

THE PREDICTIVE RELATIONSHIP BETWEEN ORAL READING FLUENCY  
AND COMPREHENSION AS IT RELATES  
TO MINORITY STUDENTS

by

SARA D. DEBOY

A DISSERTATION

Presented to the Department of Educational Methodology, Policy, and Leadership  
and the Graduate School of the University of Oregon  
in partial fulfillment of the requirements  
for the degree of  
Doctor of Education

June 2013

DISSERTATION APPROVAL PAGE

Student: Sara D. Deboy

Title: The Predictive Relationship Between Oral Reading Fluency and Comprehension as It Relates to Minority Students

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Education degree in the Department of Educational Methodology, Policy, and Leadership by:

Dr. Gina Biancarosa	Chair
Dr. Gerald Tindal	Member
Dr. Charles Martinez	Member
Dr. Audrey Lucero	Outside Member

and

Kimberly Andrews Espy	Vice President for Research & Innovation/Dean of the Graduate School
-----------------------	--

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2013

© 2013 Sara D. Deboy

## DISSERTATION ABSTRACT

Sara D. Deboy

Doctor of Education

Department of Educational Methodology, Policy, and Leadership

June 2013

Title: The Predictive Relationship Between Oral Reading Fluency and Comprehension as It Relates to Minority Students

The prominent use of DIBELS as a screening and placement tool has provided schools and districts the ability to implement interventions and best practices for students, particularly in the primary grades. Although many studies have highlighted the predictive validity of oral reading fluency (ORF) to anticipate reading performance, few have extended that research to examine the performance of ethnic and economic subgroups as compared to non-minority peers. Disaggregating the data to study specific populations can expose whether ORF's relationship with reading comprehension depends on group membership and ultimately improve the quality of the assessment. This study examines the predictive validity of DIBELS ORF for two types of comprehension scores within a racially and ethnically diverse second grade cohort.

## CURRICULUM VITAE

NAME OF AUTHOR: Sara D. Deboy

### GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon  
University of Portland, Portland, Oregon  
Western Oregon University, Monmouth, Oregon  
University of Puget Sound, Tacoma, Washington

### DEGREES AWARDED:

Doctor of Education, 2013, University of Oregon  
Master of Education, Education Administration, 2007, University of Portland  
Bachelor of Arts, Writing, 1996, University of Puget Sound

### AREAS OF SPECIAL INTEREST:

Literacy Instruction, Curriculum, and Assessment  
Literacy Intervention and Support for Struggling Readers  
Equitable Instruction and Assessment for Minority Subpopulations  
Development of Teacher and Administrative Leaders in Schools

### PROFESSIONAL EXPERIENCE:

Elementary Principal, Lake Oswego School District, 2007 to present  
Title One Specialist, Beaverton School District, 2003 to 2007  
Classroom Teacher, Beaverton School District, 1998 to 2003

### GRANTS, AWARDS, AND HONORS:

Leadership Fellow, Education Leadership Network, University of Portland, 2013  
Chair, Lake Oswego Leadership, 2011  
Graduated with Honors, 1996 & 2007  
National Merit Scholar, 1992

## ACKNOWLEDGEMENTS

I wish to thank the many teachers and leaders that have influenced my learning and encouraged my professional growth. Special thanks to Dr. Gina Biancarosa, a thoughtful and gifted advisor who guided me through this experience with grace and understanding. To Dr. Gerald Tindal, Dr. Charles Martinez, and Dr. Audrey Lucero for their expert advice and service on my committee, I am so very thankful. To Dr. Nancy Heapes, who provided immense support and advocated for me from the beginning of the program to the end. To Dr. William Korach, who provides the model of continuous self discovery and drive, and who gave me the impetus to take the first step. To my partners in learning, Jennifer Spencer-Iiams and Suzanne West, who without their friendship and collegiality, I could not have realized this goal. To my first teachers, my mother and father, who give wise counsel and unconditional love. To Dorine, for the life lessons and care she has given to my children and friendship she has given me as I took on this challenge. And finally, to my partner and heart, Gary, who teaches me about the necessity of equity and the importance of continuing to grow as a person.

Dedicated to my children, Nicolas, Sam, and Finnegan. The light and the joy of my life,  
and the greatest gifts I have ever received.

## TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION.....	1
II. REVIEW OF THEORY AND LITERATURE.....	5
Validity of Reading Comprehension Measures.....	6
ORF Validity Overall.....	8
ORF Findings Specific to Race and Ethnicity.....	10
III. METHODS.....	15
Setting.....	15
Participants.....	16
Measures.....	20
Procedures.....	24
IV. RESULTS.....	28
Descriptive Statistics.....	28
Correlations.....	31
Regression for Both Comprehension Measures.....	31
V. DISCUSSION.....	36
Summary.....	36
Study Limitations.....	40
Implications.....	43
Additional Directions for Future Research.....	45
APPENDIX: FIGURES.....	47
REFERENCES CITED.....	59



LIST OF TABLES

Table	Page
1. Frequencies - Ethnicity and Race Reporting Categories .....	17
2. Frequencies for Ethnicity and Race Regrouped.....	19
3. Frequencies for Ethnicity and Race Subpopulations Participating in Comprehension Measures.....	20
4. Descriptive Statistics for Students Assessed with easyCBM .....	29
5. Descriptive Statistics for Students Assessed with SAT 10.....	30
6. Correlations for Spring Comprehension Measures, ORF, Ethnicity, and Race with easyCBM MCRC Above and SAT 10.....	31
7. Summary of Sequential Regression Analysis for Fall ORF, Ethnicity, and Race Predicting easyCBM MCRC Scores.....	32
8. Table of Coefficients for Sequential Regression Analysis for Fall ORF, Ethnicity, and Race Predicting easyCBM Scores.....	33
9. Summary of Sequential Regression Analysis for Fall ORF, Ethnicity, and Race Predicting SAT 10 Reading Comprehension Scores.....	34
10. Table of Coefficients for Sequential Regression Analysis for Fall ORF, Ethnicity, and Race Predicting SAT 10 Scores.....	35

## LIST OF FIGURES

Figure	Page
1. DIBELS ORF Scores for Hispanic Students Participating in easyCBM Assessment.....	47
2. Spring easyCBM Multiple Choice Reading Comprehension Scores for Hispanic Students.....	48
3. DIBELS ORF Scores for Non-Hispanic White Students Participating in easyCBM Assessment.....	49
4. Spring easyCBM Multiple Choice Reading Comprehension Scores for Non-Hispanic White Students.....	50
5. DIBELS ORF Scores for Non-Hispanic Minority Students Participating in Spring easyCBM Multiple Choice Reading Comprehension.....	51
6. Spring easyCBM Multiple Choice Reading Comprehension Scores for Non-Hispanic Minority Students .....	52
7. DIBELS ORF Scores for Hispanic Students Participating in Spring SAT 10 Reading Comprehension.....	53
8. Spring SAT 10 Reading Comprehension Score for Hispanic Students.....	54
9. DIBELS ORF Scores for Non-Hispanic White Students Participating in Spring SAT 10 Reading Comprehension .....	55
10. Spring SAT 10 Reading Comprehension Score for Non-Hispanic White Students..	56
11. DIBELS ORF Scores for Non-Hispanic Minority Students Participating in Spring SAT 10 Reading Comprehension .....	57
12. Spring SAT 10 Reading Comprehension Scores for Non-Hispanic Minority Students.....	58

## CHAPTER I

### INTRODUCTION

For over 30 years, oral reading fluency has been studied as an important indicator of reading proficiency. Studies are plentiful demonstrating the predictive nature of reading fluency scores as indicators of reading comprehension ability. Much of this research has focused on criterion-related evidence in which fluency measures are correlated with other important indicators of reading ability and achievement. This research has documented both concurrent and predictive relations in associating fluency with other important reading measures. For example, in a study regarding convergent validity of DIBELS literacy skills with word reading efficiency in first grade students, Hagan-Burke, Burke, and Crowder (2006) discuss the implications for struggling readers. They argued that summative evaluations given at the end of first grade are less useful, as many intervention opportunities have been lost; therefore, predictive assessments such as DIBELS are essential for prevention-oriented models of reading instruction. By documenting concurrent relations, teachers can then use fluency on a frequent basis so that instruction can be monitored and adjusted. By determining predictive relations, teachers can know that fluent reading is predictive of eventual success on reading measures in the future.

However, most of this research has been done with intact groups and less frequently with disaggregated groups. Within studies examining the predictive and concurrent validity of DIBELS, surprisingly little attention has been directed to student characteristics and subgroups. While some studies report student demographics related to

gender, socioeconomic status, and race, they rarely disaggregate the data by subgroups to determine the reliability and validity of the measures for these populations. In this study, my primary question is about the predictive validity of DIBELS fluency measures for understudied subgroups. To what extent do second grade students from racial/ethnic minority groups have similar validity coefficients compared to their racial/ethnic majority peers? To what extent does the predictive relationship between ORF and reading comprehension depend on student characteristics, such as race or ethnicity?

Developing sensitivity to the subpopulations within a study can provide educators with the information needed to make sure we are neither over- or under-identifying students of minority status. This study focuses on the validity of DIBELS using an extant data set to determine how minority status mediates predictions of reading comprehension proficiency. Ultimately, it is important to determine if, in using these scores to place students and determine their instruction, other systematic variance (i.e., construct irrelevant variance) is occurring within the system to avoid over- and under-identifying students due to bias. Several studies describe this result as false positives (students receiving supplemental interventions when they are not needed due to not meeting benchmark on the ORF) and false negatives (students being excluded from supplemental instruction due to ORF scores when they were, in fact, in need of intervention).

There is further reason to explore this topic when considering the research that suggests that minority students have been historically over-identified for intervention programs and subject to tracking that limited their potential for success. McPherson (2010) points to a court case in 1967 in which the U.S. Federal Circuit Court *Hobson v. Hansen* was the first legal suit to question the use of school tracking in D.C. Public

Schools from 1956 to 1966. Julius Hobson alleged that the use of testing to track students based on performance on intelligence and achievement tests resulted in discrimination against African American students and low income students. In 1956, Superintendent Dr. Carl F. Hansen implemented a school tracking system in D.C. Public Schools to place students into ability-specific course trajectories (e.g., general, regular, special academic, and honors) determined by intelligence scores, academic achievement test scores, and teacher recommendations. This sorting process was reported to typically begin on the eighth day of Kindergarten, and as there was little movement between tracks, many students remained in the same curricular track from Kindergarten throughout Grade 12. Darling-Hammond (2010) suggests that the legacy of educational inequality is alive and well, with students in poverty (many of them racial and ethnic minorities) being left to cope with an opportunity gap that leaves them with a) fewer resources, b) less experienced teachers, and c) automatic tracking of coursework that prevents many from realizing post-secondary education. Even with recent efforts in the last twenty years to desegregate public schools, tracking systems leave minority students segregated “inside the classroom” in predominately White schools as they engage in separate coursework in the lower trajectory (e.g. remedial classes) while their White peers take honors or advanced placement classes (McPherson, p. 802).

