquisition, gender, ethnicity, and qualification for free and reduced lunch). The DIBELS were more highly correlated to the outcome measure for those students who were “At Risk” and therefore would not meet proficiency standards on the AIMS than those who were at “Low Risk” and likely to meet proficiency standards.
Shaw and Shaw (2002) conducted a similar study with third grade students in Colorado and their performance on the third grade Colorado State Assessment Program (CSAP). Students' spring ORF scores had a .80 correlation with the CSAP, which was given within the same month. The fall and winter ORF correlations with the CSAP were .93 and .91, respectively.

Barger (2003) compared the third grade students' performance on the North Carolina state test to spring DIBELS performance. Results suggest a high correlation (.73) between the two tests. However, unlike Wilson's (2005) results, stronger correlations were found between the two tests for students who were proficient readers, and weaker correlations for less proficient readers.

Nationwide, studies are emerging that link proficiency on the DIBELS to standards on statewide outcome measures of reading. Although the DIBELS measures are relatively new to the field of reading assessment, they have demonstrated adequate technical characteristics to be used as an indicator of overall reading achievement. Furthermore, the acquisition of equal interval data, utility of use with preliterate children, and availability of multiple equivalent forms makes it ideal for monitoring progress of early reading development.

Developmental Reading Assessment

The Developmental Reading Assessment (DRA K-3) is an individually administered test of reading interest, accuracy, and comprehension for students in kindergarten through third grade. Its publisher, the Pearson Learning Group (2001), reports that the DRA results inform instructional interventions, that the DRA can be used on an annual or semiannual basis to monitor student reading development, and it can be
administered more frequently with struggling readers to ensure continued growth. When considering the four purposes of assessment as discussed earlier, this suggests the DRA can function as both a diagnostic and progress monitoring tool (Pearson Learning Group, 2001). These statements have implications that affect the necessary characteristics of the DRA K-3. To inform instruction, it should have high levels of technical adequacy, and should measure skills that are influenced by instruction. To monitor progress, it should be sensitive to changes in student progress, easy and efficient to administer, and have alternative test forms.

The DRA is individually administered in a location separate from other students. The teacher utilizes this time to question the student regarding his or her reading habits, such as whether he or she reads at home, with whom he or she reads, and favorite books. The teacher selects three or four texts of varying 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. The level of text difficulty is the “Story Level.” As the student reads aloud, the teacher makes written comments about various aspects of the student’s reading. The teacher is allowed to ask questions to gather as much information as possible about the student’s reading (Pearson Learning Group, 2001). This entire process is referred to as the DRA conference.

The DRA purports to measure reading fluency and reading comprehension (Pearson Learning Group, 2001). While conducting the DRA conference, the teacher describes the type of phrasing the student uses during oral reading, by including notation regarding how often the student stops while reading a sentence, and whether the words are connected or read one by one. The student’s use of intonation and expression, and
adherence to grammatical rules are also descriptors the teacher uses to evaluate the student’s reading. While the student reads, the teacher notes the student’s problem solving strategies, such as the number of times the student asks for help, the number of words the teacher provides to the student, the student’s use of illustrations to determine meaning, and multiple attempts to decode a word (Pearson Learning Group, 2001). Although at face value, these observations seem to measure reading competence, empirical research is necessary to determine if the observations are reliable, and if they have implications for overall reading competence.

Comprehension is measured by the student’s ability to retell the text in a way that indicates understanding of the main ideas, key facts, characters and topics. The teacher is allowed to ask questions of the student in order to gain more information, if the teacher thinks the student knows more than he or she is providing (Pearson Learning Group, 2001). The teacher then rates each of the above areas on a rubric that is divided into the following categories: “Very Little Comprehension”, “Some Comprehension”, “Adequate Comprehension”, and “Very Good Comprehension.” Each descriptor has a corresponding number to transform the qualitative rating into a quantitative rating. For example, children’s telling might indicate “Very Good Comprehension,” worth four points, of the key facts, and indicate “Very Little Comprehension,” worth one point, of the main idea. The total points of all domains are added to obtain the Comprehension Level.

The technical characteristics of the DRA K-3 have been examined by persons or groups affiliated with the DRA publisher, Pearson Learning Inc. Weber (2000, as cited by Pearson Learning Group, 2001) conducted a study that examined the test-retest reliability of the DRA K-3. Weber analyzed the Story Level at which each participant
read independently, and only considered data from students in first through third grades. Results found test-retest correlations that ranged from .92 for first graders to .99 for third graders. Williams (1999) looked into the inter-rater reliability of the DRA K-3 among 87 teachers trained in the administration of the DRA. Results suggest that agreement between the first two raters was high (.80), but lower when accounting for the third teacher (.74). Weber (as cited by Pearson Learning Group, 2001) also looked into the accuracy of teacher judgments compared to an expert rater, showing high agreement among teachers' scoring for accuracy, phrasing and fluency, but not as high for comprehension.

