Theses and Dissertations (All)

7-24-2015

The Predictive Strength of the DIBELS Next ORF Assessment to the WIAT-III ORF and Reading Comprehension Subtests for Students Referred for Special Education Eligibility Evaluations

Jaclyn R. Pollard
Indiana University of Pennsylvania

Follow this and additional works at: http://knowledge.library.iup.edu/etd

THE PREDICTIVE STRENGTH OF THE DIBELS NEXT ORF ASSESSMENT TO THE

WIAT-III ORF AND READING COMPREHENSION SUBTESTS FOR STUDENTS

REFERRED FOR SPECIAL EDUCATION ELIGIBILITY EVALUATIONS

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Education

Jaclyn R. Pollard

Indiana University of Pennsylvania

August 2015

Indiana University of Pennsylvania
School of Graduate Studies and Research
Department of Educational and School Psychology

We hereby approve the dissertation of

Jaclyn R. Pollard

Candidate for the degree of Doctor of Education

_____05/29/15_____   _____Signature on file_____
Timothy J. Runge, Ph.D.
Associate Professor of Educational and
School Psychology, Advisor

_____05/29/15_____   _____Signature on file_____
Joseph F. Kovaleski, D.Ed.
Professor of Educational and School
Psychology

_____05/29/15_____   _____Signature on file_____
Mark J. Staszkiewicz, Ed.D.
Professor of Educational and School
Psychology

_____05/29/15_____   _____Signature on file_____
Lynanne Black, Ph.D.
Associate Professor of Educational and
School Psychology

ACCEPTED

_____Signature on file_____   _____
Randy L. Martin, Ph.D.
Dean
School of Graduate Studies and Research

Title: The Predictive Strength of the DIBELS Next ORF Assessment to the WIAT-III
        ORF and Reading Comprehension Subtests for Students Referred for Special
        Education Eligibility Evaluations

Author: Jaclyn R. Pollard

Dissertation Chair: Dr. Timothy J. Runge

Dissertation Committee Members: Dr. Joseph F. Kovaleski
                                 Dr. Mark J. Staszkiewicz
                                 Dr. Lynanne Black

        The researcher investigated the predictive strength of the Dynamic Indicators of
Basic Early Literacy Skills, Next Edition (DIBELS Next) Oral Reading Fluency (ORF)
assessment to the Wechsler Individual Achievement Test, Third Edition (WIAT-III) ORF
and Reading Comprehension subtests. The sample consisted of 75 second through
fifth grade students who participated in a special education eligibility evaluation.
Archival data included teacher-collected DIBELS Next ORF Words Correct (WC) and
Accuracy Percentage scores and school psychologist-collected WIAT-III ORF Rate,
ORF Accuracy, Reading Comprehension subtest scores. Analysis of potential sex
differences across all dependent variables revealed comparable performance between
males and females.

        Next, the researcher analyzed the predictive strength of the DIBELS Next ORF
scores for the WIAT-III ORF and Reading Comprehension scores aggregated and
disaggregated by grade. Potential predictive magnitude differences across grades were
also examined. Results suggested that the DIBELS Next ORF WC and Accuracy
Percentage scores predicted the WIAT-III ORF Rate; however, only the DIBELS Next
ORF WC scores significantly predicted the WIAT-III ORF Accuracy Percentage and

Reading Comprehension when data from all grades were aggregated. Additionally, the DIBELS Next ORF WC scores predicted the WIAT-III ORF Rate scores at the second, third, and fourth grades with the strongest correlations observed at the second and fourth grade level. The DIBELS Next ORF Accuracy Percentage scores predicted the WIAT-III ORF Accuracy scores at the second level grade. DIBELS Next ORF WC predicted the WIAT-III Accuracy Percentage scores at the third and fourth grades with no significant correlation strength differences between grades. The DIBELS Next ORF WC predicted the WIAT-III Reading Comprehension scores at the second grade and the DIBELS Next ORF Accuracy Percentage predicted the WIAT-III Reading Comprehension scores at the fourth grade with no significant correlation strength differences between grades. Despite the noted limitations in the current study, the results suggest that practitioners may want to be more selective when exclusively using the DIBELS Next ORF to make high-stakes decisions such as a referral for a special education eligibility evaluation and/or identification of an educational disability.

ACKNOWLEDGEMENTS

I would like to thank my dissertation committee for the guidance they have provided to me. They asked the difficult questions and forced me to think critically about my research. I would especially like to thank Dr. Runge, my committee chair, for the assistance and support he provided. Throughout the process Dr. Runge's feedback helped me expand my knowledge base and guide my thinking. Additionally, he always managed to provide kind words of encouragement and praise which helped keep my motivation level high.

My parents truly deserve appreciation since they have supported me throughout my life and I would not be where I am without them. For as long as I can remember, my parents have told me that I can achieve anything I put my mind to and they have always motivated me to reach my fullest potential. My mom has always served as a kind listening ear and she has provided the exact words of encouragement I needed to persevere.

I would also like to thank my husband, Leonard, for his constant love and patience. Not only was he able to make me smile after a long day or research and/or writing, but he was there to challenge me and push my skill set and knowledge to the highest level.

Lastly, I would like to thank the school psychologists in my district who took the time to compile and sanitize the data used in this study.   Without their selfless participation, this research study would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

THE PROBLEM

A federal law passed in 1975 called The Education for All Handicapped Children Act, or Public Law (PL) 94-142, served as the foundation of special education law. PL 94 - 142 required schools to identify children with disabilities and begin a referral process to determine if these children required special education services due to a disability. At that time, PL 94 - 142 mandated all children receive a non-discriminatory comprehensive assessment for special education eligibility determination. This law required the development of an Individualized Education Program (IEP) for all students found to have a disability and a demonstrated need for special education services. Several changes occurred since the original creation of PL 94 - 142; however, the basis of special education remains the same. The most current version of the act, Individuals with Disabilities Education Improvement Act (IDEIA; 2004), includes the following special educational classification categories: mental retardation[1], hearing impairment, speech or language impairment, visual impairment, emotional disturbance, orthopedic impairment, autism, traumatic brain injury, other health impairment, specific learning disability (SLD), deaf-blindness, and multiple disabilities (IDEIA, 2004) .

Of all students receiving special education services, 37% of them have an SLD classification (Institute of Education Sciences, 2013). Several different definitions of an SLD exist. IDEIA defines an SLD as:

---

[1] The term mental retardation was changed to intellectual disability (Rosa's Law, 2010).

The term specific learning disability means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which disorder may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations. Such term includes such conditions as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. Such term does not include a learning problem that is primarily a result of visual, hearing, or motor disabilities, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage(IDEIA, 2004, Sec. 602(30)(A-C)).

According to IDEIA federal regulations a child meets the criteria for an SLD if he or she does not achieve adequately for his or her age or meet state-approved grade-level standards in the following areas: oral expression, listening comprehension, written expression, basic reading skills, reading fluency skills, reading comprehension, mathematics calculation, or mathematics problem solving (Assistance to States for the Education of Children With Disabilities, 2006). In addition, a child meets the SLD criteria if he or she does not make sufficient progress in one or more of the above mentioned areas when provided scientific, research-based intervention or if the child exhibits a pattern of cognitive and/or academic strengths and weaknesses (Assistance to States for the Education of Children With Disabilities, 2006).

Prior to considering if a student meets the criteria for an SLD and needs special education services, IDEIA mandates that interventions are implemented with integrity in the regular education setting. Most schools have an early intervening team in place to

help students who are struggling in the general education classroom (Truscott, Cohen, Sams, Sanborn, & Frank, 2005).

The Early Intervening Process

The names for the early intervening team and specific processes utilized by the team changed across the years; however, the basis of the early intervening teams remains the same (M. K. Burns, Wiley, & Viglietta, 2008). Early intervening teams include several school-based professionals who determine if a child requires a recommendation for a special education eligibility evaluation. More specifically, these teams suggest interventions for students who struggle in the general education classroom and then monitor student progress. If the students do not respond to research-based interventions, the team recommends the students for special education eligibility evaluations.

Individual Problem-Solving Model

One model of an early intervening team, the individual problem-solving team model, focuses on one student at a time from initial problem identification to the resolution of the problem (Tilly, 2008). In this model, teachers recommend a student to the early intervening team if he or she is lagging behind peers in critical skill development. Once a teacher identifies a student in need of support, the early intervening team works together to define and analyze the student's problem by reviewing student data. To address the student's weaknesses, the early intervening team develops interventions and the teacher implements them in the classroom. After the teacher implements the interventions for a reasonable period of time (e.g.,4 - 8 weeks), the early intervening team analyzes academic data to determine if the student's

performance nears grade level expectations and/or if the student is making marked improvement in his or her skills. If the student does not show gains concurrent with implementation of research-based interventions, the early intervening team often recommends the student for a special education eligibility evaluation (Ikeda, Tilly, Stumme, Volmer, & Allison, 1996).

Multi-Tiered Intervention Model

An early intervening team can be used as an isolated intervention model, as mentioned above, or in conjunction with a multi-tiered model of intervention commonly referred to as response to intervention (RTI; M. K. Burns & Gibbons, 2008; McDougal, 2010).RTI is a multi-tiered system of support and service delivery.   A multi-tier intervention model involves all students, not just those who are experiencing academic difficulties. By providing research-based universal instruction at Tier I, teachers support the learning of all students. In Tier I, all students receive the standard core curriculum and complete universal screening assessments at regular intervals (e.g., two or three times a year) to measure their progress in critical academic skills such as reading, mathematics, and spelling. Educators analyze universal screening data to determine if changes need to be made in the core instruction and to guide decisions about supplemental or intensive instruction for students who need additional support (Ikeda, Neessen, & Witt, 2008; Ikeda, et al., 1996; McDougal, 2010). Some researchers suggest administering multistage screening to avoid false positives that commonly result from universal screening (Fuchs, Fuchs, & Compton, 2012). Students, whose skills are not progressing, whether identified by universal screening or a multistage screening, receive supplemental instruction in the form of targeted group interventions,

commonly referred to as Tier II interventions, to focus on their areas of need. Teachers

monitor students' skill development frequently to determine if they are making gains in

their areas of identified weakness given the addition of these targeted, Tier II

interventions. If students are still not making gains despite high fidelity implementation

of targeted skills, students receive more intense intervention referred to as Tier III

interventions. The early intervening team members analyze the student data and

suggest interventions to meet student needs. These intensive interventions, classified

as Tier III supports, are often individualized and of an intense duration and depth. The

early intervening team creates individual goals for these students and teachers monitor

students' progress. In most RTI models, the early intervening team typically

recommends students who continue to be un-responsive despite high fidelity Tier I, Tier

II, and Tier III supports for a special education eligibility evaluation(M. K. Burns &

Gibbons, 2008; McDougal, 2010).

Curriculum-based measurement (CBM). When using both the individual

problem-solving model and the multi-tiered intervention model early intervening teams

assess academic skills and growth. CBM are common tools used for assessing

academic growth (Deno, 2003; Fuchs, Tindal, & Deno, 1984; Marston, Mirkin, & Deno,

1984). CBM are short, standardized assessments with adequate psychometric qualities

that are used to measure academic skills and identify academic growth (Christ &

Silberglitt, 2007). Educators utilize CBM as universal screeners and progress

monitoring tools. Universal screening is an assessment administered to determine the

academic proficiency of all students, and teachers analyze the scores to guide the

decision making process in regards to core instruction and supplemental or intensive

instruction for students in need of additional support(Ikeda et al., 2008). Progress monitoring tools are screeners administered regularly (e.g., weekly or monthly) to measure students' improvement in a certain academic skill (M. K. Burns & Gibbons, 2008). CBM are sensitive to short-range effects or changes based on current intervention, which make them very useful as progress monitoring tools (Marston et al., 1984). The purpose of CBM is not to make diagnostic decisions, nor do they explain why a student is experiencing difficulty or exactly how to intervene. Educators mainly employ CBM as a data source for problem solving and to determine a student's response to interventions (Christ & Silberglitt, 2007; Marston et al., 1984). However, more recently, some school psychologists use CBM to make special education eligibility decisions (Deno, 2003).

Researchers originally developed CBM to help special education teachers evaluate their instruction, write IEP goals, and monitor the progress of special education students (Deno, 2003; Shinn, 2008). Since their creation, the use of CBM expanded to the general education classroom. CBM now provide all teachers a simple, psychometrically-sound tool to measure academic achievement, and these data can be used for a variety of purposes including universal screening, early intervening assessments, placement in special education, formative evaluations, and evaluations to determine if a student no longer meets the criteria for special education (Deno, 2003; Shinn, 2008). In addition, school-based professionals analyze CBM data to predict success on high-stakes assessments (Deno, 2003). High stakes assessments include, but are not limited to, assessments used during an evaluation to determine a student's eligibility for special education.

Reading CBM. CBM are standardized and validated in the areas of reading, mathematics computation and application, spelling, written expression, and early literacy and numeracy (Shinn, 2008). In the mid-1980s, researchers created a mnemonic; dynamic indicators of basic skills (DIBS; Shinn, 2008) to explain the intent of CBM since they were drastically different than standard educational practices. Years later, researchers expanded on the mnemonic to focus on literacy skills and created a CBM called the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). Although a substantial amount of research exists regarding the technical adequacy of CBM across several academic areas, the most research exists in the area of reading (Ardoin, Christ, Morena, Cormier, & Klingbeil, 2013; Christ & Silberglitt, 2007; Deno, 2003; Fuchs et al., 1984; Marston et al., 1984).

Several different types of reading CBM exist; including those measuring first or initial sound fluency, letter naming fluency, phoneme segmentation fluency,nonsense word fluency, oral reading fluency (ORF), and reading comprehension. Specifically, the utility of reading-curriculum based measures (R-CBM) or ORF assessments have the strongest research base (Ardoin et al., 2013). When administering ORF assessments, the evaluator calculates the number of words a student reads aloud correctly per minute. Research suggests that the amount of words read per minute is an excellent indicator of overall reading ability (Fuchs, Fuchs, & Hosp, 2001; Shinn, 2008 *cf* Goodman, 1969; Tierney, 1998). This type of assessment directly measures a student's ORF skills which consist of two components: rate and accuracy. Despite the fact that ORF measures directly measure reading fluency, many schools in North America use the ORF assessments for screening students who may be at risk for

reading comprehension deficits in the future (Petscher & Kim, 2011). ORF is a strong predictor of reading comprehension, which is the ultimate goal of reading (Fuchs et al., 2001).

School psychologists employ several measures to screen ORF of students. Some of the most popular CBM include easyCBM (Tindal & Alonzo, 2009), AIMSweb (Edformation, 2005), and the DIBELS, Next Edition (DIBELS Next; Good & Kaminski, 2011). These CBM include a battery of assessments that measure phonemic awareness, letter identification, phonics, ORF, and reading comprehension.

Phonemic awareness is the ability to hear, identify, and manipulate sounds in orally presented words.   Phonemic awareness includes being able to segment spoken words into its component sounds, or blend isolated sounds to produce a spoken word. Letter identification entails naming written letters. Phonics skills involve the knowledge of letter-sound correspondences or relationships between letters and phonemes. When readers use phonics skills to decode words, they must generate the specific sounds for letters and blend the sounds together to read words. ORF describes how quickly and accurately a person reads. Reading comprehension is the skill with which a person can recall and understand information obtained while reading text.

Educators use the DIBELS Next assessment to screen for students in need of reading intervention and to monitor their progress in the development of skills. Specifically, the DIBELS ORF subtest is a significant predictor of standardized test scores measuring overall reading proficiency (Deno, 2003; Goffreda & Pedersen, 2009). Therefore, the DIBELS Next ORF data are a valuable piece of information used to

determine if students need a more intense intervention and/or a referral for a special education evaluation.

Referral for Special Education Eligibility Evaluation

The DIBELS Next assessment is an example of a universal screener and progress monitoring tool educators utilize to measure academic proficiency of students. Early intervening teams analyze DIBELS Next data, along with other classroom data, to help guide their decisions regarding whether a child is in need of a special education eligibility evaluation due to lack of progress and performing below grade level in a specific skill area. When the early intervening team makes a recommendation for a student to undergo a special education eligibility evaluation due to a suspected SLD, the evaluation process begins. The new SLD identification language in IDEIA 2004 allows for various SLD identification approaches including the Psychoeducational Assessment and Response to Intervention (RTI) approaches.

Approaches to SLD Identification

Both approaches, the Psychoeducational Assessment and RTI approaches, involve analyzing previously collected student data (possibly CBM data) and denoting exclusionary criteria. Exclusionary criteria are factors such as limited English proficiency; cultural, environmental, or economic disadvantage; or other disabilities that must be ruled out as the primary cause of difficulties (Assistance to States for the Education of Children With Disabilities, 2006). Within a Psychoeducational Assessment approach, school psychologists administer standardized, individually-administered, norm-referenced tests and use this performance data to determine eligibility. Among a field of diagnostic assessments of academic skills utilized by school psychologists, one

9

commercially-available product is the Wechsler Individual Achievement Test, Third Edition (WIAT-III; Wechsler, 2009). The WIAT-III contains subtests that assess oral language, written expression, reading, and mathematics skills. When using a RTI model to determine whether an SLD exists, professionals analyze CBM, such as the DIBELS Next, which are collected throughout the implementation of a research-based intervention. Some professionals solely rely on the CBM data to make special eligibility decisions, while other professionals administer individually-administered, norm-referenced tests in addition to analyzing CBM data.

Within the Psychoeducational Assessment approach, several models exist. These models include the discrepancy model, low achievement model, or intra-individual differences model (Fletcher, Lyon, Fuchs, & Barnes, 2007).

Psychoeducational Assessment Models

The discrepancy model involves identifying if there is a discrepancy between aptitude and achievement. This is the most commonly used approach for identifying an SLD (Fletcher et al., 2007). When using this model, students complete cognitive tests to assess their aptitude and standardized achievement tests to assess their academic achievement. If a student's ability level is significantly higher than his or her achievement level, the data, within this model, suggest the presence of an SLD.

Another Psychoeducational Assessment model, the low achievement model, entails merely looking at the student's achievement level. If the student's academic level is significantly low, he or she qualifies as a student with an SLD. Professionals using this model usually establish a cut point to determine a level of low achievement (Fletcher et al., 2007).

A third Psychoeducational Assessment model, the intra-individual differences model, focuses on the identification of cognitive strengths and weaknesses. Student performance on cognitive or neuropsychological assessments helps identify cognitive strengths and weaknesses. The identified weaknesses are thought to be the underlying cause of academic struggles. According to this model, a person with an SLD has cognitive strengths in one area and cognitive weaknesses in another area resulting in academic underachievement (Fiorello & Primerano, 2005; Naglieri, 2003).

The RTI Model

The RTI model is substantially different from the Psychoeducational Assessment approaches. Within an RTI model, the focus is on documenting a lack of response to intervention despite the use of research-based interventions when determining whether a student has an SLD (Fletcher et al., 2007; Kovaleski, VanDerHeyden, & Shapiro, 2013). The RTI model includes a universal screening of all students and the regular progress monitoring of students struggling in certain areas. Teachers provide students with specific, research-validated or empirically-supported interventions to remediate their area of difficulty. If these interventions are implemented with integrity and students show academic improvements, professionals do not suspect an SLD. If the students do not show improvement in the specific skill areas targeted for intervention despite high fidelity implementation, then the academic underachievement may be due to an SLD (Fletcher et al., 2007; Kovaleski, VanDerHeyden, & Shapiro, 2013).

<center>**Statement of the Problem**</center>

The most commonly classified SLDs are in reading, as 80-90% of students classified as having an SLD exhibit reading weaknesses (Kavale & Reese,

<center>11</center>

1992;President's Commission on Excellence in Special Education, 2002). In particular, most students with SLDs experience difficulties with word recognition skills (Fletcher et al., 2007). Early intervening teams commonly administer the DIBELS Next assessment to analyze student gains in reading and recommend students for special education eligibility evaluations. Although the DIBELS Next is a reliable tool, the purpose of this study is to investigate the construct validity of the DIBELS Next. Establishment of the DIBELS Next construct validity is vital due to the credence commonly placed on the DIBELS Next data when early intervening teams are making recommendations for students to undergo special education eligibility evaluations and/or using the DIBELS Next data to make special education eligibility decisions within an RTI model.

When using the Psychoeducational Assessment model to determine if a student has an SLD, professionals analyze DIBELS Next scores during the early intervening process; however, they rely on individually-administered, norm-referenced testing data to determine special education eligibility. Given the weight placed on DIBELS Next data as an indication of whether a full psychoeducational evaluation is warranted, further investigation of the construct validity of DIBELS Next is warranted. Further, when using the RTI model for SLD identification, some professionals use individually-administered, norm-referenced tests in combination with CBM data, and other professionals solely rely on CBM data to make special education eligibility decisions. Therefore, the validity of DIBELS Next data is important as well if schools and professionals operate within an RTI model.

Support for the construct validity of the DIBELS ORF has implications for professionals and schools regardless of whether the Psychoeducational Assessment

model or RTI model is used. Within a Psychoeducational Assessment model, support

for the construct validity of the DIBELS ORF to established, traditional measures of

reading achievement (i.e., WIAT-III) will allow early intervening teams to be more

confident that DIBELS Next ORF scores are useful predictors of future performance on

standardized assessments such as the WIAT-III. Within an RTI model, strong construct

validity of the DIBELS Next lends additional credibility to using these data within the

special education decision-making process.

Contrarily, if results of this study do not support the construct validity of the

DIBELS ORF to traditional measures of reading achievement (i.e., WIAT-III), then early

intervening teams should de-emphasize the importance of DIBELS Next ORF as an

indicator of reading achievement in the determination of an SLD. Such a consideration

would have implications for decision-making within either a Psychoeducational

Assessment model or RTI model.

To empirically validate the construct validity of the DIBELS Next, the researcher

analyzed the relationship between the DIBELS Next ORF and WIAT-III ORF and

Reading Comprehension subtests.   DIBELS Next ORF assessment consists of two

scores: Words Correct (WC) and Accuracy Percentage. The WC score measures the

number of words read correctly in one minute while the Accuracy Percentage measures

the percentage of words read correctly out of the total words attempted (Good &

Kaminski, 2011). The WIAT-III ORF subtest score is a compilation of the speed and

accuracy of reading. The WIAT-III ORF subtest also has two component scores, the

Accuracy score and the Rate score. The Accuracy score measures the percentage of

words read correctly when compared to the total words read; whereas, the Rate score

measures the efficiency, or speed, of reading words correctly. The WIAT-III Reading

Comprehension subtest measures the ability to answer comprehension questions after

reading passages (Wechsler, 2009). Specifically, the researcher used multiple linear

regression (MLR) to analyze the predictive strength of the DIBELS Next ORF WC and

Accuracy Percentage scores to the WIAT-III ORF Accuracy, WIAT-III ORF Rate, and

WIAT-III Reading Comprehension subtest scores. Additionally, the researcher

examined differences between groups based on sex with regard to oral reading fluency

and comprehension skills using a multivariate analysis of variance (MANOVA).

Significance of the Problem

No Child Left Behind (NCLB; 2001) and the IDEIA (2004) indicate that progress

monitoring is an essential tool in evidence-based practices. Progress monitoring tools,

such as CBM, are used to measure the growth of students in response to evidence

based practices are to make educational classification decisions. Most school systems

do not solely use CBM to make special education eligibility determinations, although the

trend is increasing. Currently, 13% of state regulations require the sole use of RTI for

SLD identification (Maki & Floyd, 2014; Zirkel & Thomas, 2010). However, early

intervening teams, which most schools have, commonly analyze CBM data to help

determine if a student should undergo a special education eligibility evaluation (Shinn,

2008).

As of 2010, the Wechsler Individual Achievement Test, Second Edition (WIAT-II;

Wechsler, 2001) was one of the three most commonly used standardized achievement

tests (T. G. Burns, 2010). While similar data are unavailable for the WIAT-III, it is

understood to be a commonly used assessment for schools when making eligibility

determinations due to the popularity of its predecessor, the WIAT-II. The DIBELS Next ORF and WIAT-III ORF both claim to measure the same construct, ORF, although this has yet to be empirically tested. Although previous editions of the DIBELS and WIAT have been correlated in research(Munger, 2010), as of now, there is no research on the predictive strength of the DIBELS Next ORF assessment for the WIAT-III ORF or Reading Comprehension subtests.