Many of the assessments highlighted in this study have become familiar in school districts across the country as indicators for future reading ability. Placement in small reading groups and in special programs in which students are given additional time with targeted reading instruction is often based on preliminary scores obtained in the primary grades. In this context, preliminary refers to assessments completed early in the year,

typically upon entrance into the grade level. These scores serve as a base line for determining growth over the course of the year. Continuous monitoring of these scores allow for both the tracking of progress in fluency and informed decision making in placement for student groupings. Students are often grouped by general ability levels and/or skill based focuses. It is quite common to find regularly scheduled school wide formative and summative assessments being administered by teams of teachers and instructional assistants to prepare for curricular planning, team teaching, flooding of assistants, and the updates of data team assessment walls.

This is not to suggest that the use of assessments is not useful and, at times, critical to understanding student progress and ability. Clearly, it is important that struggling readers are identified as early as possible so that interventions can be employed. Schilling et al. (2007) state, “fluency measures can be used not only to identify students who appear to be having substantial difficulties learning to read, but also to assess the effectiveness of instruction and/or interventions used to promote progress in reading” (p. 431). However, it is critical that we consider not only the validity of the measures in all cases, but also the inferences drawn from the measures that result in placement in reading courses that often set the tone for subsequent courses throughout a student’s school career. If we are basing our decisions to support students with interventions solely on their fluency score, we are responsible for understanding the limitations of the measure and its relevance to students in varying ethnic and racial subgroups.

## CHAPTER II

### REVIEW OF THEORY AND LITERATURE

The review of the literature concentrated on studies that employed the use of oral reading fluency scores to predict comprehension. In considering a student's rate, phrasing, and accuracy when reading aloud, it is important to recognize the impact on understanding; when rate is slow, phrasing is confusing, and accuracy is low, comprehension in turn becomes difficult. To measure a student's proficiency in fluency is to take into account the level of accuracy, rate, and phrasing to determine if students are progressing through text productively. The use of DIBELS (Dynamic Indicators of Basic Early Literacy Skills) DORF (Oral Reading Fluency) scores are used to group students as being at risk, holding some risk, or not at risk for reading success. EasyCBM is yet another tool used to measure both reading fluency through a number of varied measures and comprehension for the purposes of predicting student growth and needs over time. The third measure examined in this study, the SAT 10, is a comprehensive set of assessments for monitoring student growth and informing instruction in reading.

Examining the validity of the easyCBM and SAT 10 measures of reading comprehension is essential to understanding the usefulness of this information as a dependent variable. Likewise, the validity of the DIBELS Oral Reading Fluency Measure as an independent variable related to the comprehension measures is important because it demonstrates the known connection between ORF and comprehension utilized for intervention purposes.

## **Validity of Reading Comprehension Measures**

An important factor in considering the validity of comprehension measures is the examination of construct validity given the wide array of skills that the term “comprehension” covers. Identifying the specific construct of a comprehension assessment can be challenging, due to the fact that the process of comprehending is complex, overt, and often unobservable (Fletcher, 2006). Different comprehension assessments may include aspects of listening comprehension, retell, summarizing, drawing inferences, identifying literal facts, and many other components.

Comprehension of passages differs from comprehension at the individual word level (decoding), and in some studies, decoding was shown to account for the greatest amount of variance (Keenan, Betjemann, & Olson, 2008). Therefore, it is important that we recognize that while we tend to refer to different comprehension assessments in a manner that suggests they are interchangeable, they may measure different constructs depending on the format of the presented text (including the difficulty of the text, which is cited as the major determinant by Fletcher (2006)) and demands of the response (Pearson and Hamm, 2005). In simple terms, as Keenan et al. (2008) suggests, “comprehension is not a unitary construct” (p. 282).

Furthermore, it is necessary to consider how the assessment results are utilized. Cutting and Scarborough (2006) conducted a study in which they compared three common comprehension assessments used in the elementary setting: the Wescheler Individual Achievement Test (subtest of reading comprehension), the Gates-MacGinitie Reading Test, and the Gray Oral Reading Test and did not find the variance described in another studies regarding oral language. Unfortunately, the discrepancy emerged among



how the tests indicated a disability in children. Cumulatively, all three tests identified 43.5% of the children as having a reading comprehension deficit; however, only 9.4% of the sample was identified by all three tests, suggesting that the assessments are either not testing for the same thing or are, at the very least, imbalanced in interpretation. Keenan et al. (2008) investigated a similar comparison between the Gray's Oral Reading Test (GORT) and other common comprehension tests (Woodcock-Johnson Passage Comprehension (WJPC) subtest, Peabody Individual Achievement Test (PIAT)) and concluded that students could answer many questions on the GORT using background knowledge alone without utilizing listening comprehension skills, while the WJPC and PIAT relied mainly on decoding and comprehension at the word level. Again, these assessments are commonly used to place children in reading intervention groups and even special education, making the validity of the measures paramount to student's academic trajectories.

The measures used in this study, easyCBM Multiple Choice Reading Comprehension measure and SAT 10 Reading Comprehension measure, have similar test formats and requirements of the student reader. Both use a multiple-choice format following an independent reading of a passage. Both are administered on the computer and are untimed. There is no retell or listening comprehension components, and the examiners are not prompting or recording responses. In a recent study of the prediction of reading comprehension scores, Andreassen and Braten (2010) describe multiple choice formats such as these as the most common of all reading comprehension assessments and cite the key variations within this format to be "text availability, length of text passages, and the types of comprehension questions", such as literal, inferential, or evaluative (p.

265). For both the easy CBM MCRC and SAT 10 RC, the text is available for reference while the student is answering the comprehension questions and the types of questions are comparable. Jamgochian, Park, Nese, Lai, Sáez, Anderson, Alonzo, and Tindal (2010) describe the administration of the easyCBM MCRC measure in detail in their technical report. The length of passages used in the MCRC measure averages 900 words, while the length of passages in the SAT 10 RC is unknown. However, it is clear that the similarity in formats is a strength in utilizing and comparing these two measures as dependent variables.

### **ORF Validity Overall**

In an empirical review of 26 references that cited reliability, validity, and classification accuracy statistics for DIBELS measures, Goffreda & DiPerna (2010) concluded that the measure of Oral Reading Fluency proved to be widely regarded as useful as a screening tool for reading success. Schilling et al. (2007) studied a group of first through third graders that were 60% African American and 81.5% economically disadvantaged who were administered DIBELS assessments and a Iowa Test of Basic Skills. The correlation between the ORF score and the total reading achievement score on the Iowa Basic was .74. While a breakdown by subgroup is not provided, some merit can be given to the fact that the sample itself represents a subgroup with relation to the general population. However, it would have been enlightening to study the predictive validity as it pertained to each racial and socioeconomic group.

A large number of additional studies featured the Oral Reading Fluency (ORF) measure as the most valid indicator of student performance on future comprehension assessments (Goffreda & DiPerna, 2010, Goffreda, DiPerna, & Pedersen, 2009; Johnson,

Jenkins, Petscher, & Catts, 2009; Riedel, 2007; Speece & Ritchey, 2005; Good & Kaminski, 2002; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Good, Simmons, & Kame'enui, 2001). Johnson et al. (2009) singled ORF out as having the highest classification accuracy of all the DIBELS measures when studying a large sample (more than 12,000 first grade students) in Florida. However, among students in the top 10% of those at risk, false positives were plentiful, meaning that students near the cut-score are often identified as being at risk and (ultimately appropriate for placement in an intervention group) when they are not. Johnson et al. point to two studies – Hixon & McGlinchey (2004) and Wiley & Deno (2005) – that found “performance differences associated with language and poverty status on (DIBELS) ORF and state standards reading tests” (p. 176). The study suggested the screening accuracy of DIBELS could be improved by including information specific to subgroups to allow for adjustments in screening cut scores.

Riedel (2007) cites two studies pertaining directly to first grade students. Cook's (2003) study reported that DIBELS ORF scores were found to have a correlation of .73 with the Stanford Achievement Test Reading Comprehension Cluster, and Roberts, Good, & Corcoran (2005) found a correlation between the VIP ORF (an alternative form of the DIBELS ORF) with the Woodcock Johnson Broad Reading Cluster of  $r=.76$ . Yet, the study was conducted with a small, homogenous sample with no socioeconomic range and zero minority students (p. 550).

While Riedel points out that the Roberts et al. (2005) study sampled an urban school with more diversity, the study focused on concurrent relationships, not predictive. His study addresses these deficiencies by including a large sample (>1,000) of first grade students in the Memphis area who were predominately African American (92%) with low

socioeconomic status (85% qualifying for free and reduced lunch). The subgroup of ELL students was excluded, as the number within the sample was small, and Riedel considered their challenges to be different due to language acquisition. Using the DIBELS Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, Oral Reading Fluency, and Retell Fluency subtests in correlation with the Terra Nova Reading subtest and Group Reading Assessment and Diagnostic Evaluation vocabulary and comprehension subtests, Riedel reported that ORF was the, “single best predictor at the end of first grade ... classifying 80% of the students correctly” (p. 555). While this appears to be true, there is no information pertaining to the performance of the subgroups in relation to their non-minority peers.

#### **ORF Findings Specific to Race and Ethnicity**

In the majority of studies reviewed, little to no information is provided about student characteristics with relation to race or ethnicity when the validity of DIBELS is discussed. Specifically, there is an absence of data related to subpopulations. On the DIBELS home website, an extensive list of references is provided with studies related to the reliability and validity of DIBELS as indicators and predictors of future performance, yet only one study has a specific focus on students from a low socio-economic subgroup. This particular study conducted by Baker, Smolkowski, Katz, Fien, Seeley, Kame’enui, and Beck (2008) had an average of 69% of the student population of 34 Oregon Reading First Schools qualifying for Free & Reduced Lunch; the DIBELS ORF administered in grade 3 had a strong correlation with the Oregon Statewide Reading Assessment ( $r = .67$ ). This suggests that ORF had significant predictive validity and could be appropriately used for this population, which had a significant amount of students living in poverty.

However, there is still no comparison between the peers within the group that had poverty status/non-poverty status, nor is there disaggregation for ethnicity or even mention of racial subgroups.