The content validity, criterion related validity, and construct validity of the DRA K-3 have also been studied. Content validity was built into the DRA through the numerous stages of field-testing during its development (Pearson Learning Group, 2001). Through this process, 127 elementary school teachers from 11 states were asked for their feedback regarding whether the content of the tool accurately reflected classroom materials. The DRA was reconstructed according to teacher suggestions. During field-testing, 52% of sampled educators rated results of the DRA as having a strong relationship to classroom instruction, and an additional 40% indicated that the DRA results had a moderate relationship to classroom instruction (Pearson Learning Group). However, the article does not describe how the sample of teachers was obtained. Therefore, the extent to which this sample is representative of the population of teachers is not known.

Weber (as cited by Pearson Learning Group, 2001) looked at the criterion related convergent validity of the DRA with scores on the Reading Comprehension portion of
the Iowa Tests of Basic Skills (ITBS) (Hoover, Dunbar, & Frisbie, 2001) for first through third grade students. Weber found that performance on the DRA was moderately indicative of the performance on reading comprehension of the ITBS (correlations of .65, .84, and .54 for first through third grade respectively).

Williams (1999) examined the construct validity of the DRA compared to the reading comprehension portion of the ITBS (Hoover, et al. 2001). Williams compared the results on both measures of 2,470 students and found a statistically significant correlation between the DRA instructional Story Level and ITBS Total Reading level (.71), Reading Comprehension measure (.68), and Vocabulary Measure (.68).

Although promising, the results of these research reports need to be considered in relationship to their sources. The studies were each conducted by someone affiliated with Pearson Inc, or with MetaMetrics, a company hired to analyze statistical data. Additionally, Williams (1999) is the only study that is viewable in its original form. Information from the other study is only available as a summary of the original work printed by the Pearson Learning Group in the DRA Technical Manual. Further objective research is needed on the DRA K-3 to determine the reliability of the above findings with independent researchers.

*Terra Nova 2*

The TN2 is a group administered, norm referenced, standardized test of student achievement. Previous versions of this test have been used in educational settings since 1950 (CTB McGraw Hill, 2006). Two of the tests within the assessment series, Vocabulary and Comprehension, were used as criterion standards for which to compare the results of the DIBELS Oral Reading Fluency and DRA tasks.
No reading assessment measure is perfect for every assessment purpose. One assessment tool that may measure achievement accurately and reliably, may be too time consuming to administer. Likewise, the primary user may not accept a measure that is easy to administer and interpret based on his or her biases. A final assessment measure may be accepted by the user, but not assess important characteristics accurately or reliably.

The present study seeks to determine the relationships between the DRA and DIBELS ORF, and how the results of these two assessments relate to TN2 performance. Using what has been presented regarding the three assessment measures under investigation, the following hypothesis are tested with this study. Significant correlations will be found between the fall, winter, and spring DRA Story Level, Comprehension Level, and Accuracy Level with the fall, winter, and spring DIBELS Oral Reading Fluency benchmark scores. Significant correlations will be found between the fall, winter, and spring DIBELS ORF scores and the TN2 Comprehension test. Significant correlations will be found between the spring DRA Story Level and the TN2 Vocabulary test. Significant predictor variables for the TN2 Comprehension task will include the DIBELS ORF and DRA Comprehension ratings. The strongest predictor variable for the TN2 Vocabulary task will be the DRA Story Level.
CHAPTER III

METHOD

Participants

The author analyzed data from second grade students attending an urban elementary school receiving Reading First grant assistance. DRA data was available for 67 students. Sixty of those students also had TN2 and DIBELS scores. Descriptive data regarding the ethnicity and socioeconomic status of the participants was not collected. However, according to the New York State Education Department (2006), 84% of the students in the school are African American or Hispanic (compared to the state average of 40%). Furthermore, 9% of students in the school have limited English proficiency (compared to the state average of 7%). Ninety percent of students in the school are eligible for free or reduced lunch (compared to the state average of 28%) (New York State Education Department).

Exclusionary criteria. All second grade students in the school described above were considered for participation in this study. Each student was administered the reading instruments investigated within this study as a part of the normal school assessment process. Participants’ data were included in this study if the student was in second grade during the assessment, and if the student was assessed with both the DIBELS ORF and DRA K-3 during the at least two of the three test periods (ex. fall and winter, fall and spring, or winter and spring), and with the TN2 in the spring. DRA data was excluded unless the protocol included information regarding the student’s grade, the date of assessment, and scoring sections from the phrasing and fluency, intonation, accuracy, and comprehension were included.
Of the 60 students who met the aforementioned exclusionary criteria, 44 had been assessed with the DRA more than one time during one or more of the three assessment periods. For example, a student may have been administered four different DRA 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 after all student data had been entered.

Qualifications of Assessors

The school based assessment teams were trained in a one-day workshop in DIBELS in the spring of 2004. Follow up sessions occurred periodically to ensure sufficient training and understanding of that test. The extent to which the assessors were trained in the administration of the DRA is not known.

The researcher is a graduate student with training in research methodology, statistical applications and reading assessment. She is trained in the handling of data with specific consideration towards the coding of the data.