If this study supports that the DIBELS Next ORF assessment as an adequate predictor the WIAT-III ORF and Reading Comprehension scores, early intervening teams can emphasize the DIBELS Next ORF scores when referring a student for a special education eligibility evaluation and professionals can be more confident when using CBM data for special education eligibility determination within the RTI model. If this study does not support the DIBELS Next ORF assessment as a strong predictor of WIAT-III ORF and Reading Comprehension scores, early intervening teams may de-emphasize DIBELS Next ORF scores when deciding whether students should undergo a special education eligibility evaluation. Additionally, professionals using solely CBM to determine special education eligibility may need to consider supplementing the CBM data with other assessment data. If the DIBELS Next ORF assessment illustrates that students are struggling significantly and the WIAT-III ORF and Reading Comprehension scores do not confirm the deficit, school districts may be wasting time and resources on unnecessary special education eligibility evaluations. If the DIBELS Next ORF assessment shows that students are not performing poorly and early intervening teams decide to not refer a student for a special education evaluation based on the students'

performance on the DIBELS Next ORF assessment school districts may be excluding students from special education eligibility evaluations who truly need to be evaluated.

More specifically the researcher investigated the sensitivity and specificity of the DIBELS Next ORF assessment by using the WIAT-III as a comparison tool. Sensitivity measures the proportion of actual positives correctly identified. If students score poorly on the DIBELS Next ORF assessment and the WIAT-III ORF and Reading Comprehension subtests, this concept would be referred to as a true positive. If students score well on the DIBELS ORF assessment and poorly on the WIAT-III ORF and Reading Comprehension subtests, this concept would be referred to as a false positive. Specificity measures the proportion of actual negatives correctly identified. If students score well on the DIBELS Next ORF assessment and well on the WIAT-III ORF and Reading Comprehension subtests, this concept would be referred to as a true negative. If the students score poorly on the DIBELS Next ORF assessment and well on the WIAT-III ORF and Reading Comprehension subtests, this concept would be referred to as a false negative. This study only investigates the possibility of true positives and false negatives since students who score well on the DIBELS Next ORF assessment are not included in the sample. Students who perform well on the DIBELS Next ORF assessment should not be referred for special education eligibility evaluations, and therefore, would not be given the WIAT-III, lessening the importance of analyzing students who score well on the DIBELS ORF in this study. Therefore, the researchers did not investigate false positives or true negatives.

Research Question

What is the predictive validity of DIBELS Next ORF WC and Accuracy Percentage scores to WIAT-III ORF Rate, ORF Accuracy, and Reading Comprehension scores, and is it the same across grade levels?

Hypotheses

The researcher hypothesized that the DIBELS Next ORF (WC and Accuracy Percentage) scores will have strong predictive strength for the WIAT-III ORF Accuracy score. Test administrators calculate both the DIBELS ORF Accuracy Percentage and WIAT-III ORF Accuracy score by subtracting the number of errors made while reading a passage from the total number of words read. Since the assessment method of oral reading accuracy is the same on both assessments, they are likely highly correlated. On the DIBELS Next ORF WC, the calculation of the number of words read correctly in a passage is an assessment of oral reading rate. Oral reading rate and accuracy are components of ORF, and therefore, the DIBELS Next ORF WC and WIAT-III ORF Accuracy scores are likely highly correlated.

The researcher hypothesized that the DIBELS Next ORF (WC and Accuracy Percentage) scores will have strong predictive strength for the WIAT-III ORF Rate score. The DIBELS Next ORF WC score is a calculation of oral reading rate resulting from counting the number words read correctly in a passage. The amount of time it took a person to read a passage on the WIAT-III ORF subtest measures oral reading rate and results in a WIAT-III ORF Rate score. Both instruments assess oral reading rate, and therefore, they are likely highly correlated. By subtracting the number of errors made while reading a passage from the total number of words read on the DIBELS Next, the test administrator calculates a DIBELS Next Accuracy Percentage score. Oral

17

reading rate and accuracy are components of ORF, and therefore, the DIBELS Next ORF Accuracy Percentage and WIAT-III ORF Rate scores are likely highly correlated.

The researcher hypothesized that the DIBELS Next ORF WC and Accuracy Percentage scores will have strong predictive strength for the WIAT-III Reading Comprehension subtest score. The WIAT-III Reading Comprehension subtest is an assessment of reading comprehension. ORF is highly correlated with reading comprehension (Fuchs et al., 2001). Because the DIBELS Next ORF WC and Accuracy Percentage scores measure two components of ORF, rate and accuracy, it is likely that the scores on the DIBELS Next ORF WC and Accuracy Percentage strongly correlate with the WIAT-III Reading Comprehension subtest.

The researcher hypothesized that no significant difference of predictive strength will exist between groups (developmental level). No research exists to support that a difference will exist based on grade level.

See Table 1 for the research question with accompanying hypotheses and variables.

Definition of Terms

Curriculum-Based Measurement (CBM)

CBM are short, standardized procedures used to measure academic performance and identify academic growth (Christ & Silberglitt, 2007). Educators administer CBM to measure achievement through both universal screening and progress monitoring efforts. CBM scores provide data for educators to make decisions regarding efficacy of early intervening interventions, placement in special education,

and evaluations to determine if a student is no longer in need of special

education(Deno, 2003; Shinn, 2008).

Table 1

Research Question, Hypotheses, and Variables

Research Question	Hypotheses	Variables
What is the predictive validity of DIBELS Next ORF WC and Accuracy Percentage scores to WIAT-III ORF Rate, ORF Accuracy, and Reading Comprehension scores, and is it the same across grade levels?	The DIBELS ORF WC and Accuracy Percentage scores will have strong predictive strength for the WIAT-III ORF Rate, ORF Accuracy, and Reading Comprehension subtest scores.	DIBELS Next ORF WC and Accuracy Percentage scores WIAT-III ORF Rate, ORF Accuracy, and Reading comprehension scores
	No grade level difference will exist.	DIBELS Next ORF WC and Accuracy Percentage scores WIAT-III ORF Rate, Accuracy, and Reading Comprehension scores Developmental Level

Note. DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct; WIAT-III = Wechsler Individual Achievement Test, Third Edition.

Developmental Level

Developmental Level, as measured by this study, refers to the chronological

grade level of the students.

Dynamic Indicators of Basic Early Literacy Skills, Next Edition (DIBELS Next)

The DIBELS Next (Good et al., 2011) is an assessment used to measure early

literacy skills of students in kindergarten through sixth grades. The DIBELS Next

consists of six tests including First Sound Fluency, Letter Naming Fluency, Phoneme

Segmentation Fluency, Nonsense Word Fleuncy, ORF, and DAZE (comprehension measure). Certain tests are prescribed for different grade levels, corresponding to the developmental nature of reading acquisition.

DIBELS Next ORF Accuracy Percentage Score

On the ORF test of the DIBELS Next assessment, the examiner measures the Accuracy Percentage score. The Accuracy Percentage score is the percentage of words read correctly aloud in one minute. The Accuracy Percentage score measures reading accurateness, which is one element of oral reading fluency.

DIBELS Next ORF Words Correct (WC) Score

The examiner calculates the WC score on ORF test of the DIBELS Next assessment. The WC score is the number of words read aloud correctly in one minute. The WC score is an assessment of reading efficiency or speed, which is another component of oral reading fluency.

Oral Reading Fluency (ORF)

ORF is a measure of the accuracy and rate with which a person reads out loud in a certain amount of time. ORF is a strong predictor of reading comprehension, which is the foremost objective of reading (Fuchs et al., 2001). ORF is this study is measured with the DIBELS Next ORF and WIAT-III ORF assessments.

Early Intervening Team

An early intervening team is a school-based multi-disciplinary team that suggests interventions for students who are having academic difficulties and monitors students' progress. If a student does not respond to research-based intervention, the team

recommends the student for a special education eligibility evaluation to determine if he or she meets the criteria for a special education disability (Tilly, 2008).

Reading Comprehension

Reading comprehension is the ability to understand and recall information read. More specifically, reading comprehension is a process of simultaneously extracting and constructing meaning from written words(Snow, 2002). Reading comprehension involves three elements: the reader, the text that is being read, and the activity in which comprehension is involved. These three elements of comprehension define a concept that occurs within a sociocultural context. The abilities of the readers, the texts that are used, and the activities in which the readers are involved with the texts are influenced by the sociocultural context. The sociocultural context influences the students' experiences and the students' experiences influence the context(Snow, 2002). Reading comprehension is the focal rationale of reading (Fuchs et al., 2001). On the WIAT-III Reading Comprehension subtest, which is used in this study, students read a passage and verbally answered open-ended questions about the passage in order to measure their reading comprehension skills.

Special Education Eligibility Evaluation

A special education eligibility evaluation is a comprehensive evaluation in which professionals analyze multiple data sources to determine if a student has an educational disability. Professionals then consider whether or not the student requires specially designed instruction due to his or her disability. If a disability is found to exist and the student requires specially designed instruction, the student qualifies for special education services.

Specific Learning Disability (SLD)

Although several definitions of an SLD exist, IDEIA delineated a federal definition.

The term specific learning disability means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which disorder may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations. Such term includes such conditions as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. Such term does not include a learning problem that is primarily a result of visual, hearing, or motor disabilities, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage(IDEIA, 2004, Sec. 602(30)(A-C)).

Universal Screening

Universal screening is an assessment administered to determine the academic proficiency of all students and to guide the decision making process in regards to core instruction and supplemental or intensive instruction for students in need of additional support (Ikeda et al., 2008).

Wechsler Individualized Achievement Test, Third Edition (WIAT-III)

The WIAT-III is an academic diagnostic assessment used with children from ages four though 19 (Breaux, 2009). The WIAT-III contains subtests that assess reading, mathematics, writing, and oral language skills. Individual students complete this norm-referenced test under standardized conditions during a comprehension evaluation for special education eligibility.

WIAT-III Oral Reading Fluency (ORF) Accuracy Score

The ORF Accuracy score is a calculation of the percentage of words read aloud correctly in given passages on the ORF subtest on the WIAT-III. The ORF Accuracy score measures precision of word reading, which is an element of ORF.

WIAT-III ORF Rate Score

To calculate the ORF Rate score, the test administrator counts the numbers of words read correctly on given passages on the ORF subtest on the WIAT-III. The ORF Rate score measures reading speed, which is an element of ORF.

WIAT-III ORF Subtest Score

The ORF subtest score is a calculation measuring both reading rate and accuracy on the WIAT-III by resulting in a score that provides a comprehensive measure of ORF. Evaluators measure the time to complete each passage and words read correctly when calculating an ORF subtest score.

WIAT-III Reading Comprehension Subtest Score

The Reading Comprehension subtest score is a calculation on the WIAT-III that measures the ability to recall what a person has read. Students read passages based on their grade level and then answer questions about each passage. Based on the accuracy of the answer, evaluators assign point values to each answer. The test creators provide samples the information that needs to be included in each answer to earn a certain point value.

Assumptions

Several assumptions exist in the current study. It is an assumption that all teachers who administered the DIBELS Next ORF assessment followed standardized

assessment procedures regarding administration and scoring. If the teachers did not follow standardized procedures the reliability of the data could be compromised. Reading specialists, general education teachers, and special education teachers administer the DIBELS Next assessment to students at the study site. All teachers receive DIBELS Next manuals when they begin teaching and it is their professional responsibility to read and familiarize themselves with the testing procedures. Reading specialists, who receive in-depth training on several reading screening measures including the DIBELS Next assessment, consult with the regular and special education teachers to provide training if requested. To use the DIBELS Next as a comparative measure, it is important that all teachers follow the standardized administration and scoring rules. The researcher is confident that all teachers followed standardized protocol due to a professional responsibility; however, the researcher did not take steps to empirically validate that teachers followed standardized administration and scoring procedures.

It is also an assumption that students did not previously see the specific DIBELS Next ORF probes used in this study. If students had prior exposure to the probes a practice effect could compromise the validity of the data. When a student receives reading interventions to target an area of reading weakness, the teacher provides the student with a DIBELS Next ORF booklet. Teachers administer the DIBELS Next ORF probes in numerical order every two weeks and are not to repeat the administration of any probe. There are 20 DIBELS Next ORF progress monitoring probes also decreasing the likelihood that students were previously exposed to the probes. Students do not maintain possession of the DIBELS Next booklets; therefore, they

would not have the opportunity to read the probes independently prior to assessment. Consequently the researcher is confident that students did not see the specific DIBELS Next ORF probes prior to this assessment.

In addition, it is an assumption that the school psychologists who administered the WAIT-III followed standardized administration procedures. School psychologists have extensive training during graduate school to solidify their practice of administering standardized assessments. Therefore, the researcher is confident that school psychologists adhered to standardized administration and scoring procedures suggesting that data are reliable.

Lastly, administration of the WIAT-III to a student should only occur once per calendar year, and therefore, students should have not previously seen the reading passages specific to their grade levels. The WIAT-III ORF and Reading Comprehension subtests include reading passages for each grade level. School psychologists usually administer the WIAT-III as part of a comprehensive evaluation or re-evaluation, and therefore, it is rare that a student takes the WIAT-III assessment more than once per year. If for any reason a student were to need to complete a standardized, norm-referenced assessment within the one year time frame, the school psychologists should administer a different assessment as per best practices. Although the researcher did not take any steps to verify this assumption the researcher is confident students did not take the WIAT-III for at least one year prior to this assessment limiting the practice effect.

Limitations

Readers should examine the results of this study while taking into account some methodological limitations. Three notable limitations include the use of a convenience sample, a small sample size, and the researcher only investigating one thread of validity evidence for the DIBELS Next.

This study utilized a convenience sample which likely threatens the external and internal validity of this study. Because the sample was not randomly selected, the ability to generalize the findings to the general population is threatened because convenience samples may not be representative of the general population. Participant selection must be random to ensure the sample accurately reflects the population of interest and for findings to generalize to a broader population.

The sample size used for this study is another limitation. It is difficult to find statistically significant relationships when using a small sample size. A greater difference is necessary to find statistical significance when using a small sample size, and therefore, statistical differences may not be found even if differences exist. Additionally, statistical tests generally necessitate a larger sample size to ensure the sample is a representative distribution of the population. If a sample size is too small, it is difficult to generalize the results of the study to all students referred for special education eligibility evaluations.

An additional limitation of this study is that it only investigates one type of validity (construct validity) for the DIBELS Next. Other lines of evidence, including different types of validity, will remain unknown for the DIBELS Next.

26

Summary

The main purpose of this study is to investigate the predictive strength of the DIBELS Next ORF assessment to the WIAT-III ORF and Reading Comprehension subtests. The researcher analyzed the predictive validity between groups based developmental level. If this study supports that the DIBELS Next ORF assessment is a reasonable predictor the WIAT-III ORF and Reading Comprehension subtest scores, early intervening teams will have stronger empirical justification for using the DIBELS Next ORF scores to make important decisions including a referral for a special education eligibility evaluation and/or determination of an educationally-relevant disability. If this study does not support the DIBELS Next ORF assessment as a strong predictor of WIAT-III ORF and Reading Comprehension subtest scores, early intervening teams may lessen the importance placed on DIBELS Next ORF scores when deciding whether or not students should undergo a special education eligibility evaluation or when considering the presence of an educationally-relevant disability.

CHAPTER 2

LITERATURE REVIEW

The 1975 Education for All Handicapped Children Act, or Public Law (PL) 94-142, was the starting point for special education law (Education for All Handicapped Children Act, 1975). PL 94-142 required school professionals to locate children with disabilities and initiate the referral process to eligibility for special education services. It mandated all children undergo a non-discriminatory comprehensive assessment for special education eligibility determination. This law also mandated each student with a disability receive an Individualized Education Program (IEP) to explain the supports and services needed to attend and benefit from a public school education. The conceptualization of providing free and appropriate public education (FAPE) also occurred at that time. Several legislative changes emerged since the passage of PL 94-142 including the revision in 1990, PL 101-476, which altered the name of the law from the Education of All Handicapped Children Act to the Individuals with Disabilities Education Act (IDEA).

No Child Left Behind (NCLB) also contributed to major educational changes which influence special education (NCLB, 2001). NCLB repeatedly stated schools must use scientific, research-based practices to help students achieve academic success. NCLB held school professionals accountable for all students' learning by putting rewards or sanctions into place based on student performance. According to NCLB, all children should be proficient in basic reading and math skills by 2014. To reach this goal, school professionals have increased their focus on monitoring student progress and providing intervention services to those students performing below grade level.

Professionals frequently use curriculum-based measurement (CBM) to gather data to make recommendations for interventions for students including the possibility of a referral for a special education eligibility evaluation.

The 2004 reauthorization of IDEA brought about several significant changes including a name change to the Individuals with Disabilities Education Improvement Act (IDEIA). The most notable change made in the IDEIA is in regard to the definition of specific learning disabilities (SLD). IDEIA stipulates that school districts are not required to establish a severe discrepancy between aptitude (IQ) and achievement when deciding if a student has an SLD (IDEIA, 2004), amounting to a dramatic shift in how students are classified as SLD. This language echoes the importance of data-based decision making and progress monitoring stated two years prior in NCLB (Tilly, 2008).

Early Intervening Process

The early intervening process involves a multi-disciplinary team that plans, implements, and monitors the efficacy of an intervention designed to remediate students' academic and/or behavioral deficits. While there are numerous problem-solving models, early intervening teams typically apply a common problem-solving process encompassing very similar steps and activities (Tilly, 2008). The team members work together to identify a problem, determine why the problem is occurring, make suggestions to remediate the problem, and determine if the intervention solved the problem (Ikeda, et al., 1996; Tilly, 2008). Teams may apply the problem-solving model to one student at a time or incorporate it with the multi-tier intervention model allowing the team to focus on a large number of students. Regardless of the type of

problem solving model used, teams frequently utilize CBM to monitor student progress while implementing an intervention. Teams then examine CBM data to determine if the intervention is having the desired effect on a student's skills or if the team should recommend a student for a special education eligibility evaluation.

Individual Student Problem-Solving Model

Some early intervening teams employ the individual problem-solving model to apply structure to their team process. The original design of the individual problem-solving model involved focusing on one student at a time. The model allows the team to collaborate around a student's challenges from initial problem identification to problem solution (Ikeda et al., 1996). The problem-solving model aims to expend resources early in the hope of remediating the problem before it intensifies. Identifying the student's area of need is the first step of the problem-solving process. Next, the team determines why the problem is occurring and what interventions may help solve the problem. An educator then implements the intervention with the student. Lastly, the team analyzes the CBM data to determine if the student showed progress or if he or she requires a special education eligibility evaluation CBM serves as a tool throughout the problem-solving process to identify the problem, determine the magnitude of the problem, establish goals for student gain, and monitor progress.

Multi-Tier Intervention Model

It is possible for an early intervening team to work in combination with a multi-tier intervention model to help students. The multi-tier intervention model, commonly referred to as Response to Intervention (RTI), includes several levels or tiers of instruction and/or intervention for students (M. K. Burns & Gibbons, 2008; Haager,

Klingner, & Vaughn, 2007; McDougal, 2010). Tier I, or primary instruction, refers to instruction that all students receive in the classroom. Tier II, or secondary intervention, provides strategic intervention to students who are performing below grade level (M. K. Burns & Gibbons, 2008). Students who do not demonstrate progress after receiving Tier II interventions receive Tier III, or tertiary interventions, which are standardized, intensive interventions (M. K. Burns & Gibbons, 2008). The most commonly implemented multi-tier intervention models focus on reading (Haager et al., 2007).

By combining a multi-tier model with an early intervening team, the problem-solving system includes all students, not just those who perform below grade level (Ikeda et al., 1996; McDougal, 2010; Tilly, 2008). Universal screenings are assessments administered to all students to measure the efficacy of the Tier I, core curriculum. Additionally, these assessment scores serve as indicators when determining which students may need additional supports, via Tiers II or III interventions (Ikeda et al., 2008). If a 80-90% of students do not meet grade level standards based on Tier I instruction, the school would likely determine that changes need to be made to the core curriculum (M. K. Burns& Gibbons, 2008; McDougal, 2010). Despite a strong scientific, research-based core curriculum, it is estimated that 10%-20%of students require Tier II and Tier III interventions to make gains (M. K. Burns & Gibbons, 2008; McDougal, 2010). The early intervening team uses universal screening data to place those students identified into intervention groups based on their strengths and weaknesses. Roughly 5%-10% of all students need Tier II interventions (M. K. Burns & Gibbons, 2008; McDougal, 2010). Teachers then monitor students' progress frequently to determine if they demonstrated skill improvement. If the student showed some

improvement, the early intervening team recommends a continued Tier II intervention or a new Tier I intervention based on the specific skill need. If the student demonstrated a lack of progress, the team recommends the student receive Tier III support for a more intensive intervention and more frequent progress monitoring. It is estimated that 1-5% of students still do not make gains with Tier II intervention and require Tier III support (M. K. Burns & Gibbons, 2008; McDougal, 2010). If students do not response to Tier III intervention the early intervening team may suggest a different intervention within the same tier. Throughout all tiers, teachers measure student response to intervention with CBM. Once teachers exhaust all possible tier interventions, the early intervening team may make a recommendation for a student to undergo a comprehensive special education eligibility evaluation to determine if the student has an SLD.

SLD Identification

The definition of an SLD lacked clarity since the creation of the term (Fletcher et al., 2007). Definitions changed over the years and no consensus exists. Recent changes to the SLD definition allow states to move away from using the aptitude-achievement discrepancy model when identifying a learning disability (Fletcher et al., 2007). When using the discrepancy model for special education identification, students with low cognitive abilities and low academic skills, often called slow learners, are not classified as students with specific learning disabilities since no discrepancy exists. The IDEIA SLD definition allows professionals to disregard the need for a discrepancy between cognitive abilities and academic skills and focus on the academic deficits of students. The IDEIA definition of an SLD reads as follows:

32

The term specific learning disability means a disorder in one or more of the basic

psychological processes involved in understanding or in using language, spoken

or written, which disorder may manifest itself in the imperfect ability to listen,

think, speak, read, write, spell, or do mathematical calculations. Such term

includes such conditions as perceptual disabilities, brain injury, minimal brain

dysfunction, dyslexia, and developmental aphasia. Such term does not include a

learning problem that is primarily a result of visual, hearing, or motor disabilities,

of mental retardation, of emotional disturbance, or of environmental, cultural, or

economic disadvantage (IDEIA, 2004, Sec. 602(30)(A-C)).

To determine if a student meets the criteria for an SLD, he or she must undergo a

comprehensive psychoeducational evaluation, which, operationally, may be very

different depending on the SLD identification model employed (Fletcher et al.,

2007).Two SLD identification models exist: Psychoeducational Assessment and RTI.

Within the Psychoeducational Assessment model, several approaches exist.

Psychoeducational Assessment Models

Discrepancy model. Despite IDEIA no longer requiring the use of the

discrepancy model, determining a discrepancy between aptitude and achievement is

the most commonly used approached for identifying an SLD (Fletcher et al., 2007).

Students complete standardized, norm-referenced cognitive tests and standardized,

norm-referenced achievement tests in reading, mathematics, writing, and oral language

to assess their academic achievement. Statistical analyses are employed to determine

if there is a significant discrepancy between ability (IQ) and achievement level. Under

this model of SLD identification, evidence of an SLD is present if a student's IQ score is significantly higher than the student's achievement score.

Several statistical analyses exist for the identification of an SLD using the discrepancy model. One such SLD discrepancy model sets a minimum difference of 15 points or one standard deviation between aptitude and achievement scores as the criteria for identification (Kavale, 2002). The regression method is another discrepancy approach for identification of an SLD. Regression methods acknowledge regression to the mean and standard error of measurement (SEM) are important to statistically consider within an aptitude-achievement discrepancy analysis. Regression to the mean is the psychometric phenomenon in which scores tend to move toward the mean when multiple assessments are administered. Further, abstract constructs such as aptitude and achievement cannot be measured perfectly; some error is always attributable to an obtained aptitude or achievement score on a psychoeducational test. Thus consideration of regression to the mean and SEM is psychometrically responsible practice when evaluating whether a significant aptitude-achievement discrepancy exists (Kavale, 2002).