One additional study (Gunn, Biglan, Smolkowski, & Ary, 2000), also conducted in Oregon, discusses the efficacy of decoding instruction for a Hispanic subpopulation, but the study is focused on instructional practice rather than the validity of the measures with respect to the ethnic subgroup. The vast majority of studies included on the home website define subgroups based on grade level and reading skill measured over time (struggling readers versus skilled readers). Several studies refer to “diversity” as the span of ability in readers, but do not focus on subpopulations related to ethnicity or race, creating a need for further studies in this area. Given the fact that many Reading First Schools in Oregon have diversity within their subpopulations, it is intriguing to search for how their results may vary from predominate groups.

Only a few studies could be located that disaggregated students with minority ethnic status to show how they were similar to and different from their Caucasian counterparts. Hintze, Callahan, Matthews, and Williams (2002) studied the predictive bias of CBM when African American and Caucasian students, grades 2-5, were compared (N=136). The Woodcock Johnson Reading Comprehension subtest was used to derive comprehension scores. As a basis for their research, they point to a study (Kranzler, Miller, and Jordan, 1999) in which CBM did not prove to act as an unbiased indicator when studying African American students in contrast with their White peers. In said study, the comprehension abilities of African American students were overestimated, thereby potentially disadvantaging them for qualification into intervention programs.

More specifically, they suggest that predictive bias resulted in under-identification of African American students using CBM reading comprehension measures. Limitations of this study suggest that the results are not generalizable due to the use of different reading passages when measuring achievement. This leads to Hintze et al. paying particular attention to this factor in their study. Hintze et al. (2002) go to great lengths to detail scoring procedures and passage level to suggest greater validity. They also account for the fact that no significant relationship existed between the studied factors and age, thereby discounting it as a nuisance. They also suggest that neither SES nor ethnicity added significantly to the prediction of reading comprehension. To their point, only age and CBM oral reading fluency scores acted as adequate predictors. Hintze et al. (2002) also conducted separate z-tests for the two ethnic groups, showing that results for both groups do not differ significantly from the overall group prediction.

In contrast, Pearce and Gayle (2009) show that within their regression analysis, socioeconomic status and ethnicity (specifically, Native American ethnicity) are significant contributors. They studied the predictive bias of DIBELS when 543 Native American and Caucasian third grade students were compared. The Dakota State Test of Educational Proficiency was used as a measure of comprehension. Pearce and Gayle (2009) present their results that demonstrate that DORF accounted for about 40% of the variance in comprehension scores, with SES status adding an additional 2% and ethnicity adds an additional 3% on top of that. In conclusion, the American Indian cohort scored approximately one standard deviation below the Caucasian subgroup. They also measure an interaction effect between DORF and ethnicity, which was not statistically significant. Pearce and Gayle (2009) point out that the proportion of false negatives within the

American Indian subgroup were significantly larger than the Caucasian subgroup ( $p < .01$ ), meaning that many students from the minority subpopulation did not receive needed interventions based on ORF scores.

Valencia, Smith, Reece, Li, Wixson, and Newman (2010) conducted a study that questioned the validity of using the ORF measure to identify of 279 students at risk in grades 2, 4, & 6. This study is particularly relevant due to the population studied, which had 55% of the students in ethnic minority subgroups and 43% qualifying for free and reduced lunch. In the study, words correct per minute were examined as part of DIBELS “risk” categories in relation to ITBS comprehension scores. Discrepancies were found that demonstrated false positives and false negatives in the placement of students for intervention, with false positives ranging from 12% to 24% of students identified as at risk with DIBELS. While the student characteristics lend themselves to the consideration that minority status could be a contributor, the data was not disaggregated to show the results in relation to different subgroups.

Throughout the research examined, it is clear that we can find a great deal of evidence pertaining to the predictive validity of DIBELS Oral Reading Fluency measures for predicting comprehension scores. However, it is clear that the majority of studies fail to disaggregate the data in a way that would allow us to examine ethnic and racial minority groups. In addition, with such significant amounts of false positives and negatives, it would be powerful to examine the data through this particular lens: how would students of ethnic minorities compare to their Caucasian peers? This leads to the following research questions:

1. What is the predictive validity of ORF for two comprehension measures?

2. To what extent are race and ethnicity correlated with ORF and comprehension performance?
3. To what extent do race and ethnicity explain comprehension performance after ORF is controlled?
4. To what extent does the relationship between ORF and comprehension depend on race and ethnicity?

The present study represents an attempt to begin to answer these questions.



## CHAPTER III

### METHODS

#### **Setting**

This study was conducted in a large, suburban school district that serves several cities and some unincorporated areas. They currently have over 20,000 students enrolled in grades K-12 in both traditional and alternative settings. While their largest ethnic group is Caucasian (53%), 47% are identified as ethnic minorities, the largest group being Hispanic (33%). In the 2010-2011 school year, one high school was identified as needing improvement by the Oregon Department of Education. Six elementary schools and one charter school were rated Outstanding, and all other schools (secondary included) were rated Satisfactory.

In accordance with the Federal No Child Left Behind Act, information is made available regarding subpopulations and their performance on the statewide assessment. For the 2010-2011 school year, the following race/ethnicity subgroups were recorded: American Indian/Alaskan Native, Asian/Pacific Islander, Black, Hispanic, White, and Multi-Racial/Multi-ethnic. In addition to race/ethnicity, the data was disaggregated for the following subgroups: students with disabilities, Limited English Proficient (LEP) students, and economically disadvantaged students. In the area of Language Arts, all student groups made adequate yearly progress (AYP) with the exception of students with disabilities and students who are categorized as LEP. However, the participation rate was not met for the subgroups of Black and students with disabilities, which limit the interpretability of that data. In considering the actual percentages of students not meeting benchmark in the area of Language Arts, it is notable that 30.6% of American

Indian/Alaskan Native students did not meet, 25.8% of Black students did not meet, and 32.6% of Hispanic students did not meet benchmark. Contrast that with the subgroups of White and Asian students, in which 12.3% and 12.8% did not meet in the same curricular area. Also in the area of Language Arts, 28.9% of economically disadvantaged students failed to meet benchmarks, while for all students 19.6% did not meet the standard. The participation rates for Black students and students with disabilities were not met, and the graduation rate for LEP and economically disadvantaged students was not in line with AYP standards.

### **Participants**

The original data gathered included first and second graders in the aforementioned district. For purposes of this study, the data includes only second graders, as they participated in all three assessments of interest: a) DIBELS Oral Reading Fluency in Fall, b) easyCBM Multiple Choice Reading Comprehension in Spring, and c) SAT 10 Reading Comprehension in Spring. Federal guidelines require the reporting of ethnicity and race. In terms of ethnicity, about one third of the students in this study identified themselves as Hispanic (Table 1). Eight different categories of race were reported, including one denoted as “unknown” which represents 4.7% of the sample. Students identifying as White make up 52.2% of the total sample, and students representing races other than White (including multiple) make up 40% of the total sample.

Table 1  
*Frequencies - Ethnicity and Race Reporting Categories (N = 1591)*

Demographic Characteristic	Category	Frequency	Percent
Ethnicity	Non-Hispanic	962	60.5
	Hispanic	506	31.8
	Missing	123	7.7
Race	American	321	20.2
	Indian/Alaskan Native	122	7.7
	Asian	31	1.9
	Black	11	0.7
	Hawaiian/Pacific Islander	830	52.2
	White	78	4.9
	Multiple	74	4.7
	Unknown	124	7.8
	Missing		

**Federal regulations on reporting ethnicity and race.** To further understand the sample we are studying, it is important to note the method by which the federal government requires families and schools to report ethnicity and race. In 2010, the year that this data was collected, the procedure for reporting ethnicity and race in Oregon was detailed in a manual written by the Education Enterprise Steering Committee, a group comprised of representatives from the governor’s office, the Oregon Department of Education, the Oregon University System, Oregon Association of Education Service Districts, and members of the K-12 school community. The manual was meant to assist Oregon school districts in implementing the federal guidelines for collecting and

reporting data regarding ethnicity and race, and it acknowledges upfront that statements in the manual may be perceived as controversial in nature (p. 2).

The new guidelines stem from changes made by the U.S. Office of Management and Budget in 1997. At that time, the federal government asked families to self identify their ethnicity and race with the option to choose several categories. In an effort to report more accurate data and improve the equitable distribution of resources, changes were implemented in 2007, making the procedure for self-reporting a two part question. Families were first to identify their ethnicity as Hispanic or Non-Hispanic. After selecting an ethnicity, the second question asked families to select one or more races to describe themselves: American Indian/Alaska Native, Asian, Black or African American, Native Hawaiian/Pacific Islander, or White. No option for Latino was listed in the race categories, thereby creating a difficulty for many families that did not identify with any of the listed racial groups. Families have been know to check any one of the races that they felt was closest to their racial identity while some have checked all and still others left the race question unmarked.

To further complicate the matter, the new federal guidelines included an “observer identification requirement.” This is explained as:

When self-identification is not possible or is refused ‘observer identification’ should be used...the USED recognizes the burden placed on school and district personnel in observer identification and that the practice may not yield data as accurate as those from self identification. However absent self-identification or existing records, observer identification is considered preferable by USED to having no data at all. (Education Enterprise Steering Committee, 2010, p. 8)

So, based on best approximation, schools must mark any questions left blank. Given the current sample, it is clear that several of the schools failed to do this, resulting in a category labeled “unknown”

In an effort to obtain the most accurate portrait of ethnicity and race within the sample, the data was recoded in the following way: a) Hispanic, meaning the initial choice of ethnicity was recorded in this manner, b) Non-Hispanic White, meaning Non-Hispanic ethnicity and White race were selected, or c) Non-Hispanic Minority, which is inclusive of all who chose the ethnicity Non-Hispanic plus any racial category other than White (inclusive of American Indian/Alaska Native, Asian, Black or African American, Native Hawaiian/Pacific Islander, and the category denoted as multiple). Those who marked their ethnicity as Hispanic but had race marked as “unknown” were counted as Hispanic. Missing data and those noted as Non-Hispanic Unknown were excluded, bringing the sample from 1591 students to 1461 students. The results of these groupings are detailed in Table 2.

Table 2  
*Frequencies for Ethnicity and Race Regrouped*

	Frequency	Percent
Hispanic	506	34.6
Non Hispanic, White	732	50.0
Non Hispanic, Minority	226	15.4

To derive the final sample, it was important to narrow this sample to the group of students participating in both DIBELS ORF in the Fall and one or both of the comprehension measures in the Spring. While the vast majority of students were assessed using DIBELS, smaller samples participated in the easyCBM MCRC measure and SAT 10 Reading Comprehension measure. Table 3 below details the final sample

that was studied; the students in this sample participated in both comprehension measures. It is worthwhile to note that the resulting sample has lower percentages of students in ethnic and racial minority groups and higher percentages of students classified Non-Hispanic White. In addition, of the 222 students participating in easyCBM and the 201 students participating in SAT 10, 198 of the students are the same, which constitutes the majority of the reduced sample.