Assumptions

Several assumptions are considered in regards to data collection. The present author assumes the teachers administered the tests according to standardized procedures. Further, although the present author checked to ensure accurate scoring computations for the DRA Comprehension section, and transferring of scores within the DRA protocol, it is assumed the teachers scored the students’ responses reliably. The students’ DIBELS and TN2 test protocols were unavailable for analysis. As a result, it is unknown the
extent to which the responses were scored accurately, and whether the DIBELS and TN2 scores were accurately entered into the database used by the district to manage the data.

**Materials**

*Developmental Reading Assessment, Kindergarten through Third Grade.* This assessment was administered three times per year (fall, winter, and spring) in the school from which data was obtained. The assessment requires students to read developmentally appropriate stories aloud to the teacher. The teacher then rates the child’s reading by considering the various attributes as previously discussed. Teachers are not limited to standardized administration rules and are encouraged to probe the student depending on the teacher’s perception of the student’s reading ability.

*Coding sheet.* Prior to coding the DRA data for each of the five categories, the test administrator had already rated the students’ performance with qualitative descriptors of each of the measured constructs, which were listed in increasing order of proficiency. The researcher transformed qualitative ratings into ordinal data using the number one to refer to the lowest proficiency, and successively higher numbers to correspond with higher levels of proficiency. This author created a coding sheet specifically for this research project. It consisted of a section for identifying information, phrasing and fluency, intonation, accuracy, and comprehension. Refer to the coding sheet attached in the Appendix as an example.

*Dynamic Indicators of Basic Early Literacy Skills.* The DIBELS are a set of standardized, individually administered tests of early reading achievement that are administered three times per year (fall, winter, and spring) or more. The specific task used for this study, DIBELS Oral Reading Fluency (ORF), measures the student’s
accuracy and fluency with connected text. The students were required to read a passage for one minute while the teacher documented the number of words read correctly (WRC). The number of WRC is considered the student’s oral reading fluency rate.

*Terra Nova 2nd Edition.* According to the publisher (CTB McGraw Hill, 2006), the content of the TN2 is reflective of curriculum guides from individual states, National ELA standards, and conceptual frameworks from the National Assessment of Educational Progress (NAEP). Significant consideration has gone into the development and construction of this test.

Research for the development followed with two phases; a tryout of potential items and then standardization of the chosen items. During the try out phase, all of the items being considered for the TN2 were administered to 100,000 students (Terra Nova 2nd Edition, 2001). Items thought to represent content for a specific grade were administered to students within that hypothesized grade, and to students one grade above and below that grade. Students’ gender and ethnicity were considered according to their responses to identify potentially biased items. Item response theory models were used for both the selected response (Three parameter Logistic Model) items and constructed response items (Two parameter Partial Credit Model).

After the process of item selection was completed, standardization procedures were employed (Terra Nova 2nd Edition, 2001). A stratified random sampling procedure using 275,000 students was used to obtain a sample of students that represents the nation’s school aged population. After standardization, test developers used the “Bookmarking Procedure” (Lewis, Green, Mitzel, Baum & Patz, 1998; Lewis, Mitzel, & Green, 1996) to allow teachers to gain information regarding a student’s achievement
compared to expected performance levels. As a result of the development procedures, teachers can use normative results, criterion referenced results, and performance level scores when evaluating a child’s academic progress.

*Procedures*

The researcher obtained the DRA protocols from the school’s district office. Copies of the original protocols were created for the researcher’s use, each student was assigned a number, and all identifying information was blacked out. Student data was then excluded based upon the aforementioned criteria. DRA codes were then entered into an SPSS database, along with the DIBELS and TN2 data, which was provided by the school district in a Microsoft Excel file. This data was analyzed using Pearson correlations and Stepwise regression analyses.
CHAPTER IV

RESULTS

Results were analyzed to investigate the relationships among the three reading tests under investigation. Descriptive statistics for the DIBELS, DRA, and TN2 are presented in Tables 1, 2, and 3 respectively. Correlation and regression analyses were then run to determine the extent to which the three tests related to each other. The first set of analyses considers how the fall, winter, and spring DIBELS results correlate with fall, winter, and spring DRA results. Furthermore, the first set of analyses examines the degree to which different portions of the fall DRA scores predict winter DIBELS performance, and to what degree the fall and winter DRA data predict spring DIBELS performance. The second set of analyses considers how the spring DIBELS data correlates with the TN2 scores. The third set of analyses considers how the spring DRA data correlates with the TN2 results. The final analyses evaluate the extent to which fall and winter DRA and DIBELS data predicts performance on the TN2 Comprehension and Vocabulary measures.
Table 1

*Descriptive Statistics for the DIBELS Oral Reading Fluency*

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>32.92</td>
<td>19.79</td>
<td>59</td>
</tr>
<tr>
<td>Winter</td>
<td>53.79</td>
<td>25.29</td>
<td>58</td>
</tr>
<tr>
<td>Spring</td>
<td>73.08</td>
<td>30.23</td>
<td>60</td>
</tr>
</tbody>
</table>
Table 2