The SEM is calculated from the standard deviation of the test and its reliability estimate and clarifies the scores earned despite statistical regression(Kavale, 2002). The confidence interval, or the statistically calculated range in which the true score might be found, is then calculated using the SEM. The standard error of the estimate (SEE) is then calculated when using the intelligence score to predict the achievement score. The SEE provides a confidence interval around the predicted achievement score. The predicted achievement score and the actual achievement score are

34

compared using the SEE to determine significant differences. Statistically, the

regression model attempts to remedy problems associated with other discrepancy

models that ignore measurement error and/or regression to the mean. As such,

regression models are deemed the most psychometrically defensible of all aptitude-

achievement discrepancy models.

Low achievement model. Another model for determining if a student has an

SLD entails analyzing data to define a student's achievement level (Fletcher et al.,

2007). Professionals measure academic achievement through the use of standardized

achievement tests and/or CBM. This model usually involves establishing a cut point for

low achievement. Cut points can range from standard scores below the 26th to 20th

percentile rank (Fletcher et al., 2007). If the student scores below the cut point, he or

she qualifies as a student with an SLD.

Intra-individual differences model. When professionals use the intra-individual

differences model for determining if a student has an SLD, they focus on the student's

profile of cognitive assessment data. Within the intra-individual differences model,

several theories exist. However, according to all theories within this model, cognitive

weaknesses are the underlying cause of academic difficulties and learning disabilities

(Fiorello & Primerano, 2005; Naglieri, 2003). A person with an SLD has cognitive

strengths; however, he or she also has weaknesses in some core cognitive processes.

Researcher established that specific cognitive weaknesses in core processes are

associated with an SLD (Fiorello & Primerano, 2005; Naglieri, 2003). Cognitive or

neuropsychological assessments help ascertain these cognitive strengths and

weaknesses while CBM and standardized achievement tests help determine academic

weaknesses. A benefit of this model is it allows for differentiation between intrinsic

causes of underachievement (cognitive processing weaknesses) and extrinsic causes

such as social or economic factors. If underachievement is a result of extrinsic factors,

cognitive processing weaknesses would not be present. Proponents of this model also

boast that the identification of the underlying cognitive processing weakness allows for

better treatment planning for students with SLD(Fiorello & Primerano, 2005; Naglieri,

2003;*cf* Burns & Gibbons, 2008; Siegel, 1989).

RTI Model

The RTI model includes a mass screening of all students and the continued

progress monitoring of students experiencing difficulties in certain areas, rather than

only using assessments administered at a single point in time as described in the three

previously-reviewed models. Professionals may use progress monitoring data,

including CBM data, collected throughout the RTI process in isolation, or in addition to

standardized achievement tests to determine a student's level of achievement. In the

RTI model, students receive Tier II and Tier III interventions focused on a problem area.

If the students respond to the intervention, professionals do not suspect that the

students have SLD. If the students do not respond to the intervention by making

progress in the specific area targeted, then academic underachievement is thought to

be due to SLD (M. K. Burns & Gibbons, 2008; McDougal, 2010).

Professionals use all models to identify a student with an SLD across all

academic areas. RTI research is supportive of using it for SLD identification in reading,

but less is known about the use of RTI for SLD identification in math and writing.

Reading Achievement

The five critical components of reading are phonemic awareness, alphabetic principle or phonics, fluency, vocabulary, and comprehension (National Institute of Child Health and Human Development, 2000). Phonemic awareness is the ability to hear and differentiate between individual sounds or phonemes that form spoken words. Letter-sound correspondence, or the association between a letter and a speech sound, is the alphabetic principle or phonics skills. Oral reading fluency (ORF) is the ability to read words quickly and accurately. Recognizing and understanding words is vocabulary. Reading comprehension involves understanding and deriving meaning from text, which is the main objective of reading. The following sections will further describe the major components of reading. The components of reading build upon each other resulting in fluent reading with comprehension.

Phonemic Awareness

Phonemic awareness is the ability to hear and manipulate spoken sounds. Phonemes are individual sounds, or the smallest unit, of oral language. Although there are 26 letters in the alphabet, there are 41-44 phonemes in the English language (Hosp & MacConnell, 2008). Students practice and teachers measure phonemic awareness skills by segmenting sounds in a word or orally blending sounds to form a word. The focus on phonemic awareness skills occurs during preschool, kindergarten, and the beginning of first grade as phonemic awareness skills are a prerequisite for phonics

skills. Prior to learning phonics skills, students learn to discern spoken sounds so they can form relationships between spoken sounds and written letters.

Phonics

Phonics, or the alphabetic principle, refers to letter-sound correspondence or the relationships between letters and phonemes. Students with strong phonemic awareness skills transfer their knowledge of sounds to print. During kindergarten, students begin focusing on basic phonics skills including the production of the sounds of individual letters and blending these sounds into one syllable words. Students focus on producing sounds to common letter combinations, reading words with common letter blends, reading one-syllable words fluently, reading words that cannot be phonetically decoded (i.e., sight words), and reading words accurately and fluently during first grade (Hosp & MacConnell, 2008). Students must have a strong understanding of phonics skills to read whole words before focusing on reading comprehension and developing fluency.

Vocabulary

Vocabulary is an understanding of word meanings. Students learn new vocabulary throughout all grade levels; however, the complexity of words increases as students advance through school. Teachers introduce a few concepts at a time and review previously learned vocabulary (Joseph, 2008). When students do not understand the definition of words, they are predisposed to have difficulty understanding text suggesting vocabulary skills are a requirement for reading comprehension skills.

Fluency

Reading fluency is the automatic decoding of words that does not require mental effort or conscious attention (LaBerge & Samuels, 1974). A reader is fluent if he or she is able to identify letter-sound correspondences efficiently, identify spelling patterns and utilize them to decode words, blend sounds together to form words, and identify words and connect them to text fluently. If a reader is fluent, he or she places his or her attention on the meaning of the text rather than laboriously decoding words.

ORF is the skill with which a person audibly reads text accurately, quickly, and with prosody (Fuchs et al., 2001). Professionals usually measure ORF by calculating rate and accuracy. The number of words read correct per minute is a measure of rate and the accuracy percentage is a measure of the number of words read correct out divided by the total number of words read. Prosody is the act of reading effortlessly with correct rhythm, intonation, and expression. Assessments that consider all faucets of ORF including rate, accuracy, and prosody are stronger predictors of reading comprehension (Valencia et al., 2010). However, prosody is not easily or objectively measured; therefore, it is not often measured in commercially-available tests (Valencia et al., 2010).ORF is currently one of the most widely used assessments to identify students who are at risk for having and/or experience reading struggles (Petscher & Kim, 2011).

Reading Comprehension

Reading comprehension is the ability to understand and recall information read. More specifically, reading comprehension is a process of simultaneously extracting and constructing meaning from written words(Snow, 2002).A student must have all the prerequisite skills, phonemic awareness, phonics, vocabulary, and fluency, to focus on

and extract meaning from the text. Reading comprehension is the main purpose of reading (Fuchs, Fuchs, & Hosp, 2001).

Relationship Between ORF and Reading Comprehension

Research shows that ORF is a strong predictor of reading comprehension (Fuchs et al., 2001; Fuchs et al., 1988; Valencia et al., 2010). More specifically, numerous researchers compared the predictive and concurrent criterion validity of ORF skills to several standardized achievement tests and state assessments (Deno, 2003; Fuchs et al., 2001; Fuchs, et al., 1988; Goffreda & Pedersen, 2009). These studies suggested the relationship between ORF and reading comprehension is strong.

ORF measures are indirect ways to predict reading comprehension skills. ORF measures, specifically the measure of words correct per minute, are sufficient measures of overall reading skills and a way to screen for students who may be at risk for reading failure (Fuchs et al., 2001; Valencia et al., 2010). A review of 14 studies indicated the correlations between the number of words read correctly in one minute and reading comprehension range from .63 to .90 (Marston, 1989). This review supports the contention that ORF assessments strongly correlate with reading comprehension skills.

Another study supports the use of CBM ORF to predict performance on state assessments and standardized group-administered assessments – thus lending credence to the construct validity of CBM ORF as a measure of reading comprehension. Shapiro, Keller, Lutz, Santora, and Hintze (2006) compared ORF CBM scores to the Pennsylvania System of School Assessment (PSSA), the Stanford Achievement Test, Ninth Edition (SAT-9; Harcourt Brace Educational Measurement, 1996), and the Metropolitan Achievement Test, Eighth Edition (MAT-8; Harcourt Brace

Educational Measurement, 2000). One district administered AIMSweb (Edformation, 2005) probes as the ORF CBM measure and the other district, utilized generic probes. The researchers compared the CBM measures completed in the fall, winter, and spring to the PSSA scores earned in the spring. All correlations between the ORF CBM scores and the PSSA scores were statically significant, *p* < .001 (Shapiro et al., 2006). In addition, all correlations between the ORF CBM assessments and SAT-9 and MAT-8 assessments were significant, *p* < .001 (Shapiro et al., 2006). More specifically, the correlation coefficients between the ORF CBM and the Comprehension subtest on the SAT-9 ranged from .47to .74 during the fall, winter, spring and the correlation coefficient between the ORF CBM and the Comprehension subtest on the MAT-8 ranged from .65to .67 during the fall, winter, and spring.

Another study illustrated the relationship of CBM ORF assessments to several different types of reading comprehension assessment methods. The results supported the conclusions that ORF has a stronger relationship with most commercially-available, standardized reading comprehension measures than measures of reading comprehension that employ different methods of assessing the construct (Fuchs et al., 1988). In that study, the researchers used the Reading Comprehension Subtest from the Stanford Achievement Test (SAT; Gardner, Rudman, Karlsen, & Merwin, 1982) as the criterion measure to compare three other direct measures of reading comprehension and a measure of ORF. The sample for this study consisted of 70 boys receiving special education for learning disabilities in the area of reading from grades 4 to 8. Of the students 69% were Caucasian and the remainder were of a minority race.

The first direct reading comprehension measure used in this study was *question answering* (Fuchs et al., 1988). The question answering technique required students to read two 400-word passages for five minutes apiece. Once the students read the passages, the evaluators orally presented the students with 10 questions and asked the students to respond orally. The most commonly used reading comprehension assessment technique on commercially standardized tests and classroom assessments is question answering (Fuchs et al., 2001). The criterion validity coefficient for the question answering technique when compared to the SAT was .82 (Fuchs et al., 1988).

Another direct reading comprehension measure used in this study was *passage recall* (Fuchs et al., 1988).Students read one of the 400-word passages over a five minute time period. Then they orally retold what they remembered from the passage during a 10 minute timeframe. The criterion validity coefficient for the passage recall method with the SAT was .70 (Fuchs et al., 1988).

The *cloze* method was another direct reading comprehension measured used in this study (Fuchs et al., 1988). Within each 400-word passage the researchers deleted every 7th word, and the students filled in the blank spaces within 10 minutes. Researchers defined correct answers as restored blanks of exact matches of words deleted, synonyms of the words deleted, and words from the same syntactic word class. The criterion validity coefficient for the cloze method with the SAT was .72 (Fuchs et al., 1988).

In addition to comparing the SAT to direct measures of reading comprehension, the researchers compared the SAT to an ORF measure (Fuchs et al., 1988). The students read the two 400-word passages and the test administrators calculated the

mean number of words read correctly per minute across both passages. The criterion validity coefficient for the ORF measure was .91 (Fuchs et al., 1988). Although the direct measures of reading comprehension correlated strongly with the criterion measure, the strongest correlation existed between the ORF scores and the Reading Comprehension subtest on the SAT.

A more recent study used a longitudinal study to compare the predictive validity the Dynamic Indicators of Basic Early Literacy Skills, Sixth Edition (DIBELS 6th; Good & Kaminski, 2002) to third grade reading comprehension measures (Munger, 2010). The reading comprehension measures included the Wechsler Individual Achievement Test, Second Edition (WIAT-II; Wechsler, 2001) Reading Comprehension subtest, the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001) Reading Comprehension subtest, and the New York State English Language Arts Test (NYSELA). The sample included 47 first grade students who completed the DIBELS 6th assessment. At the time of follow-up, when the students were in third grade, only 35 students completed the three reading comprehension measures. The DIBELS 6thORF scores and the three measures of reading comprehension validity correlation coefficients ranged from .56 to .72 (Munger, 2010). The DIBELS 6thand the NYSELA resulted in the lowest correlation coefficient of .56 whereas the DIBELS 6th and the GRADE Reading Comprehension subtest resulted in the highest correlation coefficient of .72. The correlation coefficient for the predictive relationship between the DIBELS 6th and the WIAT-II Reading Comprehension subtest was .65.The above mentioned study (Munger, 2010) supported not only the relationship between ORF and state and group administered assessments (Fuchs, Fuchs, & Maxwell, 1988; Shapiro et al., 2006), but

also the relationship between ORF and an individually-administered, standardized assessment. More importantly, Munger (2010) investigated the relationship between the previous editions of the assessments used in the current study (DIBELS Next and WIAT-III).

Developmental Level and Reading

Reading acquisition is developmental, meaning as students move through the grades their skills improve. As previously mentioned the components of reading serve as building blocks with phonemic awareness, phonics, and vocabulary building a foundation for fluency and comprehension (National Institute of Child Health and Human Development, 2000). Once students master the building blocks of reading, they are able to read more fluently and concentrate more on the meaning of the words or compression of information (LaBerge & Samuels, 1974).

More pertinent to this study, the mean DIBELS Next ORF Words Correct (WC) and Accuracy Percentage scores increase across elementary school grades despite the increased difficulty level of the passages (Good et al., 2011).   These increasing mean scores support the claim that reading is a developmental construct. On the WIAT-III, each grade level has certain start points for each subtest so students do not begin at a level of difficulty too far above or below their suspected skill level based on their grade level (Breaux, 2009). In addition, evaluators administer grade specific item sets to students on the ORF subtest and Reading Comprehension subtest (Breaux, 2009). These administration procedures reaffirm reading is a developmental construct.

Reading Assessment

CBM

Researchers originally developed CBM to measure the effectiveness of data-based program modification (DBPM), which is a special education intervention model (Deno & Mirkin, 1977).Teachers used repeated measurement data to evaluate the effectiveness of their instruction when operating under this model. Researchers conducted a federally-funded study at the University of Minnesota Institute for Research on Learning Disabilities to test the use of DBPM (Davenport et al., 2002). A major result of that study was the development of generic progress monitoring procedures in reading, spelling, and written expression. The procedures developed included: core outcomes tasks to measure performance, stimulus items, measurement procedures and scoring guidelines, and decision rules to improve academic instruction. Special education teachers initially utilized CBM to monitor their effectiveness, write IEP goals, and monitor special education student progress (Deno, 2003; Shinn, 2008).

Over the years, the use of CBM expanded to the general education classroom (Deno, 2003).Now teachers commonly administer CBM to all students as universal screeners to identify students in need of support. CBM may predict performance on important criteria such as student proficiency or high-stakes testing and enhance instructional planning by allowing teachers to analyze the data and decide what areas of student need exist. Teachers also use CBM to monitor the progress of students receiving early intervening interventions.   CBM also offer alternative special education identification measures which are more common since IDEIA changed the SLD identification criteria to include responsiveness to instruction as one set of classification criteria (IDEIA, 2004). Years after the creation of CBM, a commercial available tool, the

Dynamic Indicators of Basic Early Literacy Skills (DIBELS;Good et al., 2011) is commonly used in schools.

DIBELS Next. The DIBELS Next, the latest edition of the DIBELS series, includes several assessments measuring different aspects of reading and pre-literacy skills(Good & Kaminski, 2011). The assessments include First Sound Fluency, Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, ORF, and DAZE. The First Sound Fluency assessment is a measure of phonological awareness that assesses a child's ability to recognize and produce the initial sound in an orally presented word. The Letter Naming Fluency test involves presenting the student with a page of upper- and lower-case letters arranged in a random order and asking the student to name as many letters as he or she can. The Phoneme Segmentation measure assesses a student's ability to fluently segment three- and four-phoneme words into their individual phonemes. The Nonsense Word Fluency assessment measures the student's knowledge of the letter-sound correspondence in which letters represent their most common phonemes and the ability to blend letter-sound combinations into words. ORF assessment measures reading speed and accuracy. The DAZE assessment measures reading comprehension by presenting a student with a passage including words deleted, and the student fills in the blanks by choosing one of the words offered as possible answers.

On the DIBELS Next ORF assessment, students have one minute to read a short passage. The evaluator totals the number of words read correctly and compares that number to normative benchmarks to measure reading rate in comparison to same-grade peers resulting in a WC score. To measure accuracy, the test administrator

calculates the number of words read correctly out of the total number of words attempted in one minute resulting in an Accuracy Percentage score (Good et al., 2011). As mentioned previously, early intervening teams commonly use ORF CBM, including the DIBELS Next ORF assessment, to determine the need for a recommendation for a special education eligibility evaluation.

Standardized Individual Academic Assessment

Numerous standardized academic achievement tests exist; however, school psychologists traditionally administer one of three achievement assessments (T.G. Burns, 2010): Wechsler Individual Achievement Test, Second Edition (WIAT-II; Wechsler, 2001), Woodcock-Johnson Tests of Achievement-Third Edition (Woodcock, McGrew, & Mather, 2001), and the Kaufman Tests of Educational Achievement-Second Edition (Kaufman & Kaufman, 2004). Professionals more commonly administer the Wide Range Achievement Test, Third and Fourth Editions to adults (T. G. Burns, 2010).

The Wechsler Individual Achievement Test, Third Edition (WIAT-III), the newest form of the WIAT, aims to allow for more in-depth subtest analysis along with intervention recommendations. The latest addition includes an updated standardization sample to allow for appropriate norm-referenced comparisons with contemporary students. The WIAT-III provides subtests that correspond with the specific academic areas where an SLD could exist according to IDEIA forming a direct link with federal eligibility law. When a student experiences academic issues, the WIAT-III is a common assessment used during a special education eligibility evaluation (T. G. Burns, 2010).

WIAT-III. The development of the WIAT (Psychological Corporation, 1992) began in 1992. The four main purposes of the WIAT were to establish norm-referenced

achievement measures in eight academic areas, provide an assessment of strengths and weakness, compare cognitive and academic achievement abilities, and assess progress in academic programs such as special education (Psychological Corporation, 1992). Wechsler published two more editions of the WIAT since the origination of the test. The most current edition is the WIAT-III (Wechsler, 2009).

The WIAT-III is an academic diagnostic assessment used with children from ages four though 19 (Breaux, 2009). Professionals individually administer this assessment to students under standardized conditions. The WIAT-III consists of 16 individual subtests which measure listening, speaking, reading, writing, and mathematics skills.

Due to the methodology of the current study, a detailed description of the WIAT-III ORF and Reading Comprehension subtests is offered. Students read certain passages based on their grade level for both the ORF and Reading Comprehension subtests. The WIAT-III ORF subtest requires that students read two short passages aloud. The test administrator then measures the time to complete each passage and oral reading errors made, to calculate an ORF subtest score. The ORF subtest score contains two subcomponents: ORF Accuracy and ORF Rate (Breaux, 2009). When calculating the ORF Accuracy score, the evaluator totals the number of words the student inserts into the passage and the number of words read incorrectly. The ORF Rate score represents the amount of time it took for the student to read the passage. The ORF subtest results in three standard scores: ORF subtest score, ORF Accuracy score, and ORF Rate score. The WIAT-III Reading Comprehension subtest measures reading comprehension by having students read three passages and answer

accompanying open-ended questions (Breaux, 2009). The Reading Comprehension

subtest results in one standard score.

DIBELS Next ORF and WIAT-III Administration and Scoring Procedures

Both the DIBELS Next ORF and WIAT-III ORF assessments mentioned above

measure ORF with standardized procedures. However, the standardized procedures

differ among the two assessments. See Table 2 for the specific administration and

scoring procedure similarities and differences.

Table 2

Administration and Scoring Procedures for the DIBELS Next ORF and WIAT-III ORF

	DIBELS Next ORF	WIAT-III ORF
Administration Procedures		
Time allowed	1 minute	Completion of passage
Number of passages read and scored	1	2
Reverse to lower level passages if poor performance	No	Yes
Scoring Procedures		
Supplied words by evaluator	After 3 second wait	After 5 second wait
Inserted words	No error	Error
Repeated words	No error	No error
Substitutions/ mispronounced words	Error	Error

| Transpositions | Each word transposed is 1 error | Each transposition is 1 error despite the number of words |
| Self-corrections | No error | No error |

Note. DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WIAT-III = Wechsler Individual Achievement Test, Third Edition.

Rationale for the Study

There are two primary approaches used to identify an SLD: Psychoeducational Assessment and RTI. Within both approaches, early intervening teams use CBM data, such as the DIBELS Next, to identify which students should undergo a special education eligibility evaluation.

Professionals who use the Psychoeducational Assessment approach analyze extant data and use individually-administered standardized tests, such as the WIAT-III, when determining eligibility for special education. If this study supports that the DIBELS Next ORF assessment is a reasonable predictor the WIAT-III ORF and Reading Comprehension subtest scores, early intervening teams will have stronger empirical justification for using the DIBELS Next ORF scores to make important decisions including a referral for a special education eligibility evaluation. If this study does not support the DIBELS Next ORF assessment as a reasonable predictor of WIAT-III ORF and Reading Comprehension subtest scores, early intervening teams may minimize the importance placed on DIBELS Next ORF scores when deciding whether or not students should undergo a special education eligibility evaluation.

The RTI approach is increasing in popularity due to its prevention-based and early intervention qualities. RTI requires universal screening and progress monitoring of skills. One common commercially-available tool used for universal screening and progress monitoring is the DIBELS Next. Some professionals use the DIBELS Next

data in conjunction with standardized tests such as the WIAT-III.   However, other

professionals use only DIBELS Next data when determining if a student meets the

criteria for special education eligibility as evidencing an SLD. The sole use of DIBELS

Next data to make special education eligibility decisions is somewhat controversial

given the unknown psychometric merits of the DIBELS Next with a well-established

assessment such as the WIAT-III. If this study supports validity of DIBELS Next ORF

assessment professionals will have stronger empirical justification for using the DIBELS

Next ORF scores to make special education eligibility decisions. If this study does not

support the validity of the DIBELS Next ORF assessment professionals may need to

consider using additional data when making special education eligibility decisions under

the RTI model.

Whether professionals use the DIBELS Next as a progress monitoring tool for a

referral for a special education eligibility evaluation or for diagnostic impressions of an

SLD, strong research should support the construct validity of DIBELS Next to

established criteria such as the WIAT-III. The main purpose of this study is to

investigate the predictive strength of the DIBELS Next ORF assessment to the WIAT-III

ORF and Reading Comprehension subtests. The researcher also evaluated differences

of predictive strength between groups based developmental level.

CHAPTER 3

METHODOLOGY

Overview

The purpose of this study was to examine the predictive strength of the Dynamic Indicator of Basic Early Literacy Skills, Next Edition (DIBELS Next; Good et al., 2011) to the Oral Reading Fluency (ORF) assessment to the Wechsler Individual Achievement Test, Third Edition (WIAT-III; Wechsler, 2009) ORF and Reading Comprehension subtests. The researcher also analyzed differences of predictive strength between groups based on developmental level. This chapter includes an explanation of the research design, sample, and population used in this study. This chapter also includes a description of the procedures used for data collection along with the types of measurements including their reliability and validity. Lastly, the chapter concludes with a narrative regarding statistical analyses employed.

Research Design

The researcher utilized a correlational design to observe whether a relationship exists between variables. More specifically, the researcher examined the correlation between the DIBELS Next ORF assessment and the WIAT-III ORF subtest and WIAT-III Reading Comprehension subtest.

Population

The population for this study includes second through fifth grade students who are referred for, or participated in, special education eligibility evaluations. However, the employment of a convenience sample limits generalizability to a population of students referred for special education eligibility evaluations. While the sample is

limited to second through fifth grade students, these are the grades during which most students are evaluated for and found eligible for special education services.

Study Site

The study site is a school district in Delaware with 14 elementary schools. The school district requires that all elementary school teachers use the Scott Foresman Reading Street (Blachowicz, 2008) core reading program. The Scott Foresman Reading Street program provides explicit, systematic, instruction in all critical elements in reading: phonemic awareness, phonics, fluency, vocabulary, and comprehension. The school district also requires that all students must receive a 90-minute uninterrupted block of time for reading instruction each day.