Table 3  
*Frequencies for Ethnicity and Race Subpopulations Participating in Comprehension Measures*

	easyCBM (N = 222)		SAT 10 (N = 201)	
	Frequency	Percent	Frequency	Percent
Hispanic	58	26.1	45	22.4
Non-Hispanic White	131	59.0	127	63.2
Non-Hispanic Minority	33	14.9	29	14.4

A sample of 222 second graders who were assessed with DIBELS ORF in fall also participated in the easyCBM MCRC measure in the spring and 201 second graders participated in the SAT 10 Reading Comprehension measure in the spring. The reasons for the decrease in sample size can be explained by the design of the original data collection method. Given the vast number of assessments delivered, decisions were made to limit the total number of assessments given in every school. Administering each assessment results in additional time lost in instruction, and it was important to not overtax or fatigue students by giving multiple batteries of assessments. Again, 198 of these students participated in both comprehension measures.

### Measures

**DIBELS ORF.** The measurement system commonly employed throughout recent studies is the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). DIBELS was

developed at the University of Oregon supported by the Early Childhood Research Institute on Measuring Growth and Development through the United States Department of Education. This set of early literacy assessments is the work of Dynamic Measurement Group, Inc. which works in collaboration with the Institute for the Development of Educational Achievement at the University of Oregon. This is an endeavor with the assistance of many literacy experts, but the main authors are listed as Ruth A. Kaminski, Roland H. Good III, Deborah Laimon, Sylvia Smith, and Sheila Dill. DIBELS were developed largely on the measurement system employed by Curriculum Based Measurement (CBM) at the University of Minnesota in the 1970s.

DIBELS are appropriately named, as they are intended to be predictive indicators of children's reading performance as they progress through the school system. Riedel (2007) described DIBELS oral reading fluency measures administered at the first grade level to be significantly correlated with comprehension scores, with 80% of the students being classified at the appropriate risk level. The assessments are designed to be short, specific, and targeted to provide progress monitoring over time, and the reporting guide includes a method by which to track and graph student performance. Using this information, educators can identify students with reading deficiencies early in their school career so that interventions can be as proactive as possible. They are created on the basis of the five essential components of balanced literacy: Phonemic Awareness, Alphabetic Principle/Phonics, Oral Reading Fluency, Vocabulary, and Comprehension.

Typically, all students are measured on developmentally appropriate measures at three points during the year, coinciding with three academic terms: Fall, Winter, and Spring. Based on the assessment scores, students may be labeled as "at risk", "some risk",

and “low risk.” The term “at risk” refers to students whose scores fall in the lowest 20% of students within the district. The term “some risk” refers to students performing between the 20<sup>th</sup> and 40<sup>th</sup> percentile, and “low risk” refers to performance above the 40<sup>th</sup> percentile. As mentioned earlier, not all measures are administered at each grade level. Assessments are given based on grade level and prior performance of the student.

The assessments that comprise DIBELS are separated into six specific measures to be utilized as developmentally appropriate from Kindergarten through sixth grade. The measures administered in Kindergarten include Letter Naming Fluency, Initial Sound Fluency, Phoneme Segmentation, and Nonsense Word Fluency. Beginning in the first grade, students are given the assessments for Phoneme Segmentation, Letter Naming Fluency, Nonsense Word Fluency, Oral Reading Fluency, and Retell Fluency. No additional measures within DIBELS are given beyond this point, but assessments are administered as needed depending on the growth and ability of the student. Test-retest reliabilities have been found ranging from .92-.97 over a series of studies, and alternate-form reliabilities of different reading passages in this grade level ranged from .89 to .94. Criterion-related validity studies suggest coefficients ranging from .52-.91 (DIBELS Administration Manual, 2002).

The DIBELS Oral Reading Fluency measure (DORF) is a test of accuracy and fluency given a set of progressively difficult, leveled reading passages that correspond to grade level. In contrast to the other measures presented, the DORF passages are complete and coherent in the sense that they mimic realistic stories that students would encounter in a normal setting. The purpose of this measure is to identify students needing additional support as it relates to fluency and to monitor continuous progress over the course of the



year. Speece and Ritchey (2005) ascertained that, “reading fluency is considered critical to skilled reading, given (a) its correlational if not causal connection to comprehension (Bourassa, Levy, Dowin, & Casey, 1998; L.S. Fuchs, Fuchs, Hosp, & Jenkins, 2001; National Reading Panel, 2000), and (b) evidence that at-risk and typically developing children as early as first grade demonstrate large differences in reading fluency skills (Biemiller, 1977-1978; Deno, Fuchs, Marston, & Shin, 2001).

**EasyCBM multiple choice reading comprehension.** EasyCBM (Curriculum Based Measures) was developed at the University of Oregon as a formative assessment tool that would allow teachers to make instructional decisions in a Response to Intervention model. The assessments exist in the areas of reading comprehension, passage reading fluency, word reading fluency, letter names, letter sounds, phoneme segmenting, and mathematics. Each of the measures samples grade level curriculum to assess levels of mastery of developmentally appropriate skills.

The Multiple Choice Reading Comprehension Measure is administered at the second grade level. Based on teacher requests for authentic and identifiable texts embedded in assessments, developers initially wrote stories that closely aligned with common text passages seen in the classroom. Then to minimize cost associated with scoring, multiple choice test items were constructed to evaluate ability in literal, inferential, and evaluative comprehension.

**Stanford Achievement Test 10.** The SAT 10 Achievement Test is a series of multiple-choice formative assessments in reading, language, spelling, listening, science, social science, and math that are designed to identify students for intervention placements and to allow teachers to deliver appropriate instruction based on test results. Tests are

untimed and accommodations are available; typical administration is on the computer, but Braille, large print, and pencil-paper forms are available for students needing such accommodations. Tests are meant to encompass skills taught later in the year, and are most often given at the end of the school year. For our purposes, we will focus on the reading assessments.

The complete battery of reading subtests include assessment of sounds and letters (Kindergarten only), word study skills (beginning in grade one), word reading and sentence reading (Kindergarten through grade two), reading vocabulary (beginning in the second half of second grade), and reading comprehension (beginning in the second half of first grade). There is also a total reading score that is calculated based on the total of the subtests. The Reading Comprehension subtest incorporates multiple comprehension skills and strategies, including initial understanding, interpretation, critical analysis, and awareness and usage of reading strategies. Reading selections are intended to be multicultural and diverse in nature. Much like easyCBM MCRC, a cold read of a narrative passage is followed by a series of multiple-choice questions that measure comprehension.

## **Procedures**

**Oral reading fluency administration.** In the administration of this assessment, students read a leveled, benchmarked passage to the examiner for one minute; omissions, substitutions, and pauses lasting more than three seconds result in errors. The number of correct words per minute results in the score. The benchmark goal has been 40 words per minute in the spring of first grade and 90 in the spring of second grade with intervention

suggested for students scoring below 10 words per minute in the spring of first grade and below 50 in the spring of second grade.

**EasyCBM multiple choice reading comprehension measures administration.**

The administration of the MCRC measure occurs in a group setting and, most often, on the computer. Pencil and paper forms are available. Students on IEPs may be given appropriate accommodations. After a cold read of a text passage, students answer multiple-choice questions based on their reading (12 questions total at grade two) to assess their ability in literal and inferential comprehension. Three possible answers are offered, one being most correct. The MCRC measure increases in complexity at grade three when additional multiple-choice questions are added along with the area of evaluative comprehension being assessed.

The MCRC measure is intended for students who have the ability to read independently, hence its introduction at grade two. Students should complete the test independently with no accommodations unless mandated by an existing IEP. Supports given during the testing administration may invalidate scores and data.

**Stanford achievement test 10 administration.** Much like easyCBM MCRC, a cold read of a narrative passage is followed by a series of multiple-choice questions that measure comprehension. The passages are designed to encompass literary, informational, and functional selections, providing exposure to a variety of texts encountered in the school setting and in every day life. Tests are administered in a group setting and are intended to be completed independently without accommodations. The test is untimed. Score reports are automatically generated and delineate skills within reading

comprehension, giving performance levels in each area of below, average, high average, and above to highlight strengths and weaknesses on an individual basis.

**Data analysis.** This is a correlational study of an intact cohort of students over a one-year span. The descriptive statistics will report the means and standard deviations for the general population sample, as well as for each racial and ethnic subgroup. Next, I will report correlations between ORF scores in Fall with the two reading comprehension measures taken in Spring. Finally, I will run two sequential multiple regression analyses in which I regress each comprehension measure on ORF, followed by race and ethnicity, and then by interactions between each grouping variable and ORF.

**Threats to validity.** For internal threats to validity, it begins with selection given the minority subgroup. The question will center on how these differences in respondents are relevant. Then, history and maturation may be factors, as students would be observed over the course of a school year. Attrition will be observed, as some mortality with mobility within school populations is inevitable. The final threat to internal validity is testing & instrumentation. How tests were administered, the testing environment, the presence of multiple students in the room, if directions were followed to the letter, was there prompting by assessors – all are considerations that can invalidate the assessment.

For external threats to validity, generalizability is one consideration, given the minority subgroup and delivery of the instruction. Also, the details of student demographics, including ESL, TAG, and SPED status, may or may not represent the general population. In addition to this, of the psychological effect of being a minority in a school setting and how that impacts performance in general could be a factor in performance. It is not noted which students are receiving intervention, which can also

have an impact in terms of student efficacy. As discussed in the introduction, students that are currently on a trajectory for particular groupings often remain on those tracks, potentially impacting the perspective of the student and/or the assessor and ultimately affecting performance. There is no means of measuring these biases for this sample.

For construct validity, it is important to consider the sampling plan in terms of who was included in the sample and who was omitted, and were outliers observable or removed from the data? Scoring on DIBELS assessments are hand-scored and therefore prone to error, so that should be considered as well. The easyCBM and SAT 10 scoring are scored electronically and are less subject to this particular threat.

## CHAPTER IV

### RESULTS

The predictive validity of Oral Reading Fluency Scores for two comprehension measures with additional factors of ethnicity and race are examined in this chapter. Descriptive statistics, including the mean and standard deviations for the full sample and by race and ethnicity, are reported for ORF and easyCBM in Table 4 and for ORF and SAT 10 in Table 5. Next, bivariate correlations are examined to describe relationships among ORF, both comprehension measures, ethnicity and race. Finally, two sequential linear regression analyses are used to describe the predictive relations of ORF, race, and ethnicity to each comprehension measure and the extent to which the ORF relation is moderated by race and ethnicity.