Descriptive Statistics for the Developmental Reading Assessment

<table>
<thead>
<tr>
<th>DRA Measure</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fall (n=54)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phrasing and Fluency</td>
<td>2.31</td>
<td>1.23</td>
</tr>
<tr>
<td>Intonation</td>
<td>2.20</td>
<td>0.96</td>
</tr>
<tr>
<td>Accuracy</td>
<td>96.06</td>
<td>2.61</td>
</tr>
<tr>
<td>Comprehension Level</td>
<td>15.19</td>
<td>3.62</td>
</tr>
<tr>
<td>Story Level</td>
<td>12.35</td>
<td>7.11</td>
</tr>
<tr>
<td><strong>Winter (n=48)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phrasing and Fluency</td>
<td>2.56</td>
<td>1.23</td>
</tr>
<tr>
<td>Intonation</td>
<td>2.61</td>
<td>0.90</td>
</tr>
<tr>
<td>Accuracy</td>
<td>96.78</td>
<td>3.31</td>
</tr>
<tr>
<td>Comprehension Level</td>
<td>14.94</td>
<td>3.44</td>
</tr>
<tr>
<td>Story Level</td>
<td>17.52</td>
<td>9.41</td>
</tr>
<tr>
<td><strong>Spring (n=55)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phrasing and Fluency</td>
<td>3.67</td>
<td>1.37</td>
</tr>
<tr>
<td>Intonation</td>
<td>3.11</td>
<td>1.07</td>
</tr>
<tr>
<td>Accuracy</td>
<td>96.07</td>
<td>2.79</td>
</tr>
<tr>
<td>Comprehension Level</td>
<td>16.13</td>
<td>3.59</td>
</tr>
<tr>
<td>Story Level</td>
<td>24.04</td>
<td>11.00</td>
</tr>
</tbody>
</table>
Table 3

*Descriptive Statistics for the TN2- Second Edition (n=60)*

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN2 Vocabulary</td>
<td>581.23</td>
<td>37.07</td>
</tr>
<tr>
<td>TN2 Comprehension</td>
<td>598.23</td>
<td>27.44</td>
</tr>
</tbody>
</table>
To evaluate the extent to which the DIBELS results relate to the DRA results, Pearson correlations were calculated. 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 the fall DRA Phrasing and Fluency ratings, Accuracy levels, and Story Level. In the winter, DIBELS scores were significantly correlated all winter DRA scores. In the spring, DIBELS scores were significantly correlated at the .01 level with the spring DRA Phrasing and Fluency ratings, Intonation ratings, and Story Level.
Table 4

**Correlations between the DIBELS and the Developmental Reading Assessment.**

<table>
<thead>
<tr>
<th>Subtest</th>
<th>DIBELS Fall n=54</th>
<th>DIBELS Winter n=52</th>
<th>DIBELS Spring n=54</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall Phrasing and Fluency</td>
<td>.459 (p=.000)*</td>
<td>.409 (p=.003)</td>
<td>.436 (p=.001)</td>
</tr>
<tr>
<td>Intonation</td>
<td>.164 (p=.237)</td>
<td>.269 (p=.053)</td>
<td>.328 (p=.016)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.353 (p=.009)*</td>
<td>.259 (p=.064)</td>
<td>.151 (p=.275)</td>
</tr>
<tr>
<td>Comprehension Level</td>
<td>.253 (p=.065)</td>
<td>.195 (p=.167)</td>
<td>.195 (p=.158)</td>
</tr>
<tr>
<td>Story Level</td>
<td>.684 (p=.000)*</td>
<td>.727 (p=.000)</td>
<td>.710 (p=.000)</td>
</tr>
</tbody>
</table>

| Winter Phrasing and Fluency | .498 (p=.000)   | .586 (p=.000)*   | .610 (p=.000)     |
| Intonation                | .398 (p=.006)   | .483 (p=.001)*   | .551 (p=.000)     |
| Accuracy                  | .513 (p=.000)   | .649 (p=.000)*   | .616 (p=.000)     |
| Comprehension Level       | .293 (p=.045)   | .457 (p=.001)*   | .445 (p=.002)     |
| Story Level               | .760 (p=.000)   | .812 (p=.000)*   | .766 (p=.000)     |

| Spring Phrasing and Fluency | .333 (p=.014)   | .373 (p=.006)   | .347 (p=.009)*    |
| Intonation                 | .223 (p=.106)   | .296 (p=.032)   | .347 (p=.009)*    |
| Accuracy                   | .212 (p=.124)   | .321 (p=.019)   | .244 (p=.073)     |
| Comprehension Level        | .289 (p=.034)   | -.129 (p=.359)  | -.136 (p=.323)    |
| Story Level                | .853 (p=.000)   | .900 (p=.000)   | .868 (p=.000)*    |

*Note.* Significance levels marked with an asterisk were significant at the .01 level and were considered in this study.
After analyzing the correlation relationships, a stepwise multiple regression analysis was completed to determine possible predictor variables among the administered tests. 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 0.1.