The 14 elementary schools in the school district employ 10 school psychologists, most of whom use the WIAT-III as a component of their psychoeducational evaluation batteries. All first through fifth grade students receiving Tier II and Tier III reading support in Response to Intervention (RTI), complete the DIBELS Next to monitor their progress. Students who receive intervention for an ORF deficit complete the DIBELS Next ORF assessment to monitor their progress.

For the 2011-2012 school year (the year the assessments in this study were completed), the demographic composition of the collaborating district's student body was 47.9% Caucasian, 23.1% African American, 22.2% Hispanic/Latino, 5.7% Asian, and 1.1% classified as other. The student body was 49.2% female, 52.4% of students in the district received free/reduced-priced lunch, and 11% received special education services. Students in the school district complete a state assessment in both reading and mathematics yearly in 3rd to 10th grades. The percentage of the student body that

met state standards in reading ranged from 63.8% to 76.9% and in math ranged from 62.9% to 76.6% (Delaware Department of Education, 2013).

Sample

The sample consisted of second through fifth grade students referred for special education eligibility evaluations during the 2011-2012 school year. The researcher excluded kindergarten students from the sample because the DIBELS Next ORF and the WIAT-III ORF and Reading Comprehension subtests do not include norms for students at this grade level. First grade students do not participate in the DIBELS Next ORF assessment until January of each school year, and therefore, the researcher excluded first grade students from this study. Both males ($N = 42$) and females ($N = 33$) were included in this study. The sample only included students referred for special education eligibility evaluations since the WIAT-III assessment is a comprehensive standardized assessment tool usually only used to determine special education eligibility.

Measurement

When analyzing the measurement tools for this study, the researcher considered the reliability and validity of each measure. Several types of reliability, which is the stability of measurement across a variety of conditions, exist (Drost, 2011). Test-retest reliability is the measure of consistency between more than one test administration. Alternative-form reliability measures the consistency of scores when using different measurement instruments. Internal consistency measures the stability within an instrument. There are three common approaches to measure internal consistency: inter-rater, spit-half, and Cronbach's Alpha. When more than one person rates or

judges behavior, researchers use inter-rater reliability as a measure of the degree of agreement among raters. The spilt-half approach involves separating the items on a test into two half tests and comparing the correlation between the half tests. The Cronbach's Alpha is a method used to measure item-specific variance.

The level of acceptable reliability measures differs depending on the context.   A reliability correlation coefficient of .80 is a minimum standard used for educational screening decisions with the consensus interpretation that a reliability correlation coefficient of .90 is the minimal standard for using data to make important educational decisions, such as placement in special education programs (Salvia & Ysseldyke, 2007).

Validity refers to the extent to which an instrument measures what it is intended to measure (Drost, 2011). Construct validity refers to the degree in which a test or instrument measures what it claims to measure. Content validity determines if indicators represent the meaning of a concept as defined by the researcher. Criterion-related validity is the amount of correspondence between two test measures, one of which serves as the criterion or principle measure. When the measure and the criterion exist at the same time this is a measure of concurrent validity; whereas, when the criterion is in the future this is a measure of predictive validity.

Researchers evaluate criterion-related validity and reliability by calculating correlation coefficients, or Person's *r*. Correlation coefficients are statistical measures of the linear relationship, including the direction and strength of the relationship, between two or more variables (Stevens, 2002). A positive correlation coefficient suggests that when one variable increases the other variable increases. A negative

55

correlation coefficient suggests the variables have an inverse relationship meaning that as one variable increases the other variable decreases. The closer the correlation coefficient is to │1.0│, the stronger the relationship.

Hopkins (2000) established qualitative descriptions for the interpretation of criterion-related validity correlation coefficients. See Table 3 for the criterion-related validity correlation coefficient ranges.

Table 3

Correlation Coefficient Ranges

Range	Correlation Coefficients
Strong	Above .70
Moderate-Strong	.50 to .69
Moderate	.30 to .49
Small	.10 to .29
Very Small	Less than .09

DIBELS Next ORF

The DIBELS Next (Good et al., 2011) is an assessment that measures early literacy skills of students in kindergarten through 6th grade. The DIBELS Next consists of six tests including First Sound Fluency, Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fleuncy, ORF, and DAZE. Students complete certain assessments based on their grade level. The focus of this study is on the DIBELS Next ORF assessment.

The DIBELS Next ORF measures ORF skills or the speed and accuracy of reading. The DIBELS Next ORF results in two scores: Words Correct (WC) and

Accuracy Percentage (Good et al., 2011). The DIBELS Next WC score is the metric

used to measure oral reading rate and the DIBELS Next ORF Accuracy Percentage is

the metric used to measure oral reading accuracy in this study. On the DIBELS Next

ORF assessment, students have one minute to read a short passage aloud. When a

student skips a word, reads a word incorrectly, or if the evaluator pronounces a word for

the student after waiting three seconds for the student to attempt the word, the

evaluator counts the word as an error. When a student pronounces a word correctly

without assistance from the evaluator, the evaluator records the word as read correctly.

At the conclusion of the assessment, the evaluator totals the number of words read

correctly, resulting in the DIBELS Next ORF WC score. The evaluator calculates the

DIBELS Next ORF Accuracy Percentage score by dividing the number of words read

correctly by the total number of words the reader attempted to read and multiplies it by

100. Educators then compare the WC score and Accuracy Percentage scores to

normative benchmarks to measure reading rate and accuracy in comparison to same

grade peers to determine if the student exhibited grade-level skills.

In order to control the difficultyof the passages for each grade level, the authors

of the DIBELS Next used the Dynamic Measurement Group (DMG) Passage Difficulty

Index (Cummings, Wallin, Good, & Kaminski, 2007). The DMG Passage Difficulty Index

inspects three features of passage difficulty: word difficulty, semantic difficulty, and

syntactic difficulty. To measure the word difficulty for a passages, the characters per

word and syllables per word were examined. The median words per sentence supplied

a measure of syntactic difficulty. The percent of unique words provided a measure of

semantic difficulty. Each of these facets of passage difficulty are examined in

isolation,and in combination, to ensure that they remain within a specific range.

Passages were then validated for use at each grade level by administering 40 grade level passages to students. Of those 40 passages per grade level, outlier passages were removed and passages with the highest relaibility and valdity were selected for publication (Powell-Smith, Good, & Atkins, 2010).

DIBELS Next ORF reliability. Both components of the DIBELS Next ORF assessment, WC and Accuracy Percentage, have adequate reliability (Good et al., 2011). One study measured alternative-form, test-retest, and inter-rater reliability (Powell-Smith et al., 2011). Researchers used stratified sampling based on the middle of the year benchmark to select approximately 50% of students who reached benchmark and 50% of students who fell below benchmark to test all reliabilities. Students completed testing two weeks after the middle of the year DIBELS Next benchmark assessment. The sample used to measure the alternative-form reliability consisted of 187 students. The DIBELS Next ORF WC alternate-form reliabilities from the same grade levels ranged from .95 to .97 and the DIBELS Next ORF Accuracy Percentage alternate-form reliability from the same grade levels ranged from .76 to .85.

Good et al. (2011) further analyzed the stability of DIBELS Next ORF scores using a sample of 152 students. The DIBELS Next ORF WC test-retest reliability, for grades 2 through 5, ranged from .91 to .97 and the DIBELS Next ORF Accuracy Percentage test-retest reliability, for grades 2 through 5, ranges from .57 to .97.

Inter-rater reliability of DIBELS Next ORF was also completed using data from 122 students and two raters. The inter-rater reliability for DIBELS Next ORF WC from

grades 2 through 5 was .99 and the DIBELS Next ORF Accuracy Percentage inter-rater reliability from grades 2 through 6 ranged from .85 to .99 (Powell-Smith et al., 2011).

Overall, The DIBELS Next ORF WC alternate-form and test-retest reliabilities across all grade levels met the standard used for educational screening decisions; whereas, the DIBELS Next ORF Accuracy Percentage alternative-form and test-retest reliabilities meet the educational screening standard across most, but not all, grades. The inter-rater reliabilities for both the DIBELS Next ORF WC and Accuracy Percentage scores met the educational screening criteria across all grade levels. In conclusion, previous research supports the contention that DIBELS Next meets or exceeds the educational screening standard in most cases. Therefore, the DIBELS Next ORF is an appropriate instrument to use in this study.

DIBELS Next ORF validity. To create the DIBELS Next, researchers conducted a study to gather the necessary information to set benchmark goals and test the reliability and validity of the DIBELS Next measure (Powell-Smith et al., 2011). All students (N = 3,816) from 13 schools across five districts from five different states in the North Central Midwest and Pacific regions of the United States participated in the benchmark assessment portion of the study; however, only a subset of those students (N = 1,306) participated in the criterion-related validity portion of the study. Researchers used the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001) and the Standard 4th Grade Reading Passage in the National Assessment of Education Progress (NAEP) 2002 Special Study of Oral Reading (Daane, Campbell, Grigg, Goodman, & Oranje, 2005) to measure the criterion-related validity of the DIBELS Next ORF WC and Accuracy Percentage assessments.

Students completed the GRADE at the end of the school year, and therefore, the beginning of the year and middle of the year DIBELS Next ORF benchmark scores served as the predictive criterion-related validity measures. The predictive criterion-related validity of the fall DIBELS Next ORF WC score to the GRADE Total Test score ranged from .64 to .77 (moderate-strong to strong ranges). The middle of year DIBELS Next ORF Benchmark Assessment evidenced predictive correlations of.59 to .77 (moderate-strong to strong) to the GRADE Total Test score (Good et al., 2011). The predictive validity of the fall DIBELS Next ORF Accuracy Percentage score to the GRADE Total Test score ranged from .53 to .75 (moderate-strong to strong), with winter DIBELS Next ORF Accuracy Percentage predictive correlations to GRADE Total Test scores ranging from .47 to .80(moderate to strong; Good et al., 2011).

The end of the year DIBELS Next ORF benchmark scores served as the concurrent validity measures in this study. The concurrent validity of the DIBELS Next ORF WC score with the GRADE Total Test score, administered at the same time period, ranged from .61 to .75 (moderate-strong to strong; Good et al., 2011).Similarly, the concurrent validity of the DIBELS Next ORF Accuracy score and the GRADE Total Test score ranged from .49 to .73 (moderate to strong ranges; Good et al., 2011). Overall, the predictive and concurrent criterion-related validity for the DIBELS Next ORF WC ranged from .59 to .77 (strong-moderate to strong) and for the DIBELS Next ORF Accuracy Percentage ranged from .49 to .80 (moderate to strong) when compared to the GRADE Total Test score (Good et al., 2011).

Another researcher compared the concurrent and predictive validity the Dynamic Indicators of Basic Early Literacy Skills, Sixth Edition (DIBELS 6th; Good & Kaminski,

2002) to third grade reading comprehension measures (Munger, 2010). The reading

comprehension measures included the Wechsler Individual Achievement Test, Second

Edition (WIAT-II; Wechsler, 2001) Reading Comprehension subtest, the GRADE

(Williams, 2001) Reading Comprehension subtest, and the New York State English

Language Arts Test (NYSELA). The first grade DIBELS 6th ORF scores and the three

third grade measures of reading comprehension predictive validity correlation

coefficients ranged from .56 to .72 (Munger, 2010). The DIBELS 6th and the NYSELA

resulted in the lowest correlation coefficient of .56; whereas, the DIBELS 6th and the

GRADE Reading Comprehension subtest resulted in the highest correlation coefficient

of .72. The correlation coefficient for the predictive relationship between the DIBELS 6th

and the WIAT-II Reading Comprehension subtest was .65.Munger (2010) also

investigated the concurrent criterion-related validity of the DIBELS Next ORF in this

study. The correlations between the third grade DIBELS NEXT ORF WC and the

reading comprehension measures ranged from .70 to .85. A .77 correlation coefficient

was found for the relationship between the third grade DIBELS Next ORF WC and the

WIAT-II Reading Comprehension subtest.

According to Hopkins' (2000) validity correlation coefficient ranges, the validity

correlation coefficients for the DIBELS Next ORF fell within the moderate to strong

range. These results suggest that the DIBELS Next ORF measure has adequate

criterion-related predictive and concurrent validity with the GRADE (Good et al., 2011;

Munger, 2010), predictive validity for a state assessment(Munger, 2010), and predictive

validity for the WAIT-II(Munger, 2010). Therefore, the DIBELS Next ORF is an

appropriate instrument to use in this study.

WIAT-III ORF Subtest

The WIAT-III is an academic diagnostic assessment administered to children ages four though 19 (Breaux, 2009). The WIAT-III contains subtests that assess reading, mathematics, writing, and oral language skills. The focus of this study was on two of the reading subtests: the ORF subtest and the Reading Comprehension subtest. The WIAT-III ORF subtest measures oral reading fluency skills, including reading rate and accuracy. The WIAT-III ORF subtest consists of three scores: ORF subtest, ORF Rate, and ORF Accuracy. The appropriateness of each grade level passage was validated by using readability formulas to estimate the reading level of each ORF passage. Additionally, data were collected from students at various grade levels to confirm that the levels of the passages were appropriate (Breaux, 2009).

The WIAT-III ORF subtest requires that students read two short passages until completion. The test administrator measures the time to complete each passage and words read incorrectly to calculate an ORF subtest score. The WIAT-III ORF subtest score provides an overall assessment of ORF. Students complete items based on the students' grade levels resulting in different item sets administered to different students. Since raw scores from different item sets are not comparable as scores are based on the specific item sets administered, raw scores must be converted to weighted scores. Because this is a timed subtest and item-level data cannot be analyzed, vertical scaling was completed using the common-person design and the equal percentile method (Breaux, 2009). Vertical scaling is a method used to place scores across different grade level assessments on a common scale. The common-person design was used to equate scores by administering the different item sets to the same group. Students read

grade-level item sets and the two preceding item sets. The researcher analyzed the mean scores across adjacent item sets and score discrepancies to guarantee that the scores of each student across different item sets were consistent. Since the scores across different item sets were consistent, professionals are able to compare scores on different item sets.

The WIAT-III ORF Rate score is a component of the ORF subtest score which measures reading speed (Wechsler, 2009). How quickly a student reads the passage is the basis for the WIAT-III ORF Rate score. Pauses while reading, self-corrections, and repetitions affect the ORF Rate score. The WIAT-III ORF Accuracy score is a component of the ORF subtest score which measures reading accuracy. The words read correctly out of the total words read results in the WIAT-III ORF Accuracy score. The ORF Accuracy score accounts for errors and words read incorrectly. Again, researchers used vertical scaling (Breaux, 2009) to determine ORF Rate and Accuracy scores.

WIAT-III ORF reliability. Breaux (2009) established the WIAT-III ORF subtest test-retest reliability by administering the WIAT-III twice within 2-32 days. The sample consisted of 161 prekindergarten through grade 12 students, of which 50.9% were female. The WIAT-III ORF subtest demonstrated a test-retest age-based reliability coefficient of .93. The WIAT-III ORF Rate score had a test-retest age-based reliability coefficient of .93. The WIAT-III ORF Accuracy score had a test-retest age-based reliability coefficient of .83 (Breaux, 2009).   The study supported that reliability coefficients for the WIAT-III ORF scores exceeded the minimal screening standard.

63

Based on previous research, the WIAT-III ORF subtest demonstrates reliabilities appropriate for consideration in this study.

WIAT-III ORF validity. The validity of the WIAT-III ORF subtest, along with its two components (Rate and Accuracy) is understood to be adequate due to the extensive amount of research conducted prior to the publishing of the test (Breaux, 2009). The validity of the WIAT-III was measured against the WIAT-II; however, the previous editions of the WIAT did not include the ORF subtest. Prior research investigating the validity of the WIAT-III ORF subtest, however, could not be located. Although no research supporting the adequacy of the validity of the WIAT-III ORF subtest is available, its scores are assumed to be valid due to the extensive research conducted on other, similar measures of ORF. Therefore, the WIAT-III ORF subtest is an appropriate assessment tool to use in this study.

WIAT-III Reading Comprehension Subtest

The WIAT-III Reading Comprehension subtest measures the ability to recall information the student has read (Breaux, 2009). Test administrators derive a score by asking a student questions about the story he or she has read immediately after reading the story. Evaluators administer items to students based on their grade level. Since raw scores from different item sets are not comparable, scores are based on the specific item sets, and therefore, raw scores must be converted to weighted scores.

WIAT-III Reading Comprehension reliability. Similar to the WIAT-III ORF subtest, researchers evaluated the test-retest reliability of the Reading Comprehension subtest by administering the WIAT-III twice within 2-32 days (Breaux, 2009). Again, the sample consisted of 161 students, with 49.1% of the sample across all prekindergarten

through grade 12students identified as male. The WIAT-III Reading Comprehension subtest had a test-retest reliability coefficient of .86 (Breaux, 2009).Although the study did not demonstrate that the WIAT-III Reading Comprehension score exceeded minimal standard for using data to make important educational decisions (e.g., special education eligibility), it exceeded the minimal standard used to make screening decisions. The researcher determined that the WIAT-III Comprehension subtest is an appropriate instrument to use in this study.

WIAT-III Reading Comprehension validity. The validity of the WIAT-III Reading Comprehension subtest is adequate due to the extensive amount of research conducted prior to the publishing of the test (Breaux, 2009). Researchers investigated the validity of the subtests on the WIAT-III by analyzing the correlations of its subtests with subtests on the WIAT-II. The sample consisted of 140 students across all grade levels (Prekindergarten - grade 12). Students completed the WIAT-II and WIAT-III counterbalanced with 1 to 30 days in between the two assessments. The WIAT-II Reading Comprehension subtest and the WIAT-III Reading Comprehension subtest had a .69 correlation (Breaux, 2009). According to Hopkins' ranges, the validity of the WIAT-III Reading Comprehension subtest falls into the strong range. Therefore, the WIAT-III Reading Comprehension subtest is an appropriate assessment to use in this study.

Procedures

Archival data used in this study were obtained from a school district in Delaware. All second through fifth grade students receiving Tier II and Tier III reading support in RTI complete the DIBELS Next every two weeks to monitor their progress. Students receiving intervention for an ORF deficit complete the DIBELS Next ORF assessment to

monitor their progress. When students undergo a special education eligibility

evaluation, they participate in an individualized standardized achievement test. The

researcher included data from students who completed the DIBELS Next ORF and

WIAT-III assessments during the 2011-2012 school year. The researcher asked all of

the school psychologists working in elementary schools in the district to participate in

the data collection for this study. The school psychologists who agreed to participate

gathered the DIBELS Next ORF data collected just prior to the administration of the

WIAT-III ORF and Reading Comprehension subtests. The DIBELS Next progress

monitoring probe should have been administered no more than two weeks prior to the

WIAT-III. In the school district where the data were collected, teachers administer the

DIBELS Next ORF probes every two weeks to students who are receiving Tier II and

Tier III support. Students who undergo special education eligibility evaluations are

usually receiving Tier II or Tier III support prior the referral, with the only exception being

if the referral was a parent request.

The school psychologists who work in each school, and therefore, have access

to the data, redacted the data prior to presenting it to the researcher. The data

collected included the students' grade level, sex, DIBELS Next ORF WC score, DIBELS

Next ORF Accuracy Percentage score, WIAT-III ORF Rate score, WIAT-III ORF

Accuracy score, WIAT-III ORF subtest score, WIAT-III Reading Comprehension subtest

score, and administration dates for both assessments. The district psychologists

removed students' names from the data and each student received a randomized

identification code. The researcher assigned each psychologist the first number of the

identification code to avoid the duplication of random identification codes. Each school

psychologist maintained a separate list of identification codes in case the researcher

needed to clarify missing or erroneously entered data. The school psychologists

maintained the lists securely on password-protected computers, allowing no one else

access to the information. The researcher did not have access to personally-identifying

information of the students. The school psychologists sent the archived, anonymous

data via secure email to the researcher.

To move forward with the procedures described above, the researcher obtained

permission from both the district and the university. At the district level, the researcher

submitted a Research Approval Form to the Department of Research and Evaluation.

The researcher also completed The Indiana University of Pennsylvania Institutional

Review Board (IRB) process to gain research approval (see Appendix A).

Sample Size

This study included second through fifth grade students who participated in a

special education eligibility evaluation and completed the DIBELS Next ORF, WIAT-III

ORF subtest, and Reading Comprehension subtest. These students attended 14

elementary schools in one school district in Delaware. There were 75 participants in

this study; however, the number of scores for each assessment varied. The reason for

the variance is that psychologists choose the assessment they use during a special

education evaluation based on student needs and not all students are given all

assessments. In order for student data to be used for this study, the student must have

DIBELS Next ORFWC and Accuracy Percentage scores along with at least one WIAT-

III ORF Rate, ORF Accuracy, or Reading Comprehension score. A statistical

recommendation requires approximately 15 participants per predictor for a reliable

regression equation that will cross-validate with little loss in predictive power (Stevens, 2002). The current study met this recommendation. See Table 4 for the sample size for each assessment by grade level.

Table 4

Sample Size by Assessment and Grade Level

Assessment Score	Total	2nd Grade	3rd Grade	4th Grade	5th Grade
DIBELS Next ORF WC	75	19	20	17	19
DIBELS Next ORF Accuracy Percentage	75	19	20	17	19
WIAT-III ORF Rate	70	19	18	15	18
WIAT-III ORF Accuracy	70	19	18	15	18
WIAT-III Reading Comprehension	67	16	19	17	15

Note. DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct; WIAT-III = Wechsler Individual Achievement Test, Third Edition; 2nd grade = 11 males and 8 females; 3rd grade = 11 males and 9 females; 4th grade = 9 males and 8 females; 5th grade = 11 males and 8 females.

Statistical Analyses

The researcher calculated descriptive statistics including means, standard deviations, medians, skewness, kurtosis, and ranges for scores on the DIBELS Next ORF, WIAT-III ORF subtest, and Reading Comprehension subtest. Prior to analyzing the data to address the research question, the researcher used a MANOVA to investigate whether or not differences existed by sex. The researcher used multiple linear regression (MLR) to evaluate how well the DIBELS Next ORF scores predicted the WIAT-III scores. The researcher also used regression analysis to measure the differences between groups (developmental level) on all measures.

MANOVA is a statistical analysis that measures the differences between two or more groups of participants in regards to several dependent variables. Particularly,

researchers use a MANOVA if the dependent variables measure a similar skill or concept (Stevens, 2002). In this study, the researcher grouped the participants by sex; the dependent variables are the DIBELS Next (WC and Accuracy Percentage) and WIAT-III (ORF Rate, ORF Accuracy, and Reading Comprehension) scores in the MANOVA model.

When using a MANOVA, certain assumptions exist (Stevens, 2002). The first assumption is that the data must be ratio or interval data. Interval data includes categories with ranks. Ratio data consists of the same characteristics of interval data; however, ratio data also have an anchor point. The second assumption is the data are normally distributed. Comparing the histograms for each variable to a normal curve and evaluating skewness and kurtosis allows for the verification of normality. A rule of thumb states that if skewness is greater than negative one or positive one the distribution is highly skewed, between negative one-half and negative one or between positive one-half and positive one the distribution is moderately skewed, and if the skewness is between negative one-half and positive one-half the distribution is approximately symmetric(Bulmer, 1979). A normal distribution has a kurtosis of exactly three (Bulmer, 1979). The next assumption is the homogeneity of the covariance matrices. The test used to determine if this assumption is met is Box's M.   The Box's M tests if the covariance matrices of the dependent variables are significantly different across levels of the independent variables. Having an appropriate sample size, is another assumption. Researchers recommend about 15 participants per predictor for a reliable regression equation that will cross-validate with little loss in predictive power (Stevens, 2002). The last assumption is the independence of variables. No relation

should exist on the dependent measures to other scores. The Durbin-Watson test can be used to check this assumption. The variable, sex, analyzed in this study using a MANOVA, met the needed assumptions.

MLR is a statistical procedure used when a researcher is analyzing the prediction of one dependent variable by using multiple predictors. MLR allows for researchers to evaluate the combined prediction of the predictors along with the individual prediction of each predictor (Stevens, 2002). In the current study, the DIBELS Next ORF scores (WC and Accuracy Percentage) serve as predictors for each WIAT-III (ORF Accuracy, ORF Rate, and Reading Comprehension) scores in the MLR model.