#### **Descriptive Statistics**

Table 4 describes the mean for students participating in the easy CBM MCRC measure. The benchmark for DIBELS ORF at second grade falls at 40 words per minute (wpm). In narrowing the focus to particular subpopulations, it is notable that Hispanic students averaged 47.07 wpm compared with their Non-Hispanic White peers who averaged 65.22 wpm. The Non-Hispanic Minority group outperforms both groups, although to a lesser extent when compared with their Non-Hispanic White peers. It is important to note that the Non-Hispanic Minority group (inclusive of Native American, Asian, Black, and Multiple races) participating in both comprehension measures includes a small number of students ( $n = 33$  and  $n = 29$  respectively), and within that group, racial groups are present that historically perform at a higher level than their peers, potentially skewing the mean scores.

While the mean falls above benchmark for the Fall of second grade for all groups, the averages differ by subpopulations considerably. Similarly, the mean scores on the Spring easyCBM MCRC measures show that while the subpopulations may have been performing, on average, above benchmark on the ORF, the mean comprehension scores differ considerably by subgroup, with Hispanic students scoring below their peers on average by as much as two points or more.

Table 5 displays the mean ORF scores for students participating in the SAT 10 Reading Comprehension measure. Again, it is notable that the DIBELS ORF mean score for Hispanic students falls far below that of the Non-Hispanic White and Non-Hispanic Minority students, who have mean scores of 65.38 and 69.41 words per minute respectively. SAT 10 mean scores are also lower on average for Hispanic students.

Table 4  
*Descriptive Statistics for Students Assessed with easyCBM (N=222)*

	<i>N</i>	DIBELS		easyCBM	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Hispanic Students	58	47.07	28.06	6.67	3.49
Non-Hispanic White Students	131	65.22	36.61	9.24	2.49
Non-Hispanic Minority Students	33	66.61	39.47	8.61	2.83

Table 5  
*Descriptive Statistics for Students Assessed with SAT 10 (N=201)*

	<i>N</i>	DIBELS		SAT 10	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Hispanic Students	45	50.44	28.45	587.33	44.94
Non-Hispanic White Students	127	65.38	37.33	616.66	43.61
Non-Hispanic Minority Students	29	69.41	39.53	614.34	45.60

For purposes of understanding the distribution of comprehension scores by subgroup, histograms were created for students assessed with DIBELS ORF as well as easyCBM ( $N = 222$ ) and SAT 10 ( $N = 201$ ) and are listed in the Appendix. It is worthwhile to note the differences in the distribution of scores amongst subpopulations. As seen in Table 4, the mean scores on the easyCBM MCRC measure are lower for both the Hispanic subgroup. In studying the histograms, the distributions are similar for all groups on the DIBELS ORF, yet when studying the distributions of the easyCBM scores, White and Non-Hispanic Minority students both show a left skew, with more students performing towards the top of the scale. This is consistent with the reported mean. While histograms for the SAT 10 have more normal distributions (to be expected when using scaled scores), we still see a lower overall mean score for Hispanic students when compared with their Non-Hispanic White and Non-Hispanic Minority peers.



## Correlations

Linear bivariate correlations were computed to investigate if there was a statistically significant association between the variables. Table 6 summarizes these correlations, which were run with two-tailed significance. Of particular interest is the correlation between reading measures and race and ethnicity. Note the significance of  $p < .001$  for both Spring comprehension measures and ethnicity. While the correlations are weak to moderate, they are statistically significant.

Table 6

*Correlations for Spring Comprehension Measures, ORF, Ethnicity, and Race with easyCBM MCRC (N = 222) Above and SAT 10 (N = 201) Below*

	1. Spring comprehension measure	2. DIBELS ORF Words Read Correctly, Fall	3. Ethnicity	4. Race
easyCBM		.514***	-.355***	.305***
2			-.227***	.135*
3				-.572***
SAT 10		.636***	-.265***	.150*
2			-.180**	.074
3				-.530***

\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

Note. *ORF refers to DIBELS Oral Reading Fluency Scores and easyCBM MCRC refers to the easyCBM Multiple Choice Reading Comprehension Measure.*

## Regression for Both Comprehension Measures

Table 7 summarizes the sequential regression conducted with easyCBM MCRC scores as the dependent variable. Given the  $R^2$  below, it is clear that ethnicity and race predict easyCBM MCRC scores above and beyond ORF alone. Note that in Model 1, the DIBELS ORF score alone explains 26.4% of the variance. This is not surprising given

the previous research conducted on the predictive validity of ORF. However, when Ethnicity and Race factors are added in Model 2, an additional 6.6% of the variance is explained ( $p < .0001$ ). When the interactions between ORF and ethnicity and race are added, an additional 1.8% of the variance is explained, but this is not found to be statistically significant,  $p = .052$ .

The slope estimates in Table 8 suggest that although race does not moderate ORF's prediction of easyCBM, ethnicity does. Note that while both race and ethnicity significantly predict comprehension in the second model, only ethnicity is significant in the third model, and its interaction with ORF is significant as well. The parameter estimates from Model 3 indicate that Hispanic students get significantly fewer easyCBM items correct than do their peers ( $B = -3.46$ ) and their ORF scores are significantly more predictive of their easyCBM performance ( $B = .031$ ).

Table 7  
*Summary of Sequential Regression Analysis for Fall ORF, Ethnicity, and Race Predicting easyCBM MCRC Scores (N = 222)*

Model	$R^2$	Change Statistics				
		$\Delta R^2$	$\Delta F$	df1	df2	$p$
1	.264	.264	79.000	1	220	< .001
2	.330	.066	12.291	2	218	< .001
3	.348	.018	3.958	2	216	.052

*Note.* Model 1 predictors include Fall DIBELS Oral Reading Fluency Words Read Correctly. Model 2 predictors: Fall DIBELS Oral Reading Fluency Words Read Correctly, Race, and Ethnicity. Model 3 predictors include Fall DIBELS Oral Reading Fluency Words Read Correctly, Race, Ethnicity, and interaction of ORF with race and ethnicity.

Table 8

*Table of Coefficients for Sequential Regression Analysis for Fall ORF, Ethnicity, and Race Predicting easyCBM Scores (N = 222)*

Model	Unstandardized Coefficients		<i>t</i>	Collinearity Statistics	
	<i>B</i>	<i>SE</i>		Tolerance	VIF
(Constant)	5.838***	.344	16.970		
1 DIBELS ORF	.043***	.005	8.888	1.000	1.000
WRC					
(Constant)	6.712***	.382	17.580		
DIBELS ORF	.039***	.005	8.051	.948	1.054
2 WRC					
Ethnicity	-1.862*	.403	-4.621	.894	1.118
Race	-.684*	.486	-1.409	.938	1.066
(Constant)	7.214***	.442	16.309		
DIBELS ORF	.031***	.006	5.239	.614	1.627
3 WRC					
Ethnicity	-3.462***	.776	-4.461	.237	4.226
Race	-1.577	.961	-1.641	.235	4.252
Ethnicity/ORF	.031*	.013	2.373	.255	3.917
Race/ORF	.014	.013	1.081	.222	4.502

*Note.* WRC = Words Read Correctly

\*  $p < .05$

\*\*\*  $p < .001$

Table 9 provides the model summary for the sequential regression conducted with SAT 10 Reading Comprehension Scaled Scores as the dependent variable. Similar to the easyCBM results, we find that in Model 1, the DIBELS ORF score alone explains 40.5% of the variance. When Ethnicity and Race factors are added in Model 2 an additional 2.5% of the variance is explained ( $p < .05$ ), which is less than these variables predicted easyCBM MCRC. Also somewhat different is that in Model 2 race does not significantly predict SAT 10 scores whereas it did predict easyCBM scores. However, ethnicity does

predict SAT 10 performance, with Hispanic students scoring almost 18 points lower than their peers (see Table 10). When the interaction variables between ORF and the ethnicity and race variables are added, an additional 0.8% of the variance is explained, which is not statistically significant. These findings are further supported by the fact that neither interaction term is statistically significant (see Table 10).

Table 9  
*Summary of Sequential Regression Analysis for Fall ORF, Ethnicity, and Race Predicting SAT 10 Reading Comprehension Scores (N =201)*

Model	R <sup>2</sup>	Change Statistics				
		$\Delta R^2$	$\Delta F$	df1	df2	<i>p</i>
1	.405	.405	135.436	1	199	< .001
2	.430	.025	4.303	2	197	.015
3	.438	.008	1.367	2	195	.257

*Note.* Model 1 predictors include Fall DIBELS Oral Reading Fluency Words Read Correctly. Model 2 predictors: Fall DIBELS Oral Reading Fluency Words Read Correctly, Race, and Ethnicity. Model 3 predictors include Fall DIBELS Oral Reading Fluency Words Read Correctly, Race, Ethnicity, and interaction of ORF with race and ethnicity.

Table 10

*Table of Coefficients for Sequential Regression Analysis for Fall ORF, Ethnicity, and Race Predicting SAT 10 Scores (N = 201)*

Model	Unstandardized Coefficients		t	Collinearity Statistics	
	B	SE		Tolerance	VIF
1 (Constant)	559.737***	4.966	112.704		
DIBELS ORF WRC	.799***	.069	11.638	1.000	1.000
2 (Constant)	566.594***	5.446	104.035		
DIBELS ORF WRC	.766***	.069	11.145	.966	1.034
Ethnicity	-17.892**	6.106	-2.930	.924	1.430
Race	-5.407	7.146	-.757	.950	1.391
3 (Constant)	571.577***	6.216	91.952		
DIBELS ORF WRC	.690***	.083	8.344	.665	1.503
Ethnicity	-30.997*	12.288	-2.523	.227	4.397
Race	-22.519	14.562	-1.546	.228	4.388
Ethnicity/ORF	.237	.201	1.179	.238	4.204
Race/ORF	.251	.185	1.356	.214	4.663

*Note.* WRC = Words Read Correctly

\*  $p < .05$

\*\*  $p < .01$

\*\*\*  $p < .001$

## CHAPTER V

### DISCUSSION

#### **Summary**

Given the wide spread use and reliance on oral reading fluency (ORF) scores to identify students in need of reading intervention, this study sought to determine the predictive validity of those scores for students from various subpopulations. Through examination of the existing research, it was clear that few studies focused on the predictive validity of ORF for students by ethnicity and race. Such an examination is important because it is clear that students in minority subgroups have historically experienced tracking and/or misidentification at a higher rate than their ethnic/racial majority peers (McPherson ,2010, Darling-Hammond, 2010). To avoid placement in remediation when it is not needed and to ensure that we are providing the necessary interventions for those in need, examining ORF, which is among our most common assessments, becomes paramount.