A stepwise multiple regression analysis was done to determine the extent to which the different components of the fall DRA predicted winter DIBELS scores, and components of the fall and winter DRA predicted spring DIBELS scores. Tables 5 and 6 provide 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 ORF winter and spring scores. The fall DRA Story Level was the only significant predictor of winter DIBELS ORF. It uniquely accounted for 52.85% of the variance in the DIBELS winter score. The winter DRA Story Level accounted for the most variance (24%) in the spring DIBELS ORF. Other significant predictors to the spring DIBELS score include the winter Accuracy level (9.12%), and fall and winter Intonation ratings (4.16% and 3.92% respectively).
Table 5

Summary of Stepwise Regression Analysis for Variables Predicting Winter DIBELS Scores (n=52)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE_B$</th>
<th>$\beta$</th>
<th>Squared Semi-Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1  DRA Fall Story Level</td>
<td>2.396</td>
<td>.320</td>
<td>.727</td>
<td>.5285</td>
</tr>
</tbody>
</table>
Table 6

Summary of Stepwise Regression Analysis for Variables Predicting Spring DIBELS Scores (n=41)

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>Squared Semi-Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRA Winter Story Level</td>
<td>2.336</td>
<td>.299</td>
<td>.781</td>
<td>.6099</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRA Winter Story Level</td>
<td>2.223</td>
<td>.281</td>
<td>.743</td>
<td>.540</td>
</tr>
<tr>
<td>DRA Winter Score Accuracy</td>
<td>2.665</td>
<td>1.001</td>
<td>.250</td>
<td>.0615</td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRA Winter Story Level</td>
<td>2.054</td>
<td>.276</td>
<td>.687</td>
<td>.4290</td>
</tr>
<tr>
<td>DRA Winter Accuracy</td>
<td>3.280</td>
<td>.983</td>
<td>.308</td>
<td>.0864</td>
</tr>
<tr>
<td>DRA Fall Intonation</td>
<td>6.127</td>
<td>2.623</td>
<td>.219</td>
<td>.0420</td>
</tr>
<tr>
<td>Step 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRA Winter Story Level</td>
<td>1.737</td>
<td>.291</td>
<td>.581</td>
<td>.2440</td>
</tr>
<tr>
<td>DRA Winter Accuracy</td>
<td>3.383</td>
<td>.927</td>
<td>.318</td>
<td>.0912</td>
</tr>
<tr>
<td>DRA Fall Intonation</td>
<td>6.084</td>
<td>2.471</td>
<td>.217</td>
<td>.0416</td>
</tr>
<tr>
<td>DRA Winter Intonation</td>
<td>7.710</td>
<td>3.225</td>
<td>.224</td>
<td>.0392</td>
</tr>
</tbody>
</table>
The second set of analyses considers the extent to which the DIBELS results correlate to TN2 results. Table 7 provides the correlations between the results of the DIBELS and the TN2. At the .01 level, the results from the fall, winter, and spring DIBELS scores, all correlated significantly with the results from the spring TN2 Comprehension and Vocabulary measures.
Table 7

Correlations between DIBELS Oral Reading Fluency and TN Second Edition (n=54)

<table>
<thead>
<tr>
<th>Test</th>
<th>DIBELS Fall</th>
<th>DIBELS Winter</th>
<th>DIBELS Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN2 Vocabulary</td>
<td>.546 (p=.000)*</td>
<td>.653 (p=.000)*</td>
<td>.718 (p=.000)*</td>
</tr>
<tr>
<td>TN2 Comprehension</td>
<td>.636 (p=.000)*</td>
<td>.714 (p=.000)*</td>
<td>.709 (p=.000)*</td>
</tr>
</tbody>
</table>

Note. Significance levels marks with an asterisk were significant at the .01 level and were considered in this study.
The third set of analyses considers the extent to which the DRA results correlate to TN2 results. Table 8 provides the correlations between the results of the DRA and the TN2. At the .01 level, the results from the fall, winter, and spring DRA Story Level all correlate significantly with the results from both the spring TN2 Comprehension and Vocabulary measure. The strongest correlations were found between both TN2 scores (spring) and the spring DRA Story Level.
**Table 8**