Certain assumptions are required when using MLR. Several assumptions including using interval or ratio data, normality, equal variances, independence of variables, and linearity are the same as those needed for MANOVA (Stevens, 2002). However, some statisticians suggest that independence of variables is not a necessary assumption when using MLR (Osborne & Waters, 2002). Two additional assumptions need for MLR are homoscedasticity and non-multicollinearity. Homoscedasticity means that variance of errors is the same across all levels of the variables. The researcher visually analyzed the residual plots to check for homoscedasticity. Non-multicollinearity occurs when two or more predictors are not highly correlated. The researcher verified non-multicollinearity by checking the r-value and removing a variable if necessary. The variables in this study using MLR, including the DIBELS Next ORF and WIAT-III ORF and Reading Comprehension subtests, met the needed assumptions.

Research Question With Accompanying Statistical Analyses

What is the predictive validity of DIBELS Next ORF WC and Accuracy Percentage scores to WIAT-III ORF Rate, ORF Accuracy, and Reading Comprehension scores, and is it the same across grade levels?

The researcher used a MLR to determine how well the DIBELS ORF scores predict the WIAT-III scores. The researcher used regression analysis to determine the difference between grade levels when using the DIBELS Next ORF WC and Accuracy Percentage scores together predict the WIAT-III ORF Accuracy score. See Table 5 for the variables and statistical analyses used to analyze the research question.

Table 5

Research Question, Variables, and Statistical Analyses

Research Question	Variables	Statistical Analyses
What is the predictive validity of DIBELS Next ORF WC and Accuracy Percentage scores to WIAT-III ORF Rate, ORF Accuracy, and Reading Comprehension scores, and is it the same across grade levels?	DIBELS Next ORF WC and Accuracy Percentage scores	MLR
	WIAT-III ORF Rate, ORF Accuracy, and Reading Comprehension scores	
	Developmental Level	Regression Analysis

Note. DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct; WIAT-III = Wechsler Individual Achievement Test, Third Edition; MLR = Multiple linear regression.

Summary

An evaluation of the predictive strength of the DIBELS Next ORF the WIAT-III ORF and Reading Comprehension subtests was completed in this study. The sample consisted of 75 second through fifth grade students who participated in a special

education eligibility evaluation. The students in the sample attended 14 elementary schools in one school district in Delaware. The researcher would like to generalize the results of this study to all second through fifth grade students participating in special education eligibility evaluations. The researcher invited all the elementary school psychologists in the district to participate in the data collection for this study; however, participation was voluntary. The data collected included the students' grade level, sex, DIBELS Next ORF WC score, DIBELS Next ORF Accuracy Percentage score, WIAT-III ORF Rate score, WIAT-III ORF Accuracy score, WIAT-III Reading Comprehension subtest score, and dates of administration. The school psychologists assigned to each school sanitized the data from their school prior to presenting it to the researcher. The researcher calculated descriptive statistics including means, standard deviations, medians, skewness, and kurtosis for scores on the DIBELS Next ORF and WIAT-III ORF and Reading Comprehension subtests. The researcher used a MANOVA to measure the differences between sex on both measures. The researcher then used MLR to determine how well the DIBELS Next ORF score predicts WIAT-III scores and regression analysis to determine differences between grade levels.

CHAPTER 4

RESULTS

The purpose of this study was to examine the predictive validity of the Dynamic Indicator of Basic Early Literacy Skills, Next Edition (DIBELS Next; Good et al., 2011) Oral Reading Fluency (ORF) assessment to the Wechsler Individual Achievement Test, Third Edition (WIAT-III; Wechsler, 2009) ORF and Reading Comprehension subtests. The researcher collected archival data for students from second through fifth grade who completed special education eligibility evaluations. Eight school psychologists in the school district volunteered to gather and sanitize the data for the researcher. The students in the sample attended 14 elementary schools in one school district in Delaware. The data collected included the students' grade level, sex, DIBELS Next ORF Word Correct (WC) scores, DIBELS Next ORF Accuracy Percentage scores, WIAT-III ORF Rate scores, WIAT-III ORF Accuracy scores, WIAT-III Reading Comprehension subtest scores, and dates of administration. Specifically, this research study addressed the following question and hypothesis:

What is the predictive strength of DIBELS Next ORF WC and Accuracy Percentage scores to WIAT-III ORF Rate, ORF Accuracy, and Reading Comprehension scores, and is it the same across grade levels? The researcher hypothesized that the predictive strength of the DIBELS Next ORF scores would be strong across all grade levels and all assessments.

Complications

The researcher encountered one complication during the study. Although the sample included 75 students, data for some participants were incomplete. The obtained

data set included DIBELS Next ORF WC and Accuracy Percentage scores for all participants; however, the data set included WIAT-III ORF Rate and Accuracy scores for 70 students and WIAT-III Reading Comprehension scores for 67 students. Therefore, the number of participants from which data were used varied depending on the particular statistical analysis performed. It should be noted that there was no systematic mortality in the sample that would suggest a threat to internal validity, rather, mortality was random. Overall, the sample size was smaller than what the researcher had expected thus limiting the statistical power needed to fully appraise the presence of potentially statistically significant relationships.

Statistical Analyses

The researcher used the International Business Machine's Statistical Package for the Social Sciences 21 to analyze the data. Prior to conducting any inferential statistics, the researcher conducted preliminary analyses to determine if sex differences existed in relation to the DIBELS Next ORF and WIAT-III scores and to establish that underlying statistical assumptions for the inferential statistics employed were met. The researcher created a correlation matrix along with multiple linear regression (MLR) models to measure the associations between DIBELS Next and WIAT-III scores. The researcher created MLR models for each grade level and compared the strengths of the correlations by using a Fisher r-to-z transformation. Due to the normality assumption not being met for the one variable, nonparametric statistical procedures were also used to analyze that variable.

Descriptive Statistics

The sample included DIBELS Next ORF WC and Accuracy Percentage scores of students (*N* = 75) in second through fifth grades. It also included WIAT-III ORF Rate, ORF Accuracy, and Reading Comprehension scores from students in these same grades. The number of participants from whom data on various WIAT-III measures varied due to different test administration practices among the district school psychologists: WIAT-III ORF Rate (*n* = 70), WIAT-III ORF Accuracy (*n*= 70), and WIAT-III Comprehension *(n* = 67). Table 6 includes the descriptive statistics for each variable.

Table 6

Descriptive Statistics for Dependent Variables

	N	*M*	*SD*	Skewness	Kurtosis
DIBELS ORF WC	75	67.1	28.1	.13	-.41
DIBELS ORF Accuracy Percentage	75	91.4	9.4	-2.9	10.8
WIAT-III ORF Rate	70	83.2	10.1	-.28	.83
WIAT-III ORF Accuracy	70	83.3	12.0	.20	-.07
WIAT-III Reading Comprehension	67	87.2	9.9	.40	-.18

Note. DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct; WIAT-III = Wechsler Individual Achievement Test, Third Edition. DIBELS scores are presented as raw scores. WIAT-III scores are presented as standard scores.

Preliminary Analysis of Statistical Assumptions

Certain assumptions must be met when conducting inferential statistical analyses. Stevens (2002) offered a review of the assumptions of MLR and multiple analysis of variance (MANOVA), including using interval or ratio data, normality, linearity, equal variances, and independence of variables. However, some statisticians

suggest that independence of variables is not a necessary assumption when using MLR (Osborne & Waters, 2002). Two additional assumptions needed for MLR are homoscedasticity and non-multicollinearity (Osborne & Waters, 2002; Stevens, 2002).

First, the researcher determined that all variables including DIBELS Next, WIAT-III, grade level, and sex were ratio or interval data and that the sample size was appropriate. Next, the data were examined for outliers. The researcher completed this analysis by examining the frequency data for each variable. No outliers were noted in the dataset.

The researcher then assessed the assumption of normality by visually examining histograms, interpreting the skewness and kurtosis values for each variable, and using the Shapiro-Wilk Test of Normality as an empirical measure of normality. If the skewness of a set of data is between -0.5 and +0.5, the distribution is approximately symmetric (Bulmer, 1979). A normal distribution has a kurtosis value of exactly three (Bulmer, 1979). A data set is considered within acceptable range if the kurtosis is ±2. In the case of the Shapiro-Wilk Test, a statistically significant finding is indicative of data that are not normally distributed.

The skewness and kurtosis values were within reasonable ranges for the WIAT-III ORF Rate, ORF Accuracy, and Reading Comprehension subtests, indicating these data were normally distributed. Results from the Shapiro-Wilk test indicated these variables were normally distributed. The skewness and kurtosis values for the DIBELS Next ORF WC were within reasonable limits; however, the DIBELS Next ORF Accuracy Percentage data were highly skewed and leptokurtic. The Shapiro-Wilk Test also showed a significant result for non-normality for the DIBELS Next Accuracy Percentage

76

(W = .99, p =.00). This non-normal distribution corresponds with the national DIBELS Next ORF Accuracy Percentage normative data, as the average accuracy percentage was highly negatively skewed and leptokurtic across all grade levels. The average accuracy percentage for second through fifth grade students on the DIBELS Next ORF assessment ranged from 90 – 99% (Good et al., 2011).

Preliminary analyses suggest that all but one variable was normally distributed. While the DIBELS Next ORF Accuracy Percentage score was left-skewed and leptokurtic, the distribution of these data mirrored that of large normative data sets. Consequently, interpretations from parametric analyses using DIBELS Next ORF Accuracy Percentage scores need to be made with some caution given the violation of one underlying assumption. Due to the violation, non-parametric analyses were also used to investigate relationships involving DIBELS Next ORF Accuracy Percentage scores. See Table 6 for skewness and kurtosis values, Table 7 for Shapiro-Wilk results and Appendix B for histograms.

The researcher next assessed the assumption of linearity by examining the bivariate scatterplots and correlation matrix. All scatterplots indicated a linear relationship among the variables. Additionally, the researcher examined the correlation matrix to determine if the variables had linear relationships. All variables had a correlation coefficient of .3 or higher, suggesting linear relationships (Bulmer, 1979). Therefore, after analyzing the bivariate scatterplots and correlation matrix, the researcher determined that the assumption of linearity was upheld. Appendix C includes a bivariate scatterplot matrix for all variables, and Table 8 provides the correlation matrix.

Table 7

Shapiro-Wilk Test for Normality for Each Variable

	Shapiro-Wilk *W*	*p*
DIBELS ORF WC	.99	.87
DIBELS ORF Accuracy Percentage	.73	.00
WIAT-III ORF Rate	.97	.20
WIAT-III ORF Accuracy	.98	.23
WIAT-III Reading Comprehension	.98	.42

Note. DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct; WIAT-III = Wechsler Individual Achievement Test, Third Edition.

Table 8

Pearson Correlations for the DIBELS Next ORF, WIAT-III ORF, and WIAT-III Reading Comprehension Assessment Scores

Assessment Score	DIBELS ORF WC	DIBELS ORF Accuracy Percentage	WIAT-III Rate	WIAT-III Accuracy
DIBELS ORF WC	-			
DIBELS ORF Accuracy Percentage	.65*	-		
WIAT-III ORF Rate	.59*	.55*	-	
WIAT-III ORF Accuracy	.59*	.53*	.64*	-
WIAT-III Reading Comprehension	.44*	.41*	.47*	.47*

Note. DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct; WIAT-III = Wechsler Individual Achievement Test, Third Edition.
*p < .01

Next, the researcher assessed the assumption of equal variances. The Box's

Test of Equality of Covariance Matrices was used to check homogeneity of covariance

across groups. The results indicated that the covariance matrices were equal across

dependent variables, Box's M = 22.83, F (15, 13959.61) = 1.38; p= .15. Consequently,

the assumption of equal variances was met for all the variables.

The researcher assessed the assumption of independence of variables using the

Durbin-Watson statistical test. Despite the belief of some statisticians that this

assumption is unimportant (Statistics Solutions, 2014), the researcher tested for it due

to the lack of consensus among statisticians. No relation should exist on the dependent

measures to other scores for this assumption to be met. Durbin-Waston statistics

between 1.5 and 2.5 suggest that the residuals are not correlated (Statistics Solutions,

2014). The Durbin-Watson statistics, across all dependent measures, fell within the

range of 1.5 and 2.5. See Table 9 for Durbin-Watson statistics.

The researcher checked for homoscedasticity by visually analyzing the residual

plots. When the variance of errors is the same across all levels of the variables, the

assumption of homoscedasticity is met. For all variables, the variance of errors was

approximately the same across all variables. See Appendix D for residual plots.

To assess the assumption of multicolinearity, the researcher again examined the

correlation matrix. The purpose of this examination was to determine if any of the

DIBELS Next and WIAT-III variables have a very strong correlation (.8 or higher) with

each other (Bulmer, 1979). No variables in this study demonstrated strong correlations

with each other. See Table 8 to examine the correlation matrix. The researcher also

79

assessed multicolinearity by analyzing the variance inflation factor (VIF) of the

independent variables, DIBELS Next ORF WC and Accuracy Percentage, to determine

the strength of the relationship between these predictor values. It is suggested that a

VIF value of 10 or higher indicates that multicolinearity may exist (Stevens, 2002). The

VIF for the DIBELS Next ORF WC and Accuracy Percentage variables was below the

suggested threshold of 10 (VIF = 1.00).

Table 9

Durbin-Watson Statistics

Predictors	Dependent Variables	Durbin-Watson
DIBELS Next ORF WC and Accuracy Percentage	WIAT-III ORF Rate	1.91
DIBELS Next ORF WC and Accuracy Percentage	WIAT-III ORF Accuracy	1.94
DIBELS Next ORF WC and Accuracy Percentage	WIAT-III Reading Comprehension	1.74

Note. DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct; WIAT-III = Wechsler Individual Achievement Test, Third Edition.

Research Question Results

The researcher initially used MANOVA to determine if sex differences existed

among any of the variables used in this study. The results indicated that no significant

sex differences existed between males and females for the DIBELS Next ORF WC and

Accuracy Percentage scores or the WIAT-III ORF Rate, ORF Accuracy, or Reading

Comprehension scores, Wilks' λ = .91, F (5, 56) = .96, p = .45. See Table 10 for the

results of the MANOVA.

The researcher then used follow up analysis of variance (ANOVA) tests to analyze the sex differences among the dependent measures despite the lack of statistically significant differences noted in the MANOVA. See Table 11 for the results of the ANOVAs. Since no significant differences existed, the researcher aggregated the data across sexes to maximize statistical power in the MLR analysis.

Table 10

MANOVA for Sex Differences Among All Variables

Test Name	Value	*F*	Hypothesis *df*	*Error df*	*p*
Pillai's Trace	.08	.96	5.00	56.00	.45
Wilks' Lambda	.91	.96	5.00	56.00	.45
Hotelling's Trace	.09	.96	5.00	56.00	.45
Roy's Largest Root	.09	.96	5.00	56.00	.45

Table 11

ANOVAs for Sex Differences Across All Dependent Variables

Assessment Score	*df*	*F*	*p*
DIBELS ORF WC	1, 60	.40	.53
DIBELS ORF Accuracy Percentage	1, 60	.00	.96
WIAT-III ORF Rate	1, 60	.79	.40
WIAT-III ORF Accuracy	*1*, 60	.37	.55
WIAT-III Reading Comprehension	1, 60	.69	.41

Note. DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct; WIAT-III = Wechsler Individual Achievement Test, Third Edition.

Due to the assumption violation of normality for the DIBELS Next Accuracy

Percentage scores, the researcher conducted a nonparametric statistical analysis.

The results of the Mann-Whitney U test indicate that there is no significant difference (*U*

= 667, *p* = .78) between sexes for the DIBELS Next ORF Accuracy Percentage scores.

The nonparametric analyses of the Mann-Whitney *U*, confirmed the parametric

procedures.

Grade levels (2 -5) aggregated. Several entry methods exist for MLR including

direct, forward selection, backward elimination, hierarchical selection, and stepwise

selection (Stevens, 2002). When using the direct method, all independent variables are

entered into the equation simultaneously. This method is used when the researcher

does not know which predictor variables will create the best fit to the data. The forward

selection requires the researcher to enter the predictor variables with greatest

theoretical importance into the equation first. When using the backward elimination

method the researcher includes all predictor variables in the equation and removes the

variables that do not contribute to the regression equation. Hierarchical selection is a

version of forward selection where the predictors are grouped into sets or blocks based

on psychometric properties or theory. Once the sets are established, stepwise selection

occurs. In the stepwise MLR model, the predictor variable with the highest correlation

with the dependent variable is entered first. Subsequent predictor variables continue to

be entered to find the combination of variables that account for the highest percentage

of variance. As each variable is included, the correlation strength, or R^2 value, should

increase. When the increase in the R^2 value has a *p*> .05, the stepwise procedure

stops and no additional variables are added to the equation because the addition of more variables would not significantly contribute to the prediction model.

The researcher used stepwise MLR to analyze the relationship of the DIBELS Next ORF WC and Accuracy Percentage for the WIAT-III ORF Rate, ORF Accuracy, and Reading Comprehension subtests. This stepwise selection model was utilized because the entry of prediction variables is based on empirical data rather than theoretical assumptions. In the stepwise MLR model, the predictor variable with the highest correlation with the dependent variable is entered first rather than allowing the researcher to choose the order of the variables entered into the model. By using a statistical approach for entering variables into the model, human bias and error is minimized.

The model that best predicted the WIAT-III ORF Rate score included the DIBELS Next ORF WC and Accuracy Percentage scores, $F(2, 67) = 21.52$, $p< .01$. In this model, the DIBELS Next ORF WC score accounted for 34% of the variance in the WIAT-III ORF Rate score while the addition of the DIBELS Next Accuracy Percentage score accounted for an additional 5% of the variance. Overall, the DIBELS Next ORF WC and Accuracy Percentage score accounted for 39% of the variance of the WIAT-III ORF Rate score. Table 12 includes the results of the stepwise MLR for the WIAT-III ORF Rate variable.

The model that best predicted the WIAT-III ORF Accuracy score included only the DIBELS Next ORF WC score, $F (1, 68) = 37.14$, $p < .01$. The DIBELS Next ORF WC score accounted for 35% of the variance in the WIAT-III ORF Accuracy score. The

addition of the DIBELS Next ORF Accuracy Percentage scores did not result in a

statistically significant improvement when predicting the variance of the WIAT-III ORF

Accuracy Percentage score. Table 13 includes the results of the stepwise MLR for the

WIAT-III ORF Accuracy Percentage variable.

Table 12

Stepwise MLR for the WIAT-III ORF Rate Variable

Model	Variables	β	*SE*	R^2	*F*
1	DIBELS Next ORF WC	.21	.04	.34	35.69**
2	DIBELS Next ORF WC	.14	.05		
	DIBELS Next ORF Accuracy Percentage	.31	.14	.39	21.52**

Note. MLR = Multiple linear regression; DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct.
**p <.01

Table 13

Stepwise MLR for the WIAT-IIIORF Accuracy Variable

Model	Variables	β	*SE*	R^2	*F*
1	DIBELS Next ORF WC	.25	.04	.35	37.14**

Note. MLR = Multiple linear regression; DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct.
**p <.01

The model that best predicted the WIAT-III Reading Comprehension score

included only the DIBELS Next ORF WC score, $F (1, 65) = 15.49, p< .01$. The DIBELS

Next ORF WC score accounted for19% of the variance in the WIAT-III Reading

Comprehension score. The addition of the DIBELS Next ORF Accuracy Percentage

score did not significantly add to the prediction model that included just DIBELS Next

ORF WC. Table 14 includes the results of the stepwise MLR for the WIAT-III Reading

Comprehension variable.

Individual grade levels. The researcher further analyzed the data by grade

level to determine if the prediction models were different across grade levels (groups).

The researcher used stepwise MLR for each grade level to analyze the relationship of

the DIBELS Next ORF WC and Accuracy Percentage for the WIAT-III ORF Rate, ORF

Accuracy, and Reading Comprehension subtests. Upon completion of the MLR, the

researcher compared the strength of the regression at each grade level by using a

Fisher *r*-to-*z* transformation.

Table 14

Stepwise MLR for the WIAT-III Reading Comprehension Variable

Model	Variables	β	*SE*	R^2	*F*
1	DIBELS Next ORF WC	.17	.04	.19	15.49**

Note. MLR = Multiple linear regression; DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct.
**$p < .01$

WIAT-III ORF Rate. The model that best predicted the WIAT-III ORF Rate score

for the second grade level included the DIBELS Next ORF WC score, $F(1, 17) = 46.52$,

$p < .01$.   In this model, the DIBELS Next ORF WC score accounted for 73% of the

variance in the WIAT-III ORF Rate score. The addition of the DIBELS Next ORF

Accuracy Percentage score did not significantly add to the prediction model that

included only DIBELS Next ORF WC score.

The model that best predicted the WIAT-III ORF Rate score for the third grade

level included the DIBELS Next ORF WC score, $F(1, 16) = 5.18$, $p < .05$.   In this model,

the DIBELS Next ORF WC score accounted for 25% of the variance in the WIAT-III ORF Rate score. The addition of the DIBELS Next ORF Accuracy Percentage score did not significantly add to the prediction model that included only DIBELS Next ORF WC score.

The model that best predicted the WIAT-III ORF Rate score for the fourth grade level included the DIBELS Next ORF WC score, $F(1, 13) = 5.13$, $p< .05$. In this model, the DIBELS Next ORF WC score accounted for 28% of the variance in the WIAT-III ORF Rate score. The addition of the DIBELS Next ORF Accuracy Percentage score did not significantly add to the prediction model that included only DIBELS Next ORF WC score.

No model, including DIBELS Next ORF WR and/or DIBELS Next ORF Accuracy Percentage, predicted the WIAT-III ORF Rate score at the fifth grade level. When using stepwise MLR, neither predictive variables were entered into the model because the F value for each variable was not significant at the $p< .05$ level. These results suggested that the DIBELS Next ORF WC and Accuracy Percentage scores were not significant predictors of WAIT-III ORF Rate scores at the fifth grade level. Table 15 includes the results of the stepwise MLR for the WIAT-III ORF Rate variable by grade level.

The researcher then investigated the difference between the strength of the correlation coefficients between grade levels for the WIAT-III ORF Rate score, the DIBELS Next WC score, and the variables determined by MLR analysis to be significant predictors. The significant predictor variable for the WIAT-Rate score was the DIBELS Next ORF WC score.   The statistical analysis used to evaluate the magnitude of grade-

level differences was the Fisher *r*-to-*z* transformation. The fifth grade level was

excluded from the analysis because that prediction model was not significant. The

results indicated the prediction model for second grade (*r* = .86) was significantly

stronger than the third grade model (*r* = .50), *Z* = 2.05, *p* = .04. There was no significant

difference between the strength of the second grade model (*R* = .86) and the fourth

grade model (*r* = .53), *Z* = 1.79, *p* = .07. Additionally, there was no significant difference

between the third grade model (*r* = .50) and the fourth grade model (*r* = .53), *Z* = .13, *p*

= .90. Table 16 includes the results of the Fisher *r*-to-*z* transformation for the WIAT-III

ORF Rate score.

Table 15

Stepwise MLR for the WIAT-III ORF Rate Variable

Grade	Model	Variable	β	*SE*	R^2	*F*
2nd	1	DIBELS Next ORF WC	.39	.06	.73	46.52**
3rd	1	DIBELS Next ORF WC	.19	.08	.25	5.18**
4th	1	DIBELS Next ORF WC	.20	.09	.28	5.13*

Note. MLR = Multiple linear regression; DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct.
*p < .05
**p <.01

WIAT-III ORF Accuracy. The model that best predicted the WIAT-III ORF

Accuracy scores for the second grade included the DIBELS Next ORF Accuracy

Percentage scores, *F* (1, 17) = 14.26, *p*< .01. In this model, the DIBELS Next ORF

Accuracy Percentage scores accounted for 46% of the variance in the WIAT-III ORF

Accuracy scores. The addition of the DIBELS Next ORF WC scores did not significantly

add to the prediction model that included only DIBELS Next ORF Accuracy Percentage

scores.