The research questions for this study centered on the predictive validity of oral reading fluency measures as predictors of future reading performance for minority students. In particular, they addressed the following queries:

1. What is the predictive validity of ORF for two comprehension measures?
2. To what extent are race and ethnicity correlated with ORF and comprehension performance?
3. To what extent do race and ethnicity explain comprehension performance after ORF is controlled?

4. To what extent does the relationship between ORF and comprehension depend on race and ethnicity?

Findings for each of these questions are reviewed in the following sections and related to the findings of prior studies.

**Question 1: Predictive validity of ORF.** Consistent with previous research (Goffreda & DiPerna, 2010, Goffreda, DiPerna, & Pedersen, 2009; Johnson, Jenkins, Petscher, & Catts, 2009; Riedel, 2007; Speece & Ritchey, 2005; Good & Kaminski, 2002; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Good, Simmons, & Kame'enui, 2001), DIBELS ORF scores correlated strongly and positively with comprehension measures similar to or identical to those in the current study. Specifically, higher scores on DIBELS ORF were associated with higher scores on both the easyCBM and SAT 10 assessments. This relationship was slightly stronger for SAT 10 as compared to easyCBM in this particular sample; nonetheless, both had strong correlations. Thus, the overall premise of using this assessment as a predictor of future reading performance holds for the current sample.

**Question 2: Relations of race and ethnicity with ORF and comprehension.** Of particular interest is the correlation between the two reading measures and ethnicity (Hispanic vs. not) and race (Non-Hispanic White vs. Non-Hispanic Minority). In examining previous research, findings were unclear as to whether or not racial or ethnic minority status was correlated with student scores (Hintze et al., 2002, Pearce and Gayle, 2009, Valencia et al., 2010) due to the failure to disaggregate the samples by student characteristics. In the current study, ethnic and racial groups were disaggregated: for ORF, these correlations were weak for both of these characteristics and indicated that

Non-Hispanic White students on average read slightly faster than students who were not White and that Hispanic students on average read slightly slower than students who were not Hispanic. Similarly, for the SAT 10, correlations (albeit weak) indicated that White students on average scored slightly higher than students who were not White and Hispanic students scored slightly lower than students who were not Hispanic. For easyCBM MCRC, the general pattern of findings remained the same but were more pronounced; correlations with student characteristics were moderate as well as statistically significant.

Although student characteristics were associated with differential performance on both ORF and comprehension measures, it is not clear from these particular findings whether the predictive relationship between ORF and the comprehension measures differs for students based on ethnicity or race.

**Questions 3 and 4: Extent of dependence on race and ethnicity.** Using multiple regression analysis, the extent to which predictive validity differed by ethnicity or race was examined. A regression analysis of this nature, specific to student characteristics, was not found in previous research during the course of this study. For both comprehension measures, accounting for ethnicity and race was associated with additional variance in comprehension even after ORF was controlled. Thus, the correlations between ethnicity and race and the comprehension measures held despite the strong relationship between ORF and comprehension. It was discovered that an additional 6.6% of the variance in the easyCBM MCRC scores and an additional 2.5% in SAT 10 scores was explained by these student characteristics.



When interactions between ORF and ethnicity and race were added to investigate, it was found that these student characteristics did not significantly moderate the relationship between ORF and easyCBM comprehension, or SAT 10 comprehension. However, an additional 1.8% of the variance in easyCBM MCRC was explained when interactions were added. Although not statistically significant, this additional explained variance is arguably substantial and worrisome. Our hope would be that the relationship between ORF and the comprehension measures would stay constant regardless of student characteristics, but these results suggest otherwise for the easyCBM measure, particularly given the fact that the main and interaction effects for ethnicity were statistically significant while the ones for race were not (see Table 8).

Several reasons might explain why differences exist for the two measures in terms of variance explained by ethnicity and race. One possibility is that the sample size for the two measures is different with fewer students participating in the SAT 10. In addition to this, the frequencies for ethnicity and race differ between the two samples; a somewhat higher percentage of Non-Hispanic White students participated in the SAT 10 (63% Non-Hispanic White) when compared with the group assessed with easyCBM (59% Non-Hispanic White). This is important to consider given that the group with more Hispanic and Non-Hispanic Minority students showed higher levels of variance in comprehension associated with ethnicity and race. Also, while the measures are similar in many ways, DIBELS ORF and easyCBM MCRC are criterion-referenced measures that use predefined standards to determine student scoring while the SAT 10 is a norm-referenced test that compares peer group performance nationally to determine scores. Therefore, the resulting scores do not use different frames of reference for student performance despite

targeting the same underlying construct. As seen in the current data set, students were given the easyCBM MCRC measure at three times throughout the year; the format will be recognizable for students by the time the spring measure is administered. The SAT-10 is summative in nature, and it was administered once at the conclusion of the year. In addition, when comparing easyCBM and SAT10 comprehension measures in terms of format, there are clear differences in the presentation of the text. EasyCBM MCRC presents a single, long passage (approximately 900 words for second grade) followed by 12 questions. It is meant as a screening and progress-monitoring tool in which the questions are designed to assess struggling readers in detail; indeed, a ceiling effect is observable in the student scores within this sample. The constrained range of values for this measure contrast with those for ORF, for example, which has a theoretically unbounded range. SAT 10 presents students with several shorter passages interspersed with comprehension questions. As SAT 10 is comparing and describing a mean for students at a particular grade level, there is more sensitivity around the average, which might account for the larger  $R^2$ . However, on a practical level, the manner in which scores from both measures are used to place students remains the same, making it important to investigate both.

### **Study Limitations**

Although the students in the current sample were ethnically and racially diverse, the sample was limited in a number of ways. First, there is difficulty in creating true, comparable groups based on student characteristics. The federal guidelines defining ethnicity and race illustrate a picture that is not realistic for the sample due to the lack of any Latino racial group; one clear example of this is the disproportionate amount of

Native American students reported in the sample. By attempting to make the groups more clear by defining Hispanic, Non-Hispanic White, and Non-Hispanic Minority, the current study also combined many racial groups together that are quite distinct from one another. Such aggregation of dissimilar group can often distort mean scores.

Unfortunately, the limited numbers of students who belonged to any one racial group that comprised the Non-Hispanic Minority precluded creating multiple racial groups because the sample sizes then become very small.

Another important limitation is that once the groupings had been created and the students who had participated in the measures of interest were extracted, those who participated in the comprehension assessments were somewhat less diverse than the larger sample participating in the ORF assessments. Although diverse, the current sample did not provide large enough samples of multiple ethnic or racial groups to allow for a more nuanced understanding of how race and ethnicity might affect predictive relations between ORF and comprehension. Specifically, due to low sample sizes for disaggregated ethnicity and race groupings, students had to be categorized as Hispanic, Non-Hispanic White or Non-Hispanic Minority, encompassing American Indian/Alaskan Native, Asian, Black, Hawaiian/Pacific Islander, and Multiple Race. Thus, differences between groups that fall under “not Hispanic” were impossible to model in the current sample. Future research should recruit larger samples with better representation of multiple ethnic and racial groups to allow for a more nuanced understanding of how these student characteristics moderate the relationship the relationship between ORF and comprehension measures. It would be vastly informational to have a sample that provided accurate information for each of the racial subgroups indicated in the initial

sample. In addition, it would be incredibly useful, even if not highly likely given the current regulations, to allow people of Hispanic ethnicity additional choices within the category of race that they may find more representational of their racial background. Given the studied sample, it is difficult to capture the full range of what the words "ethnicity" and "race" define.

Clearly, with the focus on Hispanic students in this study, it is also a limitation that Limited English Proficiency status is not included or modeled. English language learners in general have lower English fluency and comprehension scores than their peers who speak English as their first language. Thus, language status could be a reason that Hispanic students shows lower mean scores across the board on each measure. It may be that once language status is controlled, the ethnicity effects may become non-significant. For future studies, the ability to account for LEP status as it correlates to performance on all measures, race, and ethnicity would shed further light on these findings. Clearly, the interaction effect presented in this study did not account for first languages. A sample of increased size that included information about LEP status in conjunction with ethnicity and race would be preferable.

Another limitation of the study is the lack of knowledge about the internal validity of the testing and instrumentation. Given that the current study represented a secondary analysis of extant data, it is difficult to know how the test setting and assessor training varied across schools and might have affected current findings. For example, it is possible that teacher bias affected scores, particularly on the DIBELS ORF, which requires rapid and simultaneous scoring as the assessment is given. Having educators assess students that are not on their caseload could reduce some of the expectation effect

that results from our prior knowledge of a student's ability. Just as work samples are often blind scored, purposefully having teams score students in alternate grade levels or classrooms could assist in reducing that threat to internal validity. The nature of the data analyzed did not permit assurances of unbiased, error-free administration procedures. Future research should take into account inter-rater reliability for measures like ORF.

Finally, no information was available regarding which students received intervention during the year. In an ideal education environment, poor performance on screening measures like ORF would lead to intervention on the part of the school, and presuming the intervention was successful the relationship between ORF and later comprehension measures would be weakened. Without information on student participation in interventions, the current analysis essentially treated all students as receiving the same instruction, which was not likely the case. Future research should take intervention status into account.

### **Implications**

As educators, we have made great strides in allowing data to assist and guide us in making daily and long-term instructional decisions. This is a positive stride in most cases. However, much growth is needed in our abilities to effectively interpret data to discover the information it is actually providing. Most educators do not have the time to investigate the predictive validity to this extent. They are (rightly) focused on the action within their classroom, and they rely on the assessments that are adopted to be accurate and unbiased. When the educators in my building study data together, they not only look at scores, but also at the strengths and weaknesses of students who are meeting or failing to meet benchmark expectations. We work to identify the skills we should reinforce and

those we should correct. This has to be an additional step taken before proceeding to placement. The rich discussions at the individual student level are often the ones that lead us to the most effective interventions. Becoming dependent on data, particularly a singular data point, to make a decision that may affect a student's school career is unhealthy. We must develop the ability to consider the subtle implications and our own bias when we assess students, and we must find ways to become more flexible about groupings and intervention practices so that students are not tracked into a particular route that is predetermined. We look to data to make our decisions less subjective. Therefore, it is important that the assessments we are utilizing apply equally to all students that we serve.

However, the results of this study suggest that while ORF is a strong predictor of comprehension scores, it only accounts for a portion of the variance. In accordance with findings from past studies (Pearce and Gayle, 2009, Kranzler, Miller, and Jordan, 1999, and Valencia, Smith, Reece, Li, Wixon, and Newman, 2010), ethnicity and race were also contributing as factors in the performance of students. In addition, this study showed an interaction effect between DORF and ethnicity and race to be statistically significant for one of the comprehension measures studied. This finding has implications for current practices of identifying and serving students in intervention settings at the elementary level.