*Correlations between the TN2 and Developmental Reading Assessment*

<table>
<thead>
<tr>
<th>DRA Subtest</th>
<th>TN2 Vocabulary</th>
<th>TN2 Comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fall (n=54)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phrasing and Fluency</td>
<td>.331 (p=.015)</td>
<td>.315 (p=.020)</td>
</tr>
<tr>
<td>Intonation</td>
<td>.292 (p=.032)</td>
<td>.262 (p=.056)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.083 (p=.553)</td>
<td>.046 (p=.743)</td>
</tr>
<tr>
<td>Comprehension Level</td>
<td>.304 (p=.013)</td>
<td>.341 (p=.006)</td>
</tr>
<tr>
<td>Story Level</td>
<td>.487 (p=.000)*</td>
<td>.569 (p=.000)*</td>
</tr>
<tr>
<td><strong>Winter (n=48)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phrasing and Fluency</td>
<td>.531 (p=.000)</td>
<td>.571 (p=.000)</td>
</tr>
<tr>
<td>Intonation</td>
<td>.518 (p=.000)</td>
<td>.497 (p=.000)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.608 (p=.000)</td>
<td>.649 (p=.000)</td>
</tr>
<tr>
<td>Comprehension Level</td>
<td>.501 (p=.000)</td>
<td>.579 (p=.000)</td>
</tr>
<tr>
<td>Story Level</td>
<td>.564 (p=.000)*</td>
<td>.542 (p=.000)*</td>
</tr>
<tr>
<td><strong>Spring (n=55)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phrasing and Fluency</td>
<td>.148 (p=.280)</td>
<td>.276 (p=.041)</td>
</tr>
<tr>
<td>Intonation</td>
<td>.163 (p=.234)</td>
<td>.260 (p=.055)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.151 (p=.271)</td>
<td>.118 (p=.390)</td>
</tr>
<tr>
<td>Comprehension Level</td>
<td>-.131 (p=.339)</td>
<td>.010 (p=.941)</td>
</tr>
<tr>
<td>Story Level</td>
<td>.685 (p=.000)*</td>
<td>.700 (p=.000)*</td>
</tr>
</tbody>
</table>

*Note.* Significance levels marked with an asterisk were significant at the .01 level and were considered in this study.
Regression analyses were then run on the TN2 Vocabulary and Comprehension scores to determine which fall and winter DRA and DIBELS assessments best predicted TN2 performance. Table 9 provides evidence of predictors for the TN2 Comprehension measure. Results indicate the winter DIBELS score is the only significant predictor of spring TN2 Comprehension performance. Table 10 provides evidence of predictors for the TN2 Vocabulary measure. Results suggest the fall DRA Comprehension Level, winter DRA Phrasing and Fluency ratings, and fall DRA Story Level are each significant predictors of performance on the TN2 Vocabulary measure. Although they are significant, they contribute minimally to the prediction of TN2 Vocabulary scores.
Table 9

*Summary of Stepwise Regression Analysis for Variables Predicting TN2 Comprehension*  
(n=39)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SEB$</th>
<th>$\beta$</th>
<th>Squared Semi-Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIBELS Winter Score</td>
<td>.545</td>
<td>.124</td>
<td>.586</td>
<td>.3433</td>
</tr>
</tbody>
</table>
Table 10

Summary of Stepwise Regression Analysis for Variables Predicting TN2 Vocabulary (n=39)

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>Squared Semi-Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRA Fall Story Level</td>
<td>1.851</td>
<td>.555</td>
<td>.481</td>
<td>.2313</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRA Fall Story Level</td>
<td>1.592</td>
<td>.536</td>
<td>.413</td>
<td>.1632</td>
</tr>
<tr>
<td>DRA Fall Comp Level</td>
<td>2.328</td>
<td>.993</td>
<td>.326</td>
<td>.1017</td>
</tr>
<tr>
<td><strong>Step 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRA Fall Story Level</td>
<td>1.070</td>
<td>.568</td>
<td>.278</td>
<td>.0600</td>
</tr>
<tr>
<td>DRA Fall Comp Score</td>
<td>2.431</td>
<td>.950</td>
<td>.340</td>
<td>.1109</td>
</tr>
<tr>
<td>DRA Winter P and F</td>
<td>6.733</td>
<td>3.188</td>
<td>.305</td>
<td>.0756</td>
</tr>
</tbody>
</table>
CHAPTER V
DISCUSSION

This research was done to investigate the relationships among three reading assessment tools; the Developmental Reading Assessment, Dynamic Indicators of Basic Early Literacy Skills Oral Reading Fluency, and the TerraNova2. Specifically, this research sought to determine the relationships between the three instruments administered in a concurrent time frame, and the extent to which the DRA and DIBELS results related to TN2 performance. In the plethora of previous research that examines reading assessment, no known studies had previously examined the relationships among these three instruments. The present study reflects an examination of these relationships for second grade students in an urban school district. Results provided information that was inconsistent with the hypotheses.

Prior to gathering data, this researcher hypothesized that the fall, winter, and spring DRA Story Level, Comprehension Level and Accuracy Level would each significantly correlate with the fall, winter, and spring DIBELS Oral Reading Fluency benchmarks. This hypothesis was made because the DRA story level appears to be an overall predictor of reading achievement, previous research suggests the DIBELS ORF is an indirect measure of comprehension, and the accuracy portion of the DRA is visually similar to the DIBELS.