Table 16

Fisher r-to-z Transformation for the Significant Predictor Variables and the WIAT-III ORF Rate Score

Grade	n	r	Z	p
2nd	19	.86	2.05	0.04*
3rd	18	.50		
2nd	19	.86	1.79	0.07
4th	15	.53		
3rd	18	.50	0.13	0.90
4th	15	.53		

Note. Significant Predictor Variable = DIBELS Next ORF WC sores.
*p < .05

The model that best predicted the WIAT-III ORF Accuracy scores for the third

grade level included the DIBELS Next ORF WC scores, $F(1, 16) = 32.70$, $p< .01$. In this

model, the DIBELS Next ORF WC scores accounted for 67% of the variance in the

WIAT-III ORF Accuracy scores. The addition of the DIBELS Next ORF Accuracy

Percentage scores did not significantly add to the prediction model that included only

DIBELS Next ORF WC scores.

The model that best predicted the WIAT-III ORF Accuracy score for the fourth

grade included the DIBELS Next ORF WC scores, $F(1, 13) = 6.12$, $p< .05$. In this

model, the DIBELS Next ORF WC scores accounted for 32% of the variance in the WIAT-III ORF Accuracy scores. The addition of the DIBELS Next ORF Accuracy Percentage scores did not significantly add to the prediction model that included only DIBELS Next ORF WC scores.

No model, including DIBELS Next ORF WR and/or DIBELS Next ORF Accuracy Percentage, predicted the WIAT-III ORF Accuracy scores at the fifth grade level. When using stepwise MLR, neither predictive variables were entered into the model because the F for each value was not significant at the< .05 level. The results suggested that the DIBELS Next ORF WC and Accuracy Percentage were not significant predictors of WAIT-III ORF Accuracy scores at the fifth grade level. Table 17 includes the results of the stepwise MLR for the WIAT-III ORF Accuracy variable by grade level.

The researcher then investigated the difference between the strength of the correlation coefficients across grade levels for the WIAT-III ORF Accuracy score and the variables determining by MLR analysis to be significant predictors. The significant predictor variable for the WIAT-III ORF Accuracy score was the DIBELS Next ORF Accuracy Percentage score at the 2nd grade level and the DIBELS Next ORF WC at the 3rd and 4th grade level.   The statistical analysis used to evaluate the magnitude of grade-differences was the Fisher r-to-z transformation. The fifth grade level was excluded from the analysis because the prediction model was not significant for that grade level. The results indicated that there was no significant difference between the strength of the correlations at each grade level. Table 18 includes the results of the Fisher r-to-z transformation for the WIAT-III ORF Accuracy score.

Table 17

Stepwise MLR for the WIAT-III ORF Accuracy Variable

Grade	Model	Variable	B	SE	R^2	F
2nd	1	DIBELS Next ORF Accuracy Percentage	.69	.182	.46	14.26**
3rd	1	DIBELS Next ORF WC	.44	.08	.67	32.70**
4th	1	DIBELS Next ORF WC	.34	.14	.32	6.12*

Note. MLR = Multiple linear regression; DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct.
*p < .05
**p < .01

WIAT–III Reading Comprehension. The model that best predicted the WIAT-III

Reading Comprehension scores for the second grade included the DIBELS Next ORF

WC scores, $F(1, 17) = 13.46$, p< .01.   In this model, the DIBELS Next ORF WC scores

accounted for 49% of the variance in the WIAT-III ORF Rate scores. The addition of the

DIBELS Next ORF Accuracy Percentage scores did not significantly add to the

prediction model that included only DIBELS Next ORF WC scores.

No model, including DIBELS Next ORF WR and/or DIBELS Next ORF Accuracy

Percentage, predicted the WIAT-III Reading Comprehension scores at the third grade

level. When using stepwise MLR, neither predictor variables were entered into the

model because the F value for each variable was not significant at the < .05 level. The

results suggested that the DIBELS Next ORF WC and Accuracy Percentage were not

significant predictors of WAIT-III Reading Comprehension scores at the third grade

level.

Table 18

Fisher r-to-z Transformation for the Significant Predictor Variable and the WIAT-III ORF Accuracy Score

Grade	N	r	Z	P
2nd	19	0.68	0.93	0.35
3rd	18	0.82		
2nd	19	0.68	0.47	0.64
4th	15	0.57		
3rd	18	0.82	1.32	0.19
4th	15	0.57		

Note. Significant Predictor Variable = DIBELS Next ORF WC (3rd and 4th grade) and DIBELS Next ORF Accuracy Percentage (2nd grade).

The model that best predicted the WIAT-III Reading Comprehension scores for the fourth grade included the DIBELS Next ORF Accuracy Percentage scores, $F(1, 15)$ = 5.53, $p< .05$. In this model, the DIBELS Next Accuracy Percentage scores accounted for 27% of the variance in the WIAT-III Reading Comprehension scores. The addition of the DIBELS Next ORF WC scores did not significantly add to the prediction model that included only DIBELS Next ORF Accuracy Percentage scores.

No model, including DIBELS Next ORF WR and/or DIBELS Next ORF Accuracy Percentage, predicted the WIAT-III Reading Comprehension scores at the fifth grade level. When using stepwise MLR, neither predictor variables were entered into the model because the F value for each variable was not significant at the< .05 level. The

results suggested that the DIBELS Next ORF WC and Accuracy Percentage were not

significant predictors for WAIT-III Reading Comprehension scores at the fifth grade

level. Table 19 includes the results of the stepwise MLR for the WIAT-III Reading

Comprehension variable by grade level.

Table 19

Stepwise MLR for the WIAT-III Reading Comprehension Variable

Grade	Model	Variables	β	SE	R^2	F
2nd	1	DIBELS Next ORF WC	.27	.073	.49	13.46**
4th	1	DIBELS Next ORF Accuracy Percentage	2.0	.85	.27	5.53*

Note. MLR = Multiple linear regression; DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WC = Words correct.
*p < .05
**p < .01

The researcher then investigated the difference between the strength of the

correlation coefficients across grade levels for the WIAT-III Reading Comprehension

score and the variables determined by MLR analyses to be significant predictors.   The

significant predictor variable for the WIAT-III Reading Comprehension score was the

DIBELS Next ORF WC score at the 2nd grade level and the DIBELS Next ORF

Accuracy Percentage at the 4th grade level. The statistical analysis used to determine

the magnitude of grade-level differences was the Fisher *r*-to-*z* transformation. Only the

second and fourth grade correlations were included in the analysis because the

prediction was not significant at the third and fifth grade level. The results indicated that

there was no significant difference between the strength of the correlations at each

grade level. Table 20 includes the results of the Fisher *r*-to-*z* transformation for the

WIAT-III Reading Comprehension variable.

Table 20

Fisher r-to-z Transformation for the Significant Predictor Variable and the WIAT-III
Reading Comprehension Score

Grade	*N*	*r*	*Z*	*P*
2nd	.16	0.70	0.76	0.45
4th	17	0.52		

Note. Significant Predictor Variable = DIBELS Next ORF WC (2nd grade) and DIBELS Next ORF
Accuracy Percentage (4th grade).

       Due to the violation of the normality assumption for the WIAT-III ORF Accuracy

scores, the researcher conducted a nonparametric statistical procedure to investigate

the relationships between the WIAT-III ORF Accuracy scores and the predictor

variables (DIBELS Next ORF WC and Accuracy Percentage). Specifically, the

researcher used a Spearman's Rho to analyze the correlations. The results of the

nonparametric procedure confirmed the results of the parametric procedure. Both

procedures, the Spearman's Rho and the Pearson Product Correlation, indicated strong

correlations between the WIAT-III Accuracy scores and the predictor scores (DIBELS

Next ORF WC and Accuracy Percentage.

Summary

       The purpose of this study was to evaluate the predictive strength of the DIBELS

Next ORF to the WIAT-III ORF and Reading Comprehension subtests. Early

intervening teams commonly use the DIBELS Next ORF assessment data to make

important decisions including a referral for a special education eligibility evaluation

and/or a determination of an SLD. WIAT-III data are commonly used in the process of

determining whether a student has a disability warranting special education services.

Given that DIBELS Next data are often used to determine whether a student should be

referred for a comprehensive psychoeducational evaluation using the WIAT-III, there is

a need to empirically establish the predictive relationship of DIBELS Next to WIAT-III. If

the results of the current study supported that DIBELS Next ORF was a strong predictor

of the WIAT-III ORF and Reading Comprehension subtests, school-based teams would

have stronger empirical justification for using the DIBELS Next ORF to make important

decisions. If this study does not support the DIBELS Next ORF assessment as a strong

predictor of WIAT-III ORF and Reading Comprehension scores, early intervening teams

may de-emphasize DIBELS Next ORF scores when deciding whether students should

undergo a special education eligibility evaluation. Additionally, professionals using

solely CBM data to determine special education eligibility may need to consider

supplementing the CBM data with other assessment data. If the DIBELS Next ORF

assessment shows that students are performing adequately and early intervening teams

decide to not refer a student for a special education evaluation, school districts may be

excluding students from special education eligibility evaluations who truly need to be

evaluated.

The researcher conducted preliminary data reviews and tests for underlying

assumptions relevant to the inferential statistics employed. These results indicated that

the data met the appropriate assumptions for most data sets. The DIBELS ORF Next

Accuracy Percentage variable was highly skewed and had a positive kurtosis; however,

this profile corresponds with the national DIBELS Next ORF Accuracy Percentage

normative data set (Good et al., 2011).Additionally, the researcher used nonparametric

procedures to confirm the parametric procedures to analyze the DIBELS Next ORF

Accuracy Percentage variable. Both parametric and nonparametric analyses were used

to determine if the data should be disaggregated by sex and the results revealed no sex

differences on the DIBELS Next ORF WC or Accuracy Percentage or the WIAT-III ORF

or Reading Comprehension scores. Therefore, the researcher analyzed the data

aggregated across sex. To measure the associations between the DIBELS Next and

WIAT-III scores, the researcher created a correlation matrix and computed a series of

stepwise MLRs to evaluate the strength of various DIBELS Next metrics predicting

various WIAT-III reading metrics.

The stepwise MLR results suggested that the prediction model including the

DIBELS Next ORF WC and Accuracy Percentage scores when predicting the WIAT-III

ORF Rate scores fit the data best. However, the prediction model including only the

DIBELS Next WC scores when predicting both the WIAT-III ORF Accuracy Percentage

and Reading Comprehension scores best fit the obtained data. In order to determine

differences at each grade level, the researcher compared separate MLR models at each

grade level. The MLR analysis indicated significant correlations for the WIAT-III ORF

Rate variable at the second, third, and fourth grade level with the strongest correlations

being at the second and fourth grade level. For the WIAT-IIIORF Accuracy variable,

significant correlations again existed at the second, third, and fourth grade level with no

significant correlation strength differences between grade levels. A significant

correlation for the WIAT-III Reading Comprehension variable existed at the second and

fourth grade levels.   Second and fourth grades were the only grade levels where

significant correlations existed for all WIAT-III variables. No significant correlations

existed between the DIBELS Next ORF and WIAT-III ORF and Reading

Comprehension scores at the fifth grade level. See table 21 for an overview of the

strongest predictive correlation at each grade level.

Table 21

Strongest Predictive Correlation at Each Grade Level

Grade	WIAT-III ORF Rate	WIAT-III ORF Accuracy	WIAT-III Reading Comprehension
2nd	DIBLES Next ORF WC	DIBELS Next ORF Accuracy Percentage	DIBELS Next ORF WC
3rd	DIBLES Next ORF WC	DIBELS Next ORF WC	None
4th	DIBLES Next ORF WC	DIBELS Next ORF WC	DIBELS Next ORF Accuracy Percentage
5th	None	None	None

Note. DIBELS Next = Dynamic Indicators of Basic Early Literacy Skills, Next Edition; ORF = Oral reading fluency; WIAT-III = Wechsler Individual Achievement Test, Third Edition; WC = Words correct.

CHAPTER 5

DISCUSSION

The purpose of this study was to determine the predictive strength of Dynamic Indicators of Basic Early Literacy Skills, Next Edition (DIBELS Next) Oral Reading Fluency (ORF) Words Correct (WC) and Accuracy Percentage scores to Wechsler Individual Achievement Test, Third Edition (WIAT-III) ORF Rate, ORF Accuracy, and Reading Comprehension scores. Additionally, the researcher investigated whether the predictive strength of DIBELS Next ORF scores to corresponding WIAT-III scores were the same across grade levels. The researcher hypothesized that the predictive strength of the DIBELS Next ORF scores for the WIAT-III ORF and Reading Comprehension would be strong across all grade levels and all metrics. Establishment of the DIBELS Next construct validity is vital due to the credence commonly placed on the DIBELS Next data when early intervening teams are making recommendations for students to undergo special education eligibility evaluations and/or using the DIBELS Next data to make specific learning disability (SLD) eligibility decisions within a Response to Intervention (RTI) model.

The sample for the study included second through fifth grade students referred for special education eligibility evaluations during the 2011-2012 school year. Both males and females were included in this study. The study site consisted of one school district in Delaware with 14 elementary schools. The student body for the school district during the 2011-2012 school year was 49.2% female, 52.4% of students in the district received free/reduced-priced lunch, and 11% received special education services. At the study site, all second through fifth grade students receiving Tier II and Tier III

reading support in RTI complete the DIBELS Next every two weeks to monitor their

progress. The district school psychologists gathered the DIBELS Next ORF data

collected just prior to the administration of the WIAT-III ORF and Reading

Comprehension subtests administered during the special education eligibility evaluation,

which was archival data. The district school psychologists removed students' names

from the data and each student received a random identification code. The school

psychologists gave the archived, anonymous data to the researcher.

Prior to conducting the primary set of statistical analyses for answering the

research questions, the researcher used a multivariate analysis of variance (MANOVA),

analysis of variance (ANOVA), and a Mann-Whitney U to analyze the data for potential

sex differences on DIBELS Next ORF WC and Accuracy Percentage scores and WIAT-

III ORF Rate, ORF Accuracy, and Reading Comprehension scores. Next, the

researcher computed Pearson Product Correlations and Spearman's Rho, when

applicable, to measure the associations between DIBELS Next and WIAT-III scores.

Finally, multiple linear regression (MLR) models were created for each grade level and

comparisons of the strengths of the correlations across grade levels were accomplished

using a Fisher *r*-to-*z* transformation.

Summary of Findings

Sex Differences

Research results have been inconclusive in regards to sex differences in reading

skills. Numerous studies documented that females have higher reading achievement

(Ayers, 1909 as cited in Below et al., 2010; Davenport et al., 2002; Gates, 1961;

Klecker, 2006). Further, some research suggests that boys are more likely to receive a

SLD classification than their female counterparts (Coutinho & Oswald, 2005;Wehmeyer & Schwartz, 2001). Rates of reading SLD ranged from 18.0% to21.6% in males and 7.9% to13.0% in females across four studies (Rutter et al., 2004).   Results from SLD prevalence rate studies indicate that boys are more likely to have substantial reading deficiencies necessitating special education services. Contrary to the results of those studies, other research suggests males and females are comparable in their reading skills (Below et al., 2010).Given these equivocal results, data from the current study were reviewed for potential sex differences in any of the dependent variables.

A MANOVA, ANOVA, and Mann-Whitney U analyses were performed in the current study to determine if the data should be disaggregated by sex. Results revealed no sex differences on the DIBELS Next ORF WC or Accuracy Percentage or the WIAT-III ORF or Reading Comprehension scores. Therefore, the researcher analyzed the data aggregated across sex.

The current research adds to the inconclusive body of research regarding potential sex differences in reading skills. Some previous research supported a significant difference in reading skills based on sex (Ayers, 1909 as cited in Below et al., 2010; Coutinho & Oswald, 2005; Gates, 1961; Klecker, 2006; Rutter et al., 2004; Schwartz, 2001; Wehmeyer; Davenport et al., 2002;); however, other research did not support this sex difference (Below et al., 2010). While the current study supports Below's (2010) findings, the matter of sex differences in reading skills is not yet resolved. Clearly, additional research in this area is recommended.

Predictive Validity of the DIBELS Next ORF to the WIAT-III ORF and Reading Comprehension

  Grade levels (2-5) aggregated. As predicted, the two ORF assessments (WIAT-III ORF and DIBELS Next ORF), which purportedly measure the same construct, were significantly correlated in this sample. The DIBELS ORF (WC and Accuracy Percentage) scores had moderate to strong correlations with the WIAT-III ORF (Rate and Accuracy) scores. More specifically, the best fit model to predict the WIAT-III Rate scores included both the DIBELS Next ORF WC and Accuracy Percentage. If students are strong readers and they are reading at a fast rate, they are usually reading accurately. Students who are struggling readers tend to read slowly and inaccurately. The best-fit model to predict the WIAT-III Accuracy Percentage scores included only the DIBELS Next WC. It is important to note that the DIBELS Next WC is the stronger predictor for both the WIAT-III ORF Accuracy and Rate scores. Additionally, the DIBELS Next WC is the single-best predictor of the WIAT-III ORF Accuracy scores.

  Previous research indicated that ORF is a strong predictor of reading comprehension (Fuchs et al., 2001; Fuchs et al., 1988; Valencia et al., 2010). A review of 14 studies indicated the correlation coefficients between the number of words read correctly in one minute and reading comprehension ranged from .63 to .90 (Marston, 1989).Numerous researchers compared the predictive and concurrent criterion-related validity of ORF skills to several standardized achievement tests and state assessments (Deno, 2003; Fuchs, Fuchs, & Maxwell, 1988; Shapiro, Keller, Lutz, Santora, & Hintze, 2006). The data collected in these studies supported ORF as a strong predictor of reading comprehension.

One study supports the use of CBM ORF to predict performance on state assessments and standardized group-administered assessments. Shapiro et al.(2006) compared ORF CBM scores to the Pennsylvania System of School Assessment (PSSA), the Stanford Achievement Test, Ninth Edition (SAT-9; Harcourt Brace Educational Measurement, 1996), and the Metropolitan Achievement Test, Eighth Edition (MAT-8; Harcourt Brace Educational Measurement, 2000). The researchers compared the CBM measures completed in the fall and winter to the PSSA scores earned in the spring. The predictive validity correlations between the ORF CBM scores and the PSSA ranged from .25 to .69 and the concurrent validity coefficients ranged from .62 to .67. In addition, the researchers investigated the predictive and concurrent validity coefficients for the ORF CBM assessments using the SAT-9 and MAT-8 assessments as additional criterion measures. The predictive correlation coefficients between the ORF CBM and the Comprehension subtest on the SAT-9 ranged from .59 to .72, whereas the concurrent correlation coefficient ranged from .58 to .70. The predictive correlation coefficients between the ORF CBM and the Comprehension subtest on the MAT-8 was .67 during both fall and winter administrations. The concurrent validity coefficients between the ORF CBM and the MAT-8 Comprehension test was .65. These results confirm moderate to strong validity evidence for the previous version of the DIBELS. Further the results of the present study, using DIBELS Next data, are consistent with previous results in which an earlier edition of DIBELS was utilized.

The results of another study supported the conclusion that ORF has strong concurrent criterion-related validity for reading comprehension skills (Fuchs et al.,

1988). One study used the Stanford Achievement Test (SAT; Gardner, Rudman,

Karlsen, & Merwin, 1982) as the criterion measure to compare three other direct

measures of reading comprehension and a measure of ORF. The criterion validity

coefficients for the three direct measures of reading comprehension (passage recall,

cloze method, and question answering) ranged from .70 to .82. The criterion validity

coefficient for the ORF measure was .91 (Fuchs et al., 1988). Although the direct

measures of reading comprehension correlated strongly with the criterion measure, the

strongest correlation existed between the ORF scores and the Reading Comprehension

subtest on the SAT. The correlation coefficients in the current student for the DIBELS

Next ORF (WC and Accuracy Percentages) scores to the WIAT-III Reading

Comprehension scores supported a slightly weaker correlation than did the correlation

coefficients in the Fuchs et al. (1988) study. This difference in correlation strength is

appropriate since the current study investigated predictive validity and the earlier study

investigated concurrent validity.

More specifically, previous research studies have investigated the predictive and

concurrent criterion-related validity for the DIBELS ORF assessment (Munger, 2010;

Powell-Smith et al., 2011). A longitudinal study completed by Munger (2010) compared

the predictive criterion-related validity of the Dynamic Indicators of Basic Early Literacy

Skills, Sixth Edition (DIBELS 6th; Good & Kaminski, 2002) to third grade reading

comprehension measures. The reading comprehension measures included the

Wechsler Individual Achievement Test, Second Edition (WIAT-II; Wechsler, 2001)

Reading Comprehension subtest, the Group Reading Assessment and Diagnostic

Evaluation (GRADE; Williams, 2001) Reading Comprehension subtest, and the New

York State English Language Arts Test (NYSELA). The correlations between the first grade DIBELS 6th ORF WC and all of the reading comprehension measures ranged from .56 to .72. The correlation coefficient for the predictive relationship between the DIBELS 6thORF WC and the WIAT-II Reading Comprehension subtest was .66. Munger (2010) also investigated the concurrent criterion-related validity of the DIBELS Next ORF in this study. The correlations between the third grade DIBELS NEXT ORF WC and the reading comprehension measures ranged from .70 to .85. A .77 correlation coefficient was found for the relationship between the third grade DIBELS Next ORF WC and the WIAT-II Reading Comprehension subtest.

Although correlations were significant in both the Munger (2010) and current studies, the former study illustrated slightly stronger correlations between the DIBELS ORF and WIAT Reading Comprehension than the correlations in the current study. However, when comparing the two studies, several differences between Munger's (2010) study and the current study should be noted. The two research studies cannot be directly compared as both assessments used in the current study are newer versions of the assessments used in the former study. Maturation effects, an internal threat to validity in the Munger study, were not as threatening to the validity of results in the current study. Additionally, the strength of the correlation may differ due to the Flynn Effect (Flynn, 1994) which suggests that higher average ORF and reading comprehension scores over time may negatively influence the validity of comparing current results with previous literature.

A more recent study investigated the criterion-related validity of the current edition of the DIBELS, the DIBELS Next ORF assessment (Powell-Smith et al., 2011).

Researchers used the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001) and the Standard 4th Grade Reading Passage in the National Assessment of Education Progress (NAEP) 2002 Special Study of Oral Reading (Daane, Campbell, Grigg, Goodman, & Oranje, 2005) as the criterion measures. Students completed the GRADE at the end of the school year, and therefore, the beginning of the year and middle of the year DIBELS Next ORF benchmark scores served as the predictor measures. The predictive criterion-related validity of the fall DIBELS Next ORF WC scores to the GRADE Total Test scores ranged from .64to .77. The middle-of-year DIBELS Next ORF Benchmark Assessment evidenced predictive correlations of.59 to .77 to the GRADE Total Test scores. The predictive validity of the fall DIBELS Next ORF Accuracy Percentage scores to the GRADE Total Test scores ranged from .53 to .75with winter DIBELS Next ORF Accuracy Percentage predictive correlations to GRADE Total Test scores ranging from .47 to .80. The concurrent validity of the DIBELS Next ORF WC scores with the GRADE Total Test scores ranged from .61 to .75. Similarly, the concurrent validity of the DIBELS Next ORF Accuracy scores and the GRADE Total Test scores ranged from .49 to .73. Overall, the predictive and concurrent criterion-related validity for the DIBELS Next ORF WC ranged from .59 to .77 and for the DIBELS Next ORF Accuracy Percentage ranged from .49 to .80 when compared to the GRADE Total Test score. The correlation between the DIBELS Next ORF scores and the WIAT-III Reading Comprehension scores in the current study demonstrated a similar strength to the study conducted by Powell-Smith et al. (2011). The correlation between DIBELS Next ORF scores and the WIAT-III Reading

Comprehension scores was slightly weaker in the current study than the correlation between the DIBELS Next ORF and the GRADE Total Test score in the 2011 study.

In the current study, the researcher investigated the predictive validity of the DIBELS Next ORF WC and Accuracy Percentage for the WIAT-III Reading Comprehension score. Despite the fact that the DIBELS ORF and WIAT-III Reading Comprehension scores measure different constructs, it was predicted that the correlations between the two measures would be significant since previous research supported that reading fluency and reading comprehension skills are highly correlated (Deno, 2003; Fuchs et al., 2001, 1988; Fuchs, Fuchs, & Maxwell, 1988;Goffreda & Pedersen, 2009; Marston, 1989; Shapiro et al., 2006; Valencia et al., 2010). As predicted, current results evidenced moderate correlations between the DIBELS Next ORF Rate scores with the WIAT-III Reading Comprehension scores. The best-fit model to predict the WIAT-III Reading Comprehension scores included only the DIBELS Next WC as the sole predictor. Although the percent of variance for the predicted WIAT-III Reading Comprehension was lower than for the WIAT-III ORF Accuracy and Rate, the DIBELS Next ORF WC variable was a statistically significant predictor.