One implication is to search for assessments that could give us additional data that could assist us in identifying students for intervention. The obvious drawback to this is the time investment – both to test and to interpret the test results. Nevertheless, such assessment could be conducted judiciously so that both skills and general ability are

assessed, thereby striking more of a balance between general ability grouping and skill-based groupings that could be shorter term. By focusing on skill deficits, we may naturally discover that groups would be become more flexible, and students that we may have anticipated to be proficient could demonstrate a gap in their understanding, while still others could surprise us with their grasp of the particular skill or concept. This avoids the problems that tracking presents. The greatest mistake we can make in practice is to utilize one assessment as the ultimate contributor to our decisions in the classroom.

On the broader level, I agree with Johnson et al. (2009) that including information about specific subgroups by developing local norms could be helpful in making adjustments to cut scores for different subpopulations. Given the somewhat inconsistent findings regarding whether ethnicity or race moderate relations between ORF and comprehension, it seems like gaining a local understanding of the relations would be most important. While we simultaneously strive to produce and test new versions of these assessments that demonstrate less variance with regards to ethnicity and race, we could also use what we know to level the playing field for minorities. The effort to reduce bias within the assessments themselves is, of course, paramount.

### **Additional Directions for Future Research**

The results of this study suggest the need for further research. Studying larger sample sizes in more diverse districts can further inform conclusions drawn from this and all previous research. Including specific information about ethnic and racial subgroups could improve the accuracy of the correlations presented; gathering an increasingly accurate picture of the demographics within our schools is complicated, yet favorable to our understanding of the effectiveness of our assessments. The current study suggests

that at least for some comprehension measures, predictive validity of ORF is moderated by student characteristics like race and ethnicity. The current findings suggest the importance of using multiple comprehension measures when investigating ORF and comprehension. In fact, the inconsistency of prior research regarding whether race and ethnicity moderate the ORF-comprehension relationship may be due, at least in part, to the use of different comprehension measures across studies. Use of multiple comprehension measures, in concert with an understanding of how those measures differ, may contribute to our understanding by making more apparent when and for which types of measures moderation exists.

Of course, the current study only scratches the surface in terms of investigating such differential relationships. Future research should engage larger, more diverse samples to allow for including specific information about ethnic and racial subgroups to improve our understanding of how relationships between ORF and comprehension may vary depending on group membership. It also would be interesting to explore the role of limited English proficiency in addition to ethnicity and race and how they jointly impact assessment results and predictive validity. Including student socioeconomic status, a well-known correlate to many demographic variables, is strongly encouraged as it could alter the extent to which moderation effects are found.



APPENDIX

FIGURES

Figure 1

*DIBELS ORF Scores for Hispanic Students Participating in easyCBM Assessment*

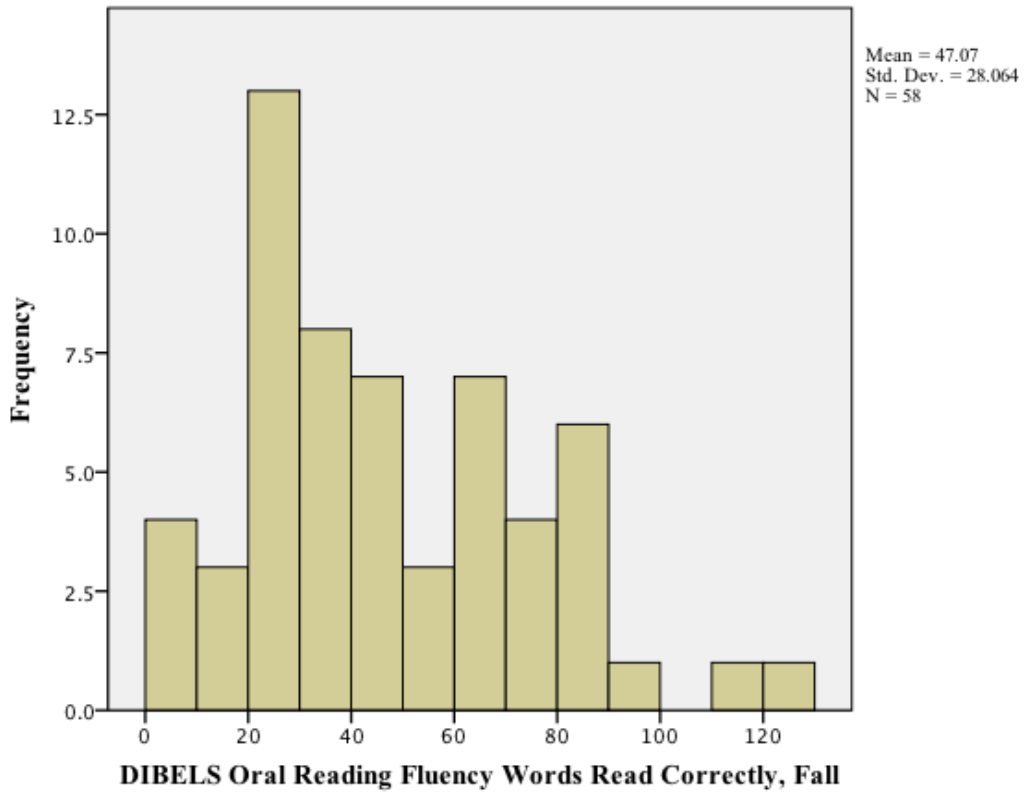


Figure 2  
*Spring easyCBM Multiple Choice Reading Comprehension Scores for Hispanic Students*

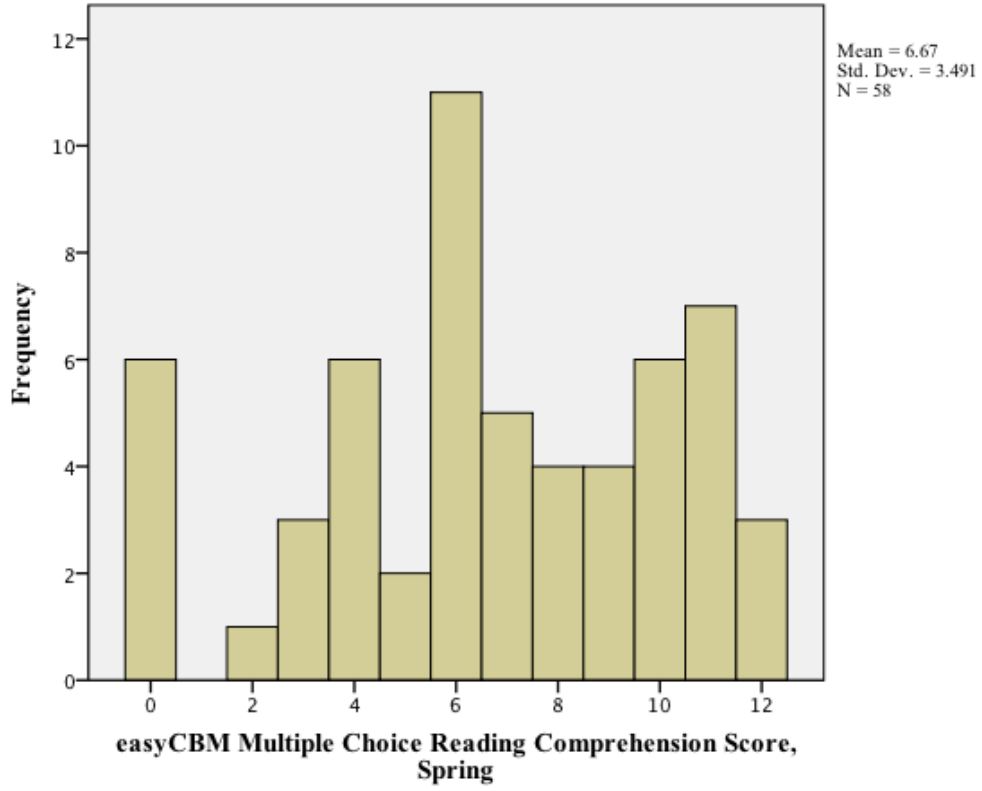


Figure 3  
*DIBELS ORF Scores for Non-Hispanic White Students Participating in easyCBM Assessment*

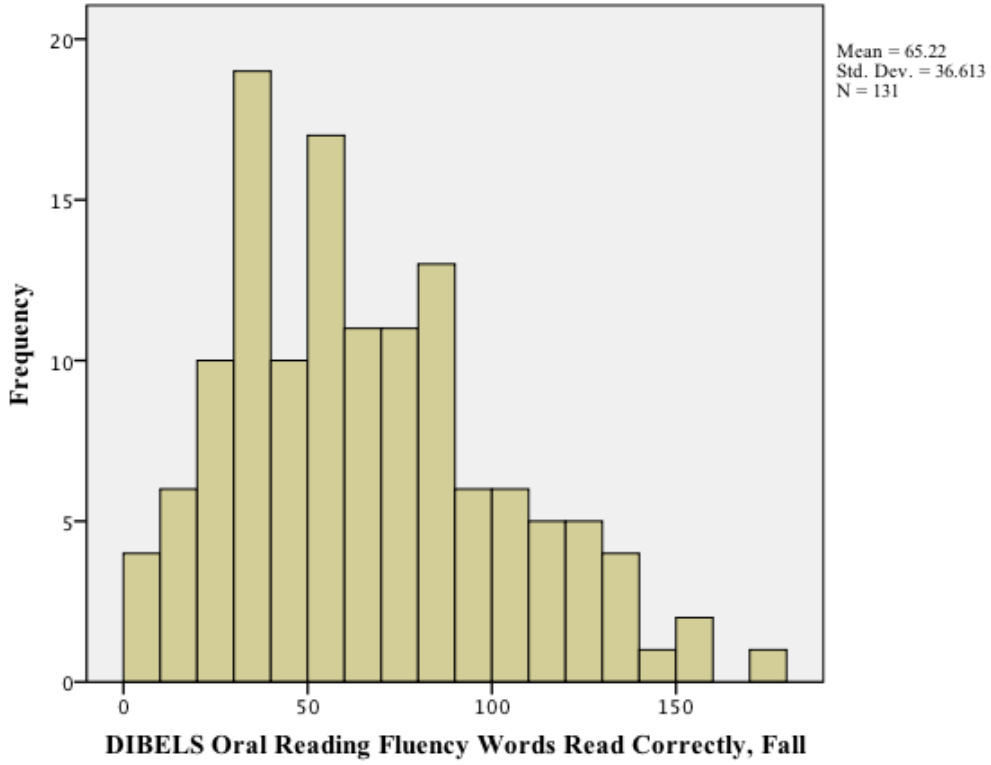


Figure 4  
*Spring easyCBM Multiple Choice Reading Comprehension Score for Non-Hispanic White Students*