Relationships between the DRA and DIBELS. Results suggest a significant correlation between the fall DIBELS and fall DRA Phrasing and Fluency ratings, Accuracy levels, and Story Levels. Surprisingly, the Phrasing and Fluency ratings were significantly correlated, but the Comprehension Levels were not. This may be because as
a student becomes a more proficient reader, his or her ability to phrase written language likely increases regardless of text difficulty, whereas comprehension might be more likely to vary with increasing text difficulty. In addition, when examining the data closely, scores from “Comprehension Level” remained relatively constant across various text difficulties. This result also might point to a lack of reliability in judging comprehension. Since the DRA text level is chosen by the teacher, it is likely the teacher chose a level he or she assumed would fall at the child’s instructional level in order to give the best diagnostic information.

Winter DIBELS scores significantly correlated with the winter DRA Accuracy level, Intonation rating, Phrasing and Fluency rating, Comprehension Level, and Story Level. It is reasonable that Intonation ratings offered significant correlations in the same manner from which Phrasing and Fluency rating have been described. A child’s ability to use intonation may improve regardless of text difficulty. An explanation for why Comprehension Level offered a significant correlation is not entirely clear. It is possible this is a product of teacher subjectivity when scoring the DRA. For example, because teachers are encouraged to question students more or less depending on the perception of a given child, the teacher might question high performing readers more, and therefore rate higher levels of comprehension, and question low performing readers less, and therefore rate lower levels of comprehension.

Spring DIBELS ORF scores correlated significantly with Spring DRA Intonation, Phrasing and Fluency and Story Level. Similar to results of other correlations, it is likely that a student’s ability to use intonation and appropriate phrasing will improve as the
child becomes a more proficient reader. Furthermore, the child’s teacher would be more likely to select a more difficult text, or higher story level for a more proficient reader.

A stepwise multiple regression analysis examined fall DRA data to determine predictors of the winter DIBELS score, and fall and winter DRA data to determine predictors of the spring DIBELS score. Results suggest the fall DRA Story Level is the only significant predictor, accounting for 52.85% of the variance of the winter DIBELS score. Results of the second regression analysis were different. When all fall and winter DRA data were analyzed to determine which variables predicted the spring DIBELS score, the winter DRA Story level accounted for the most variance (24.40%). However, other variables also offered significant predictions (winter Accuracy level 9.12%, winter Intonation rating 3.92%, and fall Intonation rating 4.16%). When this information is combined, the results suggest that of the DRA sub measures, Story Level is the strongest and most consistent factor in finding relationships and making predictions with the DRA.

The extent to which the predictor variables are related is not known. As a result, the variables that contribute minimally to prediction may be a function of error.

*Relationships between the DIBELS and TN2.* Prior to conducting the analysis, this researcher hypothesized that spring DIBELS scores would significantly correlate with TN2 Comprehension scores, but not with TN2 Vocabulary scores. This was based on previous research that has suggested the TN2 Comprehension and DIBELS ORF are both valid measures of comprehension. However, the DIBELS are not known to be reflective of vocabulary development, as is reportedly measured by the TN2 Vocabulary test. In partial confirmation of the hypothesis, results suggested the fall, winter, and spring DIBELS scores correlated significantly with both the TN2 Vocabulary and
Comprehension tests. These results offer supporting evidence to the existing literature that suggests the DIBELS is an indicator of comprehension. Additionally, either the DIBELS have some correlation with vocabulary development, or the TN2 Vocabulary measure requires overall reading ability rather than strictly assessing vocabulary.

*Relationships between the DRA and TN2.* Prior to analyzing the data, it was hypothesized that the spring DRA Comprehension Level would correlate with the TN2 Comprehension test. This was hypothesized based on the assumption that “Comprehension Level” was an indicator of overall reading comprehension. This hypothesis is not supported through this analysis. However, the results suggest significant correlations between the TN2 Comprehension measure and the spring DRA Story Level. Furthermore, results suggest significant correlations between the TN2 Vocabulary measure and spring DRA Story Level. Although not the target of this investigation, the fall and winter DRA Story Levels were also significantly correlated with both the TN2 Comprehension and Vocabulary measures. This may offer support to the explanation that the TN2 Vocabulary test requires overall reading ability. No significant correlations were found between the TN2 Vocabulary and Comprehension scores and other spring DRA scores. As discussed before, analysis of Comprehension Level scores suggests they did not fluctuate significantly with increasing or decreasing difficulty of texts. If the teacher’s choice of texts for assessments were based on perceptions of a given child’s instructional reading level, ratings of Comprehension scores would remain relatively stable.

*Predictors of TN2 performance.* All fall and winter DRA and DIBELS scores were analyzed to determine which variables best predicted performance on the TN2
Comprehension and Vocabulary measures. Results of the analysis suggest the winter DIBELS score is the best predictor of performance on the TN2 Comprehension measure. It accounts for 34.33% of the variance in TN2 Comprehension scores. This is not surprising considering the established relationship between DIBELS and reading comprehension. Several variables emerged as significant, albeit weak, predictors of TN2 Vocabulary scores. The fall DRA Comprehension Level uniquely contributed 11.09% of the variance, the winter DRA Phrasing and Fluency accounted for 7.56% of the variance, and the fall DRA Story Level accounted for 6.00% of the variance. However, none of these variables were significant at the designated level (.01). Because it is not known whether these predictor variables are inter-correlated, and in consideration of the minimal levels of prediction, the results of the TN2 Vocabulary regression analysis might be a function of error. This may suggest that neither the DRA nor DIBELS measure the same constructs as the TN2 Vocabulary test and provides evidence of discriminant validity for both the DIBELS and DRA.