Separate grade levels. The researcher disaggregated the data by grade level, which allowed for an analysis of the predictive validity of DIBELS Next to WIAT-III at each grade level. When predicting the WIAT-III ORF Rate scores, the DIBELS Next ORF WC scores was significant for second, third, and fourth grades. The DIBELS Next ORF WC or Accuracy Percentage scores were not a significant predictor for the WIAT-III ORF Rate scores at the fifth grade level. Overall, the prediction models for second and fourth grade were stronger than the prediction model for third grade.

When predicting the WIAT-III ORF Accuracy scores, the DIBELS Next ORF Accuracy Percentage scores were a significant predictor for second grade. For the third and fourth grade levels, the DIBELS Next ORF WC was a significant predictor of the WIAT-III ORF Accuracy scores. The DIBELS Next ORF WC or Accuracy Percentage scores were not a significant predictor for the WIAT-III ORF Accuracy scores at the fifth grade level. There was no significant difference between the strength of the correlations at the second, third, and fourth grade levels.

When predicting the WIAT-III Reading Comprehension scores, the DIBELS Next ORF WC scores were significant for second grade. At the fourth grade level, the DIBELS Next ORF Accuracy Percentage scores were strong predictors of the WIAT-III Reading Comprehension scores. Neither the DIBELS Next ORF WC nor Accuracy Percentage scores were predictive of the WIAT-III Reading Comprehension scores at the third or fifth grade level. There was no significant difference between the strength of the correlations at the second and fourth grade levels.

Results varied across grade levels in the current study. At the second grade level, the DIBELS Next ORF WC scores were significant when predicting the WIAT-III ORF Rate and Reading Comprehension scores; whereas, the DIBELS Next ORF Accuracy Percentages were significant when predicting the WIAT-III Accuracy Percentage scores. The DIBELS Next ORF WC scores were significant when predicting the WIAT-III ORF Rate and ORF Accuracy scores at the third and fourth grade level. However, the DIBELS Next Accuracy Percentage scores were the only significant predictor for the WIAT-III Reading Comprehension scores at the fourth grade

level. No DIBELS ORF scores were predictive of the WIAT-III ORF or Reading Comprehension scores at the fifth grade level.

One possible factor that contributed to the lack of prediction at the fifth grade level is that most students who undergo initial special education eligibility evaluations during fifth grade are not being assessed for an SLD in ORF; rather, they are more often being assessed for a reading comprehension or other higher level application of skills disability. Students with significant deficits in ORF are usually assessed at an earlier grade level. The lack of prediction at the fifth grade level may also be due to a random error in sampling because of less variability in the population at the fifth grade level. This artificial restriction of range may exist because students who were weaker readers were identified for special education at an earlier grade level, and therefore, not included in the current sample. Another possible reason for the restricted range is that good instruction and maturation between the earlier grades and fifth grade resulted in a restricted range of the data.

Data from a previous study resulted in inconsistent correlations at the fifth grade level when comparing ORF CBM data to group standardized assessments (Shapiro et al., 2006). The strength of the predictive criterion-related validity correlation coefficients when comparing fifth grade ORF CBM data to the state assessment ranged from .25 for one district to .70 in another district. When comparing the correlation between ORF CBM scores to the SAT-9 and the MAT-8 comprehension scores for second and fourth grade students, the predictive criterion-related validity correlation coefficients ranged from .59 to .74. The current research results suggest that no MLR model including the

DIBELS Next ORF Rate and Accuracy scores predicted the WIAT-III ORF or Reading Comprehension scores.

Previous researchers have investigated the predictive criterion-related validity for the DIBELS ORF assessment (Munger, 2010; Powell-Smith et al., 2011). Munger (2010) compared the predictive criterion-related validity of the first grade DIBELS 6th Edition to third grade reading comprehension measures including the WIAT-II Reading Comprehension subtest, the GRADE Reading Comprehension subtest, and NYSELA. The correlations between the DIBELS 6th Edition ORF WC and all of the reading comprehension measures ranged from moderate to strong. The correlation coefficient for the predictive relationship between the DIBELS 6th Edition ORF WC (first grade) and the WIAT-II Reading Comprehension subtest (third grade) was .66. The relationship between the predictive measure (DIBELS 6th Edition ORF WC) and the 3rd grade criterion measure (WIAT-II Reading Comprehension was strong, whereas no predictive MLR model including the DIBELS Next WC and Accuracy Percentage predicted the WIAT-III Reading Comprehension scores at the 3rd grade level. The predictive measure DIBELS 6th Edition ORF WC in the Munger (2010) study was administered two years prior to the WIAT-II, as opposed to the approximate two-week latency between the DIBELS Next ORF WC and Accuracy Percentage and the WIAT-III Reading Comprehension utilized in the present study. Therefore, comparison of results from these two studies should be made with caution.

Researched conducted a more recent study to investigate the criterion-related validity of the current edition of the DIBELS, the DIBELS Next ORF assessment, to the GRADE and the Standard 4th Grade Reading Passage in the NAEP 2002 Special Study

of Oral Reading as the criterion measures(Powell-Smith et al., 2011). Students

completed the GRADE at the end of the school year, and therefore, the beginning of the

year and middle of the year DIBELS Next ORF benchmark scores served as the

predictive criterion-related validity measures. The predictive criterion-related validity of

the fall and winter DIBELS Next ORF WC score to the GRADE Total Test score fell in

the moderate-strong to strong range. The predictive validity of the fall and winter

DIBELS Next ORF Accuracy Percentage score to the GRADE Total Test score fell in

the moderate to strong range. Overall, the predictive criterion-related validity for the

DIBELS Next ORF WC ranged from moderate-strong to strong and for the DIBELS Next

ORF Accuracy Percentage ranged from moderate to strong when compared to the

GRADE Total Test score.   In the current study, the correlation strengths for all grade

level results (second through fifth grade) when using ORF scores to predict reading

comprehension scores fell within the moderate-strong to strong range. The MLR

models, including the DIBELS Next ORF Rate and Accuracy Percentage scores, in the

current study were only significant for the second and fourth grade levels when

predicting the WIAT-III Reading Comprehension scores.

Implications for the Field of School Psychology

There are two primary approaches used to identify an SLD: Psychoeducational

Assessment and Response to Intervention (RTI). Within both approaches, early

intervening teams use curriculum based measurement (CBM) data, such as the DIBELS

Next, to identify which students should undergo a special education eligibility evaluation.

Professionals who use the Psychoeducational Assessment approach analyze extant

data and individually-administered standardized tests, such as the WIAT-III, when

109

determining eligibility for special education. When using the RTI approach, professionals commonly use the DIBELS Next to screen for disabilities and monitor progress within an implemented intervention. Some professionals use the DIBELS Next data in conjunction with standardized tests such as the WIAT-III; although, other professionals use only DIBELS Next data when determining if a student meets the criteria for special education eligibility as evidencing an SLD.

The results of the current study have implications for both SLD determination models; however, the implications should be interpreted with caution due to the small sample size used in the current study. These results are important within the context of a Psychoeducational Assessment model because they determine the extent to which school personnel can confidently predict how a student will perform when undergoing a formal psychoedcuational evaluation that includes the WIAT-III. Within an RTI model, these results are helpful in evaluating the degree to which a comprehensive evaluation is warranted. Overall, the current results help determine the confidence level that school personnel should have for the DIBELS ORF assessment as a valid measure of reading fluency and rate and a predictive measure for reading comprehension.

The results of this study support the DIBELS Next ORF assessment as a significant predictor to the WIAT-III ORF and Reading Comprehension subtest scores when analyzing data as a group for grades 2 to 5. However, when analyzing the grades separately, correlations were not significant at every grade level, suggesting that practitioners should be cautious when interpreting DIBELS Next ORF data. More specifically, the DIBELS Next ORF Accuracy Percentage and WC scores were not predictive of the WIAT-III ORF or Reading Comprehension scores at the fifth grade

level. Additionally, the DIBELS Next ORF WC and Accuracy Percentage scores were only predictive at the second and fourth grade level for the WIAT-III Reading Comprehension scores. Although the results provide stronger empirical justification for early intervening teams to use the DIBELS Next ORF scores to make important decisions including a referral for a special education eligibility evaluation or identification of a disability, these decisions, especially for fifth grade students, should be made with caution.

Since the results of this study do not support the DIBELS Next ORF score as strong predictors of the WIAT-III ORF and Reading Comprehension scores, future practitioners should consider alternative screening measures for oral reading fluency and reading comprehension at the fifth grade level. More specifically, they should consider oral reading assessments allowing the evaluator to hear the child read for a longer amount of time. The one minute time limit on the DIBELS Next ORF assessment may have been a factor for the lack of predictive strength in the current study. When screening for reading comprehension, practitioners should consider using a brief cloze technique or a brief computerized multiple choice reading comprehension measure. When evaluating fifth grade students for special education practitioners should use a standardized, norm-reference measure to assess oral reading fluency and reading comprehension. Due to the lack of support for the predictive validity of the DIBELS Next ORF scores for the WIAT-III Reading Comprehension scores at the third grade level, practitioners may also want to consider an alternative screening measure for reading comprehension deficits at this grade level. More specifically, they should also consider using a brief cloze technique or a brief computerized multiple choice reading

comprehension measure. When evaluating third grade students for special education practitioners should use a standardized, norm-referenced measure to assess reading comprehension.

The results of the current study additionally suggest that practitioners may want to be more selective when choosing to use the DIBELS Next ORF to make important decisions such as a referral for a special education eligibility evaluation and/or an identification of an SLD in reading. The current results validated the use of DIBELS Next ORF data as predictors of the WIAT-III ORF subtest for second through fourth grade and for the WIAT-III Reading Comprehension subtest for second and fourth graders. However, the use of the DIBELS Next-III ORF assessment as a predictor of the WIAT-III ORF or Reading Comprehension subtests was not validated for fifth grade students. Practitioners should be cautious when analyzing DIBELS Next ORF data to determine if students should be referred for special education eligibility evaluations or when exclusively used for SLD identification. Future research may be needed to determine a brief assessment, similar to the DIBELS Next, for pre-referral teams and psychologists functioning under the RTI model for SLD eligibility. Additionally, researchers should determine combinations of predictors that will overall improve the predictive validity of WIAT-III scores, rather than relying on one stand-alone measure. Using only one predictive measure, such as the DIBELS Next, underscores the ethical obligations that school teams have to analyze multiple data sources when making a special education eligibility determination.

Although the current results indicate that the DIBELS Next ORF assessment has strong predictive validity for the WIAT-III ORF subtest, practitioners should proceed with

caution when using the DIBELS Next ORF assessment as the sole standardized measurement to determine an SLD in the areas of oral reading fluency and reading comprehension. Although the DIBELS Next ORF assessment appears to be a good predictor of the WIAT-III ORF subtest for grades 2 through 4, the DIBELS Next ORF assessment is a brief assessment that does not provide the practitioner much opportunity to analyze a student's reading skills. The educator administering the DIBELS Next ORF only hears the student read for one minute and most likely will not hear a student read an entire passage. In the current study, the researcher only collected one DIBELS Next ORF data point for each student. It is recommended that practitioners analyze the trend of more than one DIBELS Next ORF points when assessing a student's reading skills and making high-stakes decisions.   Previous literature supports the use of multiple CBM data points (approximately 7 to 10) when analyzing student growth; however, not enough empirical evidence exists to determine the exact number of data points necessary to accurately analyze a trend (Ardoin et al., 2013). The WIAT-III requires a student to read aloud at least two passages until completion allowing a practitioner the opportunity to hear a student read for a longer amount of time and to conduct a more thorough error analysis.

Even though the DIBELS Next ORF assessment predicted the scores on the WIAT-III Reading Comprehension subtest for grades 2 and 4, practitioners should proceed with caution when solely using the DIBELS Next ORF scores to identify an SLD in reading comprehension. ORF skills are predictive of reading comprehension skills but ORF skills and reading comprehension skills remain different constructs (Fuchs et al., 2001; Fuchs et al., 1988; Valencia et al., 2010). It is recommended that

practitioners use an assessment that directly measures reading comprehension when identifying a student with an SLD in reading comprehension.

Although this study only used one type of ORF CBM (i.e., the DIBELS Next ORF) it is reasonable to conclude that results would be similar for many other ORF CBM measures. All commercially-available ORF CBM measures assess ORF similarly: the number of words read correctly per minute and the accuracy percentage of oral reading; however, each product use different passages representing passages of potentially different difficulty levels. Consequently, the predictive criterion-related validity of other ORF CBM measures remains unknown. Despite the similar administration and scoring techniques shared among ORF CBM, future research is recommended using different ORF CBM measures as the predictors of WIAT-III ORF and Reading Comprehension subtests.

Limitations

A few limitations to the current study be considered when making generalizations to other samples and when integrating these results with previous research. A limitation of this study is that it only investigated one type of validity (predictive criterion-related validity) for the DIBELS Next ORF. More psychometric evidence is needed for the DIBELS Next, especially from authors independent of the product. Several criterion-related validity research studies exist for the previous editions of the DIBELS; however, DIBELS Next criterion-related research is sparse. Content validity is important because it establishes the extent to which an assessment measures the content or subject area it is intended to assess. Construct validity is the extent to which an assessment accurately measures an underlying theoretical concept. Content and construct validity

114

research would provide needed empirical support that the DIBELS Next is, in fact, a measure of oral reading fluency skills. The current study provides some independent evidence of the psychometric merits of DIBELS Next.

Three limitations exist regarding the sample. First, the small sample size used in this study undermined the statistical power needed to fully appraise the presence of statistically significant relationships, possibly causing differences to not be found even if they do exist.   Additionally, the researcher used a convenience sample, which limits the generalizability of the current findings to the general population. Lastly, the sample only consisted of students referred for special education eligibility evaluations. Therefore, it is unknown whether or not the DIBELS Next ORF scores are strong predictors for the WIAT-III ORF and Reading Comprehension scores among typically-developing students. Typically-developing students, however, would most likely not need to be evaluated with the WIAT-III, lowering the importance of understanding the predictive strength of the DIBELS Next ORF for the WIAT-III ORF and Reading Comprehension scores for that population. While the current sample was skewed, it included students with DIBELS Next ORF scores who most often are evaluated using the WIAT-III. These methodological shortcomings regarding the sample limit generalization to a broader population.

Recommendations for Future Research

Researchers may be interested in investigating other types of validity beyond the current analysis of predictive criterion-related validity. More specifically, researchers could assess the concurrent criterion-related validity by measuring the correspondence between the DIBELS Next ORF and another test, which would serve as the criterion or

principle measure, administered at the same time. Additionally, an investigation of content validity and construct validity would determine if the DIBELS Next ORF assessment represents the construct of oral reading fluency. Further research may provide educational professionals with empirical support to conclude that the DIBELS Next ORF assessment accurately measures ORF skills.

Researchers may also want to assess the predictive criterion-related validity with a more expansive the sample that the one used in the current study. They may want to focus on increasing the sample size, randomizing the sample selection, or including typically-developing students. A larger sample size would result in substantially stronger statistical power and more confidence in interpreting results. A randomized sampling technique would allow findings to generalize to a broader population than possible in the current study. Lastly, a randomized sample should include students who are being evaluated for special education eligibility along with typically-developing students. This expansion of the sample will allow researchers to determine if the DIBELS Next ORF scores are a strong predictor for typically-progressing students in addition to those for whom educators have concerns about reading skill acquisition. Without further research, practitioners cannot be confident the DIBELS Next ORF assessments accurately identify all struggling readers. The absence of comprehensive psychometric study of DIBELS Next ORF renders high-stakes decisions about students (i.e., identification of SLD or referral to comprehensive psychoeducational evaluation) suspect at best.

Summary

This study evaluated the predictive strength of the DIBELS Next ORF to the WIAT-III ORF and Reading Comprehension subtests. The results suggested that the DIBELS Next ORF WC and Accuracy Percentage scores predicted the WIAT-III ORF Rate; however, only the DIBELS Next ORF WC scores significantly predicted the WIAT-III ORF Accuracy Percentage and Reading Comprehension when data from all grades were aggregated. Additionally, the DIBELS Next ORF WC scores predicted the WIAT-III ORF Rate scores at the second, third, and fourth grade level with the strongest correlations being at the second and fourth grade level. The DIBELS Next ORF Accuracy Percentage scores predicted the WIAT-III ORF Accuracy scores at the second level, and the DIBELS Next ORF WC predicted the WIAT-III Accuracy Percentage scores at the third and fourth grade level with no significant correlation strength differences between grade levels. The DIBELS Next ORF WC predicted the WIAT-III Reading Comprehension scores at the second grade level and the DIBELS Next ORF Accuracy Percentage predicted the WIAT-III Reading Comprehension scores at the fourth grade level with no significant correlation strength differences between grade levels.

This chapter also addressed limitations of the current study, recommendations for future research, and implications for the field of school psychology. For the current study, limitations exist regarding the investigation of only one type of validity and a narrow sample from which conclusions were drawn. Future researchers should investigate other types of validity beyond the current analysis of predictive criterion-related validity. Future work is needed to appraise the psychometric merits of DIBELS

117

Next ORF using a larger sample size, a randomized sample, or inclusion of typically-developing students. Despite the limitations of the current study, results suggest that practitioners may want to be more selective when exclusively using the DIBELS Next ORF to make high-stakes decisions such as a referral for a special education eligibility evaluation and/or an SLD identification.

References

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*, 1-18. doi:10.1016/j.jsp.2012.09.0049

Assistance to States for the Education of Children With Disabilities, 34 C.F.R. § 300 (2006).

Blachowicz, C., L., Leu, D. L., Juel, C., Pearson, P. D., Paratore, J. R., Sebesta, S. L., … Kame'enui, E. J. (2008). *Scott Foresman Reading Street.* Upper Saddle River, NJ: Scott Foresman.

Breaux, C. K. (2009). *WIAT-III technical manual*. San Antonio, TX: NCS Pearson.

Bulmer, M. G. (1979). *Principals of statistics*. Mineola, NY: Dover.

Burns, M. K., & Gibbons, K. A. (2008). *Implementing response-to-intervention in elementary and secondary schools*. New York, NY: Routledge.

Burns, M. K., Wiley, H. J., & Viglietta, E. (2008). Best practices in implementing effective problem-solving teams. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 1633-1644). Bethesda, MD: National Association of School Psychologists.

Burns, T. G. (2010). Wechsler Individual Achievement Test-III: What is the 'gold standard' for measuring academic achievement? *Applied Neuropsychology, 17*, 234-236. doi:10.1080/09084282.210.499803

Christ, T. J., & Silberglitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review, 36*, 130-146.

Coutinho, M. J., & Oswald, D. P. (2005). State variation in gender proportionality in special educartion: Findings and recommendations. *Remedial and Special Education, 26*, 7-15. doi:10.1177/07419325050260010201

Cummings, K. D., Wallin, J., Good, R. H., & Kaminski, R. A. (2007). *The DMG Passage Difficulty Index*. Eugene, OR: Dynamic Measurement Group.

Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading*. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences.

Davenport, E. C., Davison, M. L., Kwak, N., Peterson, K., Irish, M. L., Chan, C. K., . . . ChenWu, Y. (2002). *The Minnesota Basic Skills Test: Performance gaps for 1996 to 2001 on the reading and mathematics tests, by gender, ethnicity, limited English proficiency, individualized education plans, and socio-economic status*. Minneapolis, MN: Office of Educational Accountability, College of Education and Human Development, University of Minnesota.

Delaware Department of Education. (2013). School profiles. Retreived from http://profiles.doe.k12.de.us/SchoolProfiles

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*, 184-192. doi:10.1177/00224669030370030801

Deno, S.L., & Merkin, P. K. (1977*). Data-based program modification: A manual*.

 Reston, VA: Counsel for Exceptional Children.

Drost, E. A. (2011). Validity and reliability in social science research. *Education*

 Research and Perspectives, 38(1), 105-123.

Edformation. (2005). *AIMSweb progress monitoring and improvement system.*

 Retrieved from http://www.aimsweb.com/

Education for All Handicapped Children Act, Pub. L. No. 94-142 § U. S. C. 1400 et seq.,

 6 Stat. (1975).

Fiorello, C. A., & Primerano, D. (2005). Research into practice: Cattell-Horn-Carroll

 cognitive assessment in practice: Eligibility and program development issues.

 Psychology in the Schools, 42, 525-536. doi: 10.1002/pits.20089

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities:*

 From identification to intervention. New York, NY: Guildford Press.

Flynn, J. R. (1994). IQ gains over time. In R. J. Sternberg (Ed.), *Encyclopedia of human*

 intelligence (pp. 617-623). New York, NY: Macmillan.

Fuchs, L. S., Fuchs, D., & Hosp, M. K. (2001). Oral reading fluency as an indicator of

 reading competence: A theoretical, emperical, and historical analysis. *Scientific*

 Studies of Reading, 5, 239-256. doi:10.1207/S1532799XSSR0503_3

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading

 comprehension measures. *Remedial and Special Education, 9*, 20-28.

 doi:10.1177/074193258800900206

Fuchs, L. S., Tindal, G., & Deno, S. L. (1984). Methodological issues in curriculum-based reading assessment. *Assessment for Effective Intervention, 9*, 191-207. doi:10.1177/073724778400900401

Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, J. C. (1982). *Stanford Achievement Test*. Iowa City, IA: Harcourt Brace Jovanovich.

Goffreda, C. T., & DiPerna, J. C. (2010). An empirical review of psychometric evidence for the dynamic indicators of basic early literacy skills. *School Psychology Review, 39*, 463-483.

Goffreda, C. T., & Pedersen, J. A. (2009). Preventative screening for early readers: Predictive validity of the Dynamic Indicators of Early Literacy Skills (DIBELS). *Psychology in the Schools*, 46, 539-552.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Education Achievement.

Good, R. H., & Kaminski, R. A. (2011). *Dynamic indicators of basic early literacy skills* (Next ed.). Eugene, OR: Dynamic Measurement.

Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. J. (2011). *DIBELS Next technical manual*. Eugene, OR: Dynamic Measurement.

Goodman, K. S. (1969). Analysis of oral reading miscues: Applied psycholinguistics. *Reading Research Quarterly, 5*, 9-30.

Haager, D., Klingner, J., & Vaughn, S. (2007). *Evidence-based reding practices for response to intervention*. Baltimore, MD: Paul H. Brookes.

Harcourt Brace Educational Measurement. (1996). *Stanford Achievement Test* (9th ed.). San Antonio, TX: Harcourt Assessment.

Harcourt Brace Educational Measurement. (2000). *Metropolitan Achievement Test* (8th ed.). San Antonio, TX: Harcourt Assessment. doi:10.1002/dys.389

Hopkins, W. G. (2000). *A new view of statistics*. Retrieved from http://www.sportsci.org/resource/stats/effectmag.html

Hosp, M. K., & MacConnell, K. L. (Eds.). (2008). Best practices in curriculum-based evaluation in early reading. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 103-114). Bethesda, MD: National Association of School Psychologists.

Ikeda, M. J., Neessen, E., & Witt, J. C. (2008). Best practices in universal screening. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 377-398). Bethesda, MD: National Association of School Psychologists.

Ikeda, M. J., Tilly, W. D., Stumme, J., Volmer, L., & Allison, R. (1996). Agency-wide implimentations of problem solving consultation: Foundations, current implementation, and future directions. *School Psychology Quarterly, 11*, 228-243.

Individuals With Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446 § U. S. C. 1400 et seq. (2004).

Institute of Education Sciences. (2013). National center for education statistics. Retreived from http://nces.ed.gov/fastfacts/display.asp?id=64

Joseph, L. M. (2008). Best practices on interventions for students with reading problems. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology*

(5th ed., pp. 1163-1180). Bethesda, MD: National Association of School

Psychologists.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman test of educational achievement*.

San Antonio, TX: Pearson.

Kavale, K. A. (2002). Discrepancy models in the identification of learning disability. In R.

Bradley, L. Danielson & D. P. Hallahan (Eds.), *Indentification of learning*

disabilities: Research to practice (pp. 369-426). Mahwah, NJ: Lawernce Erlbaum

Associates.