Implications for Theory and Practice

Due to the dearth of empirical, objective research on the DRA, hypotheses regarding how it functions to measure reading acquisition in young children are highly tentative. One must consider the findings of this research compared to the findings of Madelaine and Wheldall (2005). That research suggested that only 15% of teachers could accurately identify the three lowest readers in his classroom. Within that research, the use of R-CBM is more accurate at identifying low readers than through the sole use of teacher judgment. The results of the current research partially support the information provided by Madelaine and Wheldall. In the present study, the measure of oral reading
fluency had a stronger relationship with an outcome measure of reading comprehension than the measure that relies more on teacher judgment. However, this relationship did not stand when comparing the measure of ORF with an outcome measure of vocabulary development. Within the current study, the DRA sub measure that most consistently offered valuable information was that which is tentatively based on teacher judgment, the DRA Story Level.

These conclusions offer some considerations for those who require an understanding of various reading assessments within a school district. Namely, if trying to predict performance on a standardized reading comprehension test for second grade students, using DIBELS is likely more effective than using any component of the DRA. For a district already using the DRA, the sub component that offers the most information regarding a second grader’s reading ability is the student’s “Story Level.” All other scores offer information that is partially dependent on the Story Level. Neither the DRA nor DIBELS tests are adequate if trying to predict second grade students’ performance on a standardized test of Vocabulary. Additional assessments should be considered for use if trying to predict Vocabulary performance.

**Limitations.** There are several limitations to this study that affect the confidence one has in the outcomes. Primarily this was historical data. As a result, the data collection process, the extent to which data collectors were trained to administer the DRA, and how the data was originally placed into an electronic database are all unknown. Regarding 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’ decisions regarding DRA text selection, and could possibly influence the teachers’ scoring of the DRA. For example, a teacher who had previously done DIBELS with a student might select a more appropriate text level for a student, or might score a student’s response differently based on the knowledge gained from administering DIBELS. Although these issues raise concerns regarding the internal validity of this study, this is likely a true reflection of how information is obtained and used within a large school system.

A second limitation relates to the amount of data that had to be excluded by meeting exclusionary criteria. 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 student within each assessment period. For example, many students took the DRA two or more times on a single day, yet only one test protocol was analyzed.

A third 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, 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 correlation analyses. As a result, the extent to which the results of correlation analyses would be replicated in future studies is not known. Another consideration related to the analysis of data is in regards to the use of the Stepwise multiple regression procedure. The relationships between the predictor variables entered into the stepwise equation are 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 related to the criterion based on other evidence.

A final limitation considers the manner in which test protocols were handled. Protocols were coded according to objective criteria; however, no interrater reliability measures were used to ensure reliable coding. Additionally, after coding all protocols, no interrater reliability measures were used to ensure accurate transference of data into electronic format.

*Directions for Future Research*

These limitations present issues that need to be addressed in future research. Specifically, more controls on data collection and treatment will increase the internal validity of the results. Furthermore, questions addressed in this study should be addressed with current state tests of academic standards. Within the results of this study, the relationship among the DRA Phrasing and Fluency, Intonation, and Accuracy needs to be investigated to determine the potentiality of them all measuring a single factor. Beyond this investigation, the constructs of the TN2 Vocabulary test should be considered. This could yield potential benefits in developing a tool that predicts vocabulary development, an essential component of successful reading. Finally, the current research did not differentiate between the poor and proficient readers in the sample. A future study could differentiate between these two types of readers to better determine whether reliable and valid assessment of both types of readers requires different assessment methods.
References


Kame’enui, E.J., Francis, D., Fuchs, L., Good, R., O’Connor, R., Simmons, D. et al. (2002). Secretary’s Leadership Academy, Reading First Initiative, Assessment Committee Presentation. US Department of Education: Washington, DC.


Shaywitz, S.E., Fletcher, J.M., Holahan, J.M., Schneider, A.E., Marchione, K.E.,
longitudinal study at adolescence. American Academy of Pediatrics, 104, 6,
1351-1359
Spache, G. (1953). A new readability formula for primary-grade reading material, The
Elementary School Journal, 55, 410-413.
Schools, 42(6), 593-602.
Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual
Stanovich, K.E. (2000). Process in understanding reading: Scientific foundations and
new frontiers. New York: Guilford
CTB/McGraw-Hill.
US Department of Education. (2005) Reading first: Over $4.3 billion to improve the
reading skills of young children. Retrieved 15 October 2005 from

Appendix

Coding Sheet 2nd Grade

Student's ID #:

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

Grade Level:
Below Grade Level
6, 8, 10, 12, 14
On Grade Level
16, 20
Above Grade Level
28, 30, 34