Kavale, K. A., & Reese, L. (1992). The character of learning disabilities: An Iowa

profile. *Learning Disability Quarterly, 15*, 74-94.

Kovaleski, J. F., VanDerHeyden, A. M., & Shapiro, E. S. (2013). *The RtI approach to*

evaluating learning disabilities. New York, NY: Guilford Press.

LaBerge, D., & Samuels, S. (1974). Toward a therory of automatic information

processing in reading. *Cognitive Psychology, 6*, 293-323.

Maki, K. E., & Floyd, R. (2014, February). *Comparing state specific learning*

disability (SLD) eligibility criteria and prevalence rates. Paper session presented

at the National Association of School Psychologists Annual Convention,

Washington, D.C.

Marston, 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 (pp. 18-78). New

York, NY: Guilford Press.

Marston, D., Mirkin, P., & Deno, S. (1984). Curriculum-based measurement: An
alternative to traditional screening, referral, and identification. *The Journal of
Special Education, 18,* 109-117. doi:10.1177/002246698401800204

McDougal, J. L., Graney, S. B., Wright, J. A., & Ardoin, S. P. (2010). *RTI in practice: A
practical guide to implementing effective evidence-based interventions in your
school.* Hoboken, NJ: John Wiley & Sons.

Munger, K. A. (2010). *A longitudinal follow-up study of the Dynamic Indicators of Basic
Early Literacy Skills (DIBELS) as a predictor of third grade reading comrpehenion*
(Doctoral dissertation). Available from ProQuest Dissertations and Theses
database. (UMI No. 3459383)

Naglieri, J. A. (2003). Current advances in assessment and intervention for children with
learning disabilities. *Advances in Learning and Behavioral Disabilities, 16*, 163-
190.

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 Pulication No. 00-4769). Retrived from
http://www.nichd.nih.gov/publications/nrp/smallbook.cfm? renderforprint=1

No Child Left Behind Act of 2001, Pub. L. No. 107-110 § U.S.C. 20 (2001).

Osborne, Jason & Elaine Waters (2002). Four assumptions of multiple regression that
researchers should always test. *Practical Assessment, Research & Evaluation*,
8(2). Retrieved from http://PAREonline.net/getvn.asp?v=8&n=2

Petscher, Y., & Kim, Y. S. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal School Psychology, 49*, 107-129. doi:10.1016/j.jsp.2010.09.004

Powell-Smith, K. A., Good, R. H., Atkins, T. (2010). *DIBELS Next oral reading fluency readability study* (Tech. Report No . 7). Eugene, OR: Dynamic Measurement.

Powell-Smith, K. A., Good, R. H., Latimer, R. J., Dewey, E. N., & Kaminski, R. A. (2011). *DIBELS Next benchmark goals* study (Tech. Report No . 11). Eugene, OR: Dynamic Measurement.

Psychological Corporatation. (1992). *Wechsler Individual Achievement*. San Antonio, TX: Author.

President's Commission on Excellence in Special Educaiton. (2002). *A new era: Revitalizing special education for children and their families.* Jessup, MD: U.S. Department of Education.

Rosa's Law, Pub. L. No. 111-256§ S. 781 (2010).

Rutter, M., Caspi, A., Fergusson, D., Goodman, L. J., Maughan, B., Moffitt, T. E., . . . Carroll, J. (2004). Sex differences in developmental reading disability: New findings from 4 epidemolgical studies. *Journal of American Medical Association, 291*, 2007-2012.

Salvia, J., & Ysseldyke, J. E. (2007). *Assessment in special and inclusive education* (10th ed.). Boston, MA: Houghton Mifflin.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santora, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and

standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*, 19-35. doi:10.1177/0734282905285237

Shinn, M. R. (2008). Best practices in using curriculum-based measurement in a problem solving model. *Best practices in school psychology* (5th ed., pp. 234 - 262). Bethesda, MD: National Association of School Psychologists.

Siegel, L. S. (1989). IQ is irrelevant to the definition of learning disabilties. *Journal of Learning Disabilities, 22*, 469-478. doi:10.1177/002221948902200803

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension.* Santa Monica, CA: RAND Education.

Statistics Solutions (2014). *Assumptions of multiple linear regression*. Retrived from http://www.statisticssolutions.com/assumptions-of-multiple-linear-regression/

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawerence Erlbaum Associates.

Tierney, R. J. (1998). Literacy assessment reform: Shifting beliefs, principled possiblities, and emerging practices. *The Reading Teacher, 51*, 374-390.

Tilly, W. D. (2008). The evolution of school psychology to science-based practice: Problem solving and the three-tiered model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 17 - 36). Bethesda, MD: National Association of School Psychologists.

Tindal, G., & Alonzo, J. (2009). *easyCBM*. Eugene: University of Oregon.

Truscott, S. D., Cohen, C. E., Sams, D. P., Sanborn, K. J., & Frank, A. J. (2005). The curent state(s) of prereferral intervention teams: A report from two national

surveys. *Remedial and Special Education, 26*, 130-140.

doi:10.1177/07419325050260030101

Valencia, S. W., Smith, A. T., Reece, A. M., Li, M., Wixson, K. K., & Newman, H. (2010).

Oral reading fluency assessmnet: Issues of construct, criterion, and

consequential validity. *Reading Research Quarterly, 45*, 270-291.

doi:10.1598/RRQ.45.3.1

Wechsler, D. (2001). *Wechsler Individual Achievement Test* (2nd ed.). San Antonio, TX:

Pearson.

Wechsler, D. (2009). *Wechsler Individual Achievement Test* (3rd ed.). San Antonio, TX:

Pearson.

Wehmeyer, M., & Schwartz, M. (2001). Disproportionate representation of males in

special education services: Biology, behavior, or bias? *Education and Treatment

of Children, 24*, 28-45.

Williams, K. T. (2001). *Group Reading Assessment and Diagnostic Evaluation*. New

York, NY: Pearson.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson Tests of

Achievement* (3rd ed.). Itasca, IL: Riverside.

Zirkel, P. A., & Thomas, L. B. (2010). State laws and guidelines for implementing RTI.

Teaching Exceptional Children, 43(1), 60–73.

Appendix A

Institutional Review Board (IRB) Approval Letter

Indiana University of Pennsylvania
www.iup.edu

Institutional Review Board for the
Protection of Human Subjects
School of Graduate Studies and Research
Stright Hall, Room 113
210 South Tenth Street
Indiana, Pennsylvania 15705-1048

P 724-357-7730
F 724-357-2715
irb-research@iup.edu
www.iup.edu/irb

April 19, 2013

Jaclyn Purcell
229 Missouri Lane
Kulpmont, PA 17834

Dear Ms. Purcell:

Your proposed research project, "The Predictive Strength of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Next Oral Reading Fluency (ORF) Assessment to the Wechsler Individual Achievement Test, Third Edition (WIATT-III) Oral Reading Fluence (ORF) Subtest and the Reading Fluency (ORF) Subtest and the Reading Comprehension Subtest for Students Referred for Special Education Evaluations," (Log No. 13-092) has been reviewed by the IRB and is approved as an expedited review for the period of April 18, 2013 to April 18, 2014.

It is also important for you to note that IUP adheres strictly to Federal Policy that requires you to notify the IRB promptly regarding:

1. any additions or changes in procedures you might wish for your study (additions or changes must be approved by the IRB before they are implemented),
2. any events that affect the safety or well-being of subjects, and
3. any modifications of your study or other responses that are necessitated by any events reported in (2).

Should you need to continue your research beyond April 18, 2014 you will need to file additional information for continuing review. Please contact the IRB office at (724) 357-7730 or come to Room 113, Stright Hall for further information.

Although your human subjects review process is complete, the School of Graduate Studies and Research requires submission and approval of a Research Topic Approval Form (RTAF) before you can begin your research. If you have not yet submitted your RTAF, the form can be found at
http://www.iup.edu/page.aspx?id=91683 .

This letter indicates the IRB's approval of your protocol. IRB approval does not supersede or obviate compliance with any other University policies, including, but not limited to, policies regarding program enrollment, topic approval, and conduct of university-affiliated activities.

I wish you success as you pursue this important endeavor.

Sincerely,

John A. Mills, Ph.D., ABPP
Chairperson, Institutional Review Board for the Protection of Human Subjects
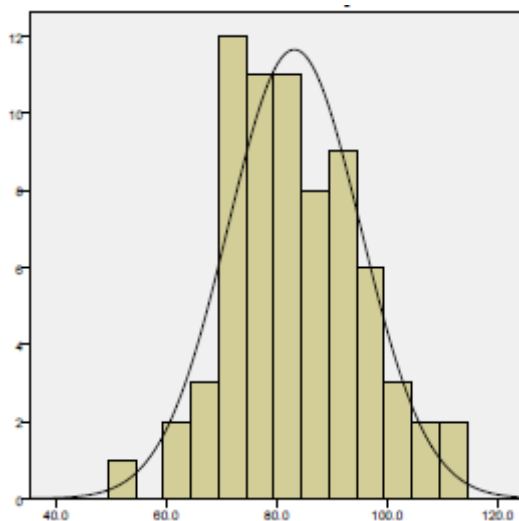Professor of Psychology

JAM:jeb

Cc: Dr. Timothy Runge, Dissertation Advisor
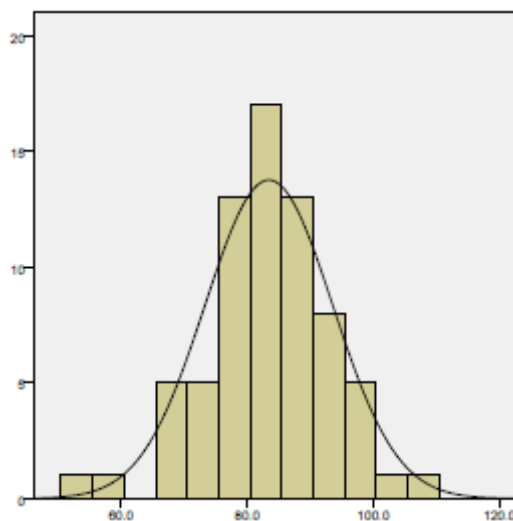   Ms. Brenda Boal, Secretary

Appendix B

Histograms for WIAT-III ORF Accuracy, Rate, and Reading Comprehension; DIBELS
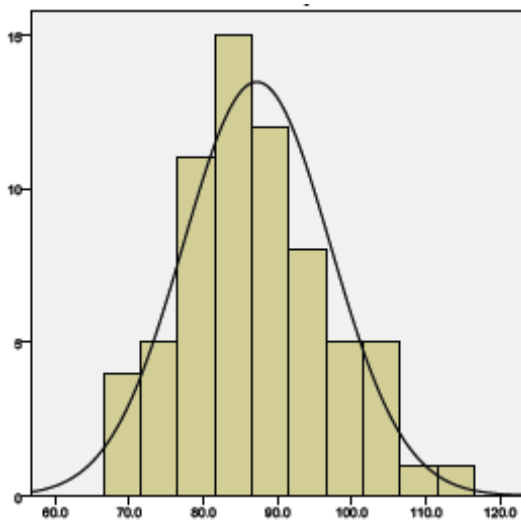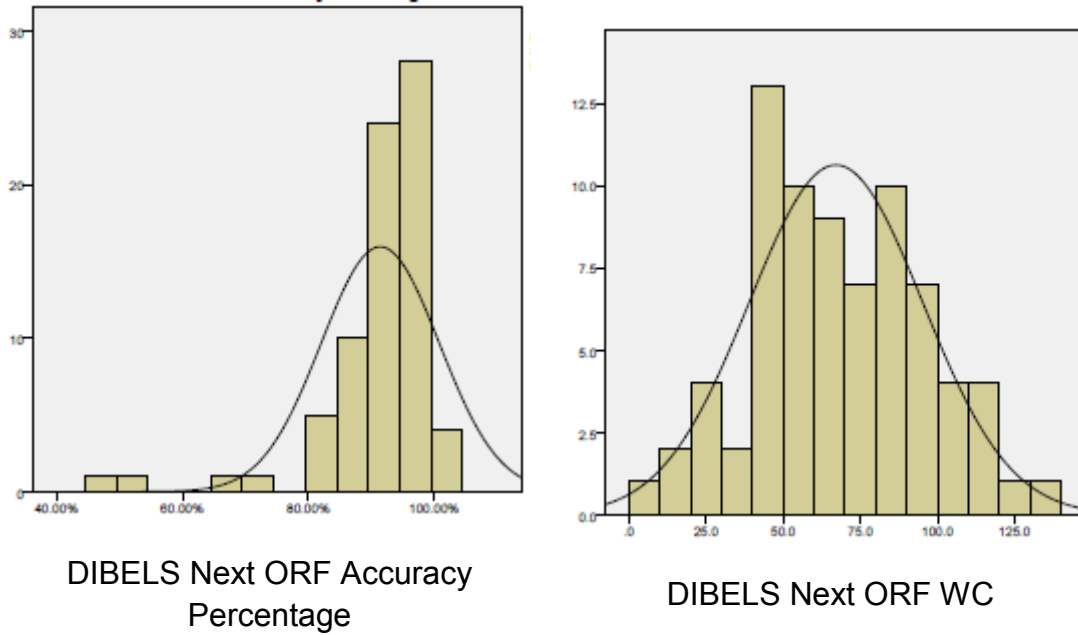
Next ORF Word Correct (WC) and Accuracy Percentage

WIAT-III ORF Accuracy

WIAT-III ORF Rate

WIAT-III Reading Comprehension

Figure 1. WIAT-III histograms.

Appendix B

Histograms for WIAT-III ORF Accuracy, Rate, and Reading Comprehension;

DIBELS Next ORF Word Correct (WC) and Accuracy Percentage

DIBELS Next ORF Accuracy
Percentage

DIBELS Next ORF WC

Figure 2. DIBELS Next histograms.

Appendix C

Bivariate Scatterplot Matrix for WIAT-III ORF Rate, Accuracy, Reading Comprehension; DIBELS Next ORF Words Correct (WC) and Accuracy Percentage
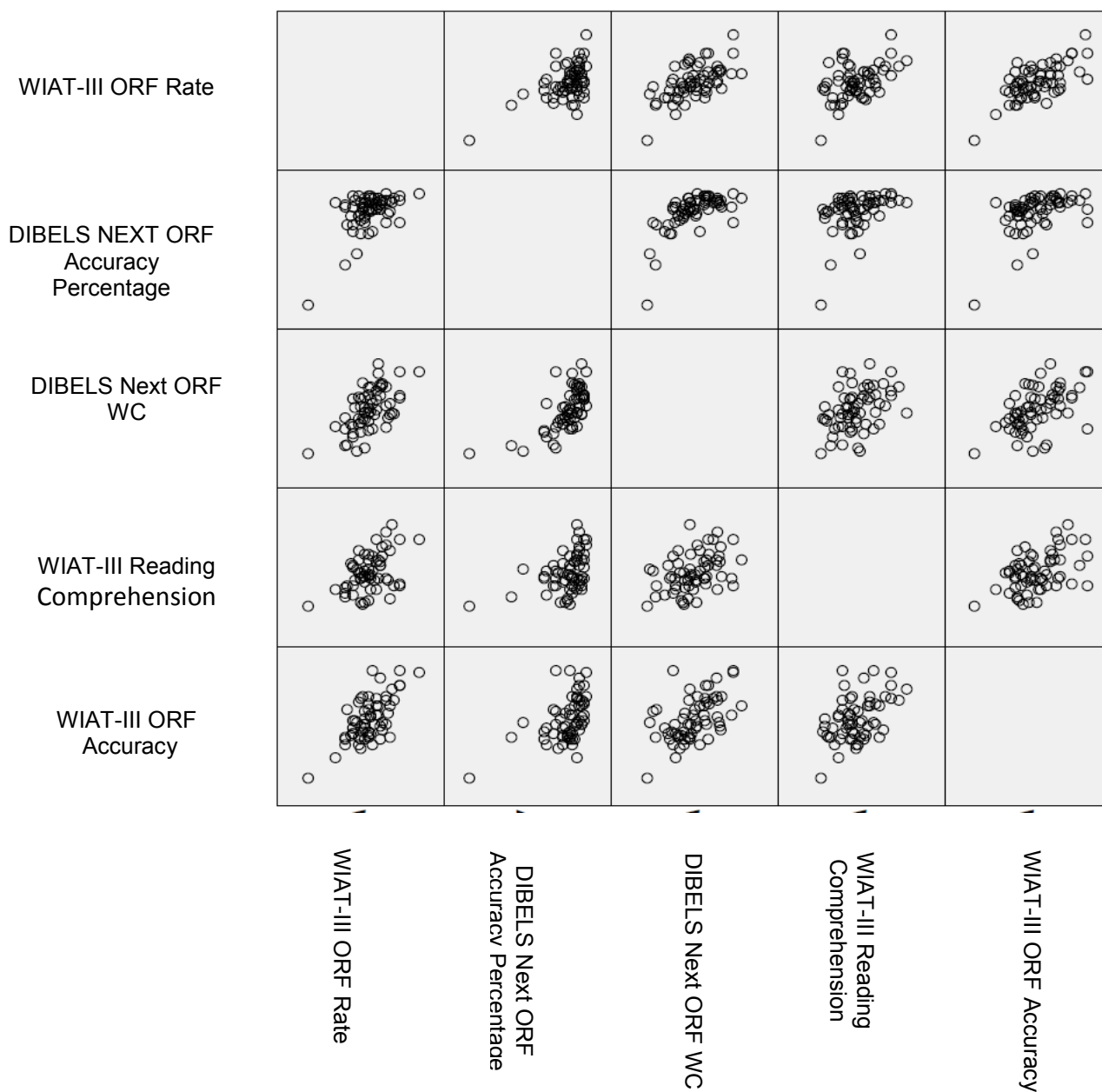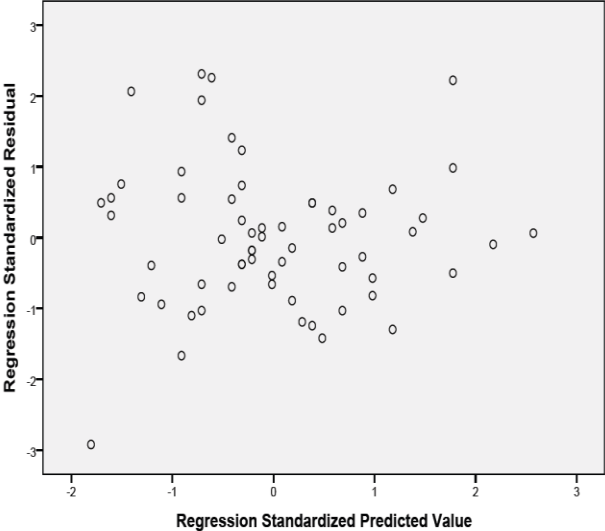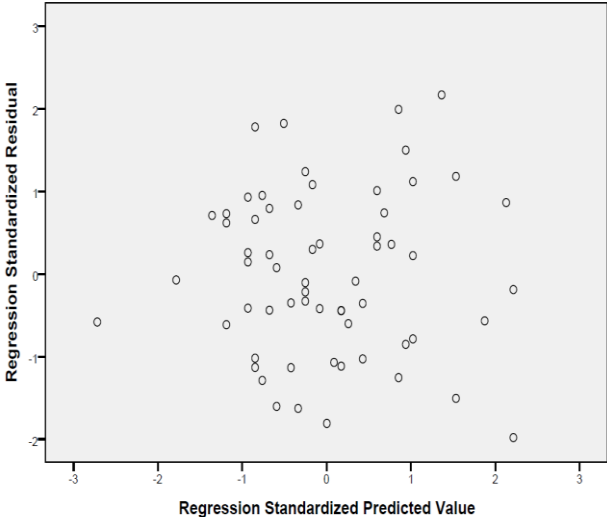

Figure 3. Bivariate scatterplot matrix for all variables.

Appendix D

Residual Plots for WIAT-III ORF Rate, Accuracy, Reading Comprehension; DIBELS Next ORF Words Correct (WC) and Accuracy Percentage
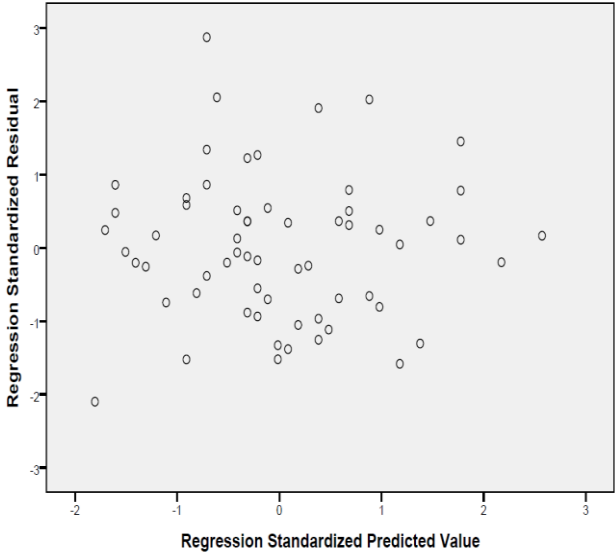
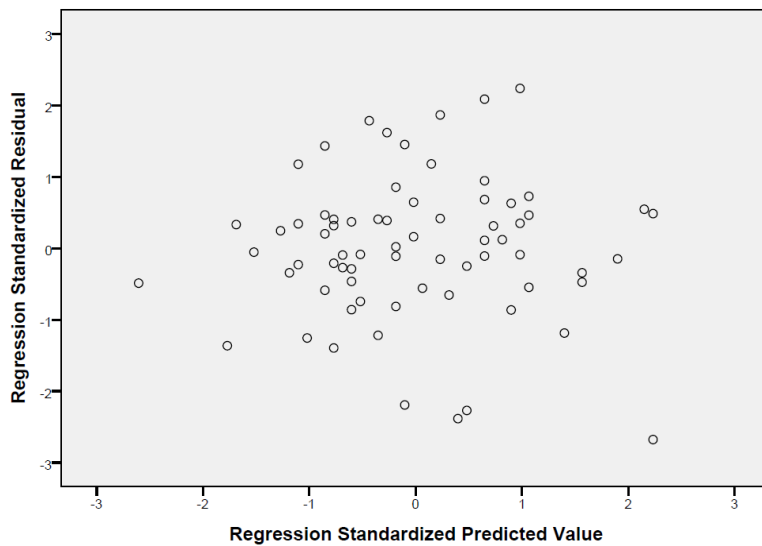WIAT-III ORF Accuracy
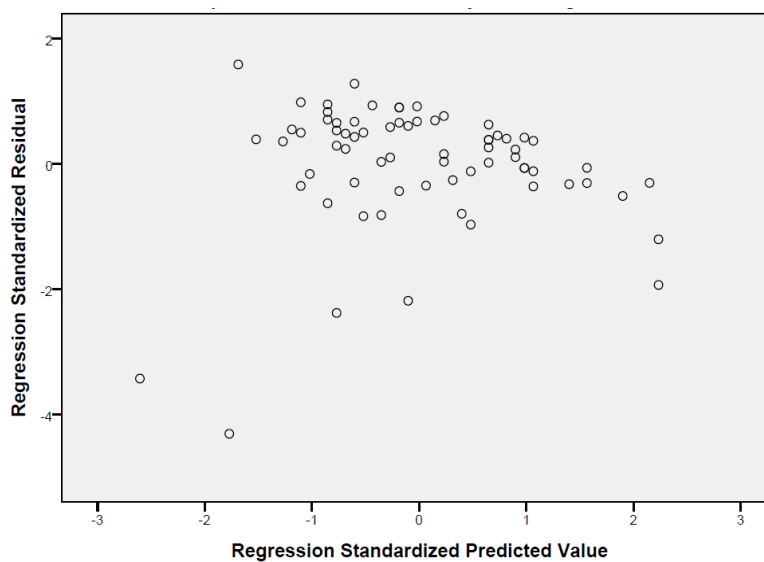
WIAT-III ORF Rate

WIAT-III Reading Comprehension

Figure 4. WIAT-III residual plots.

Appendix D

Residual Plots for WIAT-III ORF Rate, Accuracy, Reading Comprehension; DIBELS Next ORF Words Correct (WC) and Accuracy Percentage

DIBELS Next ORF WC

DIBELS Next ORF Accuracy Percentage

Figure 5. DIBELS Next residual plots.