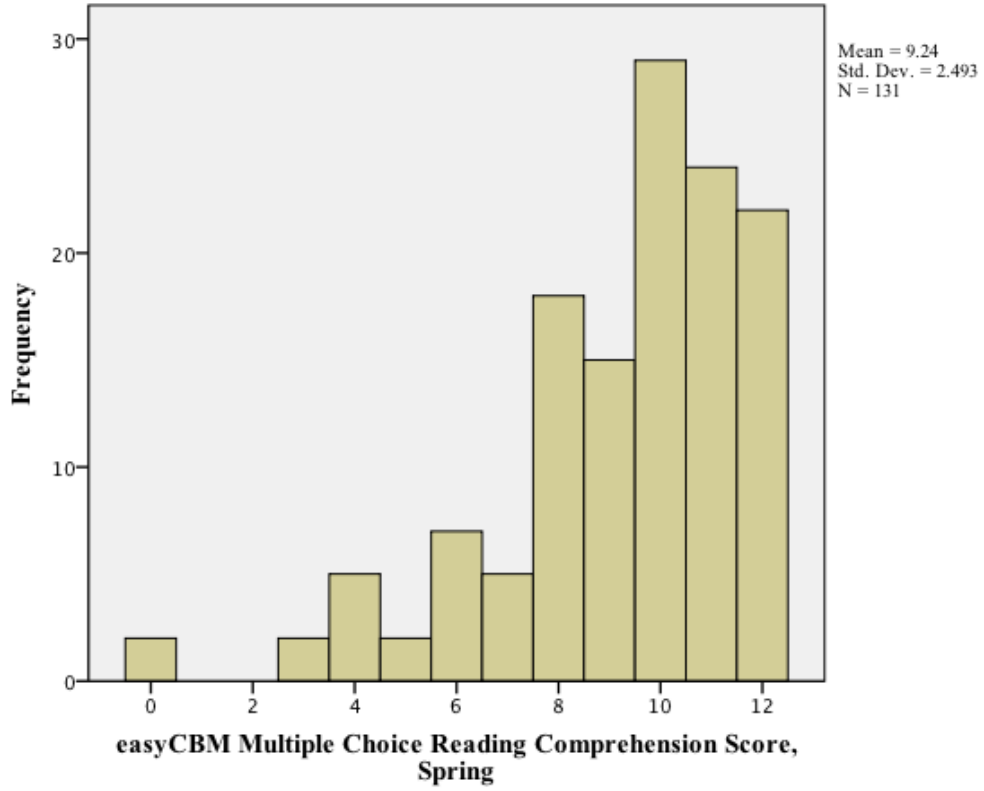


Figure 5  
*DIBELS ORF Scores for Non-Hispanic Minority Students Participating in Spring  
easyCBM Multiple Choice Reading Comprehension*

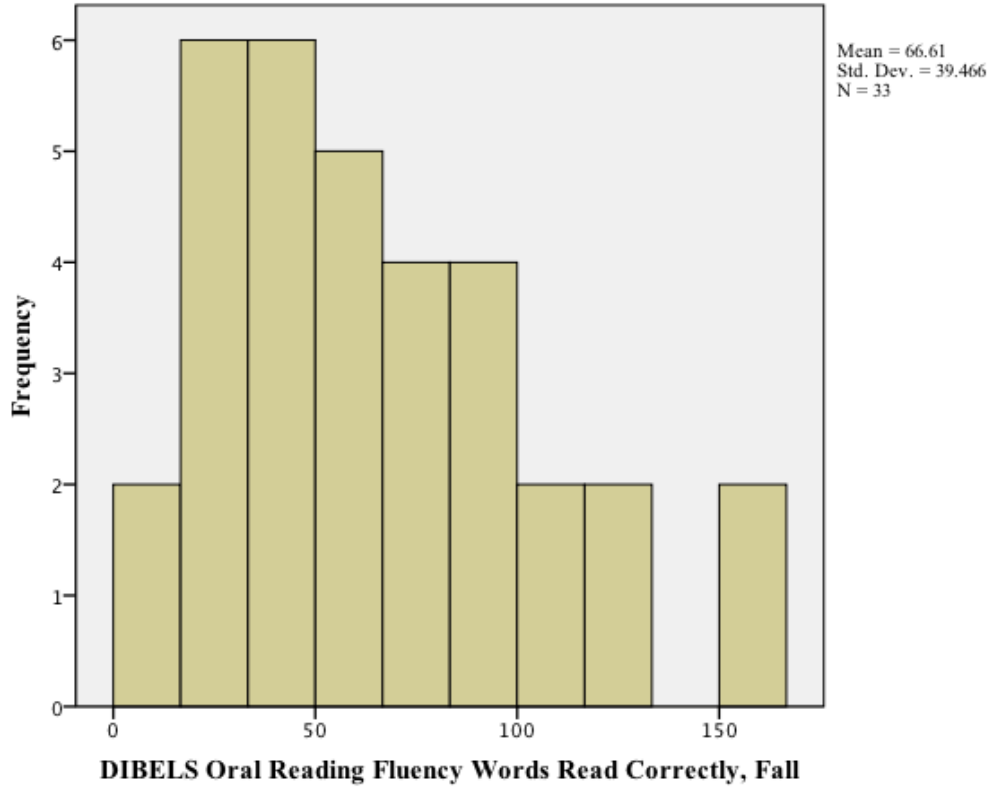


Figure 6  
*Spring easyCBM Multiple Choice Reading Comprehension Scores for Non-Hispanic Minority Students*

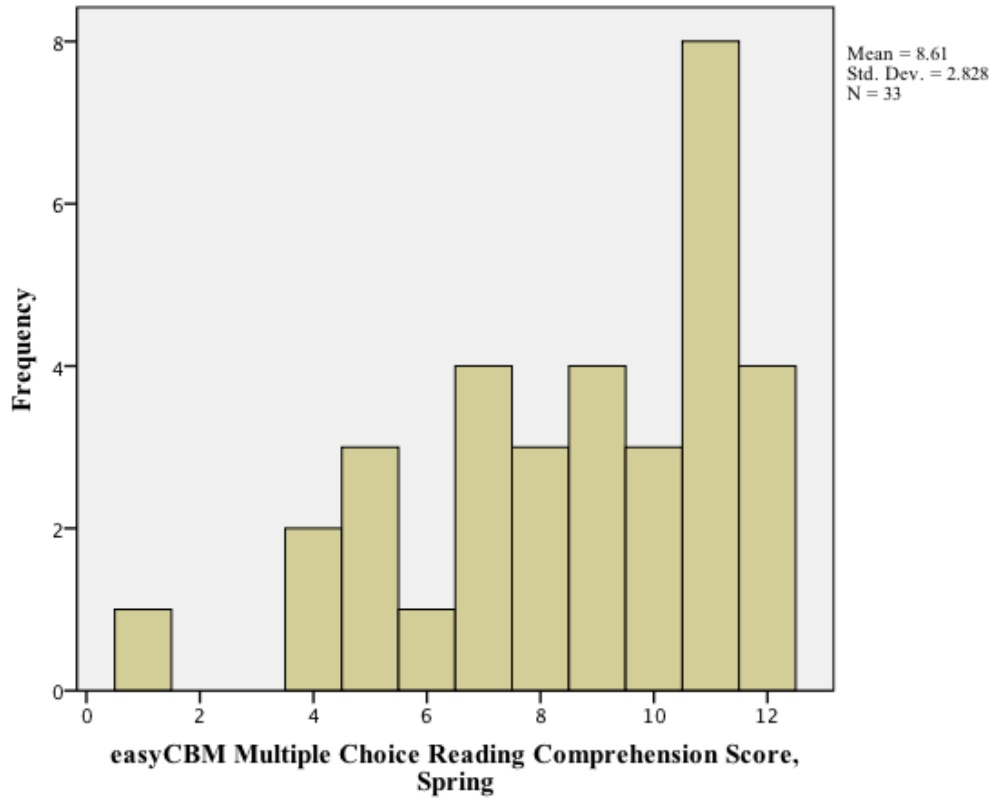


Figure 7  
*DIBELS ORF Scores for Hispanic Students Participating in Spring SAT 10 Reading Comprehension*

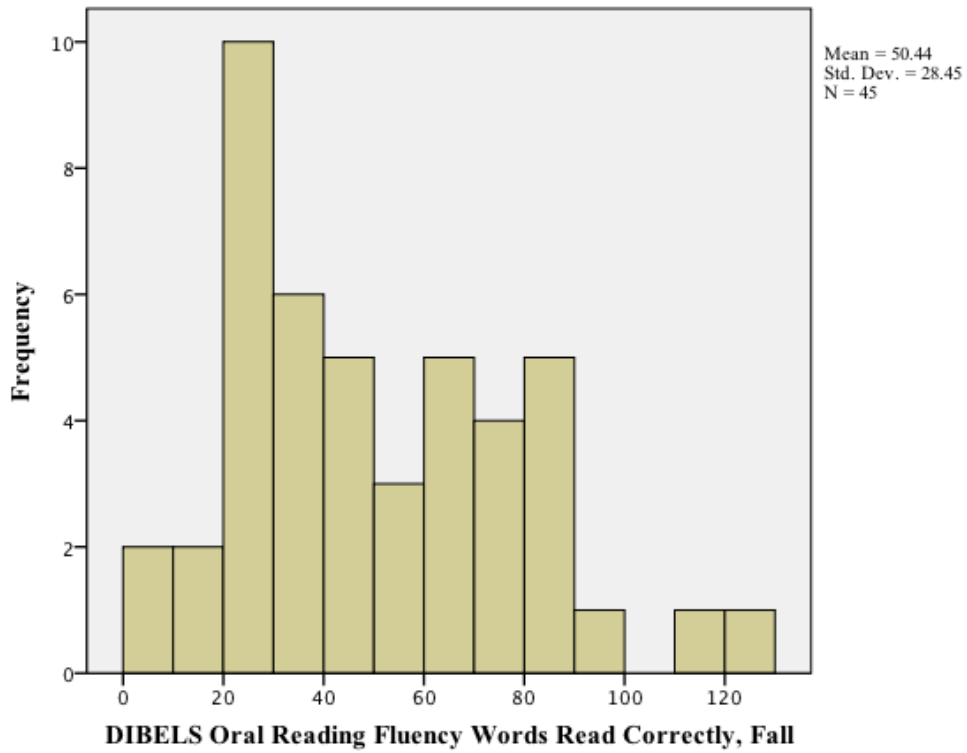


Figure 8  
*Spring SAT 10 Reading Comprehension Score for Hispanic Students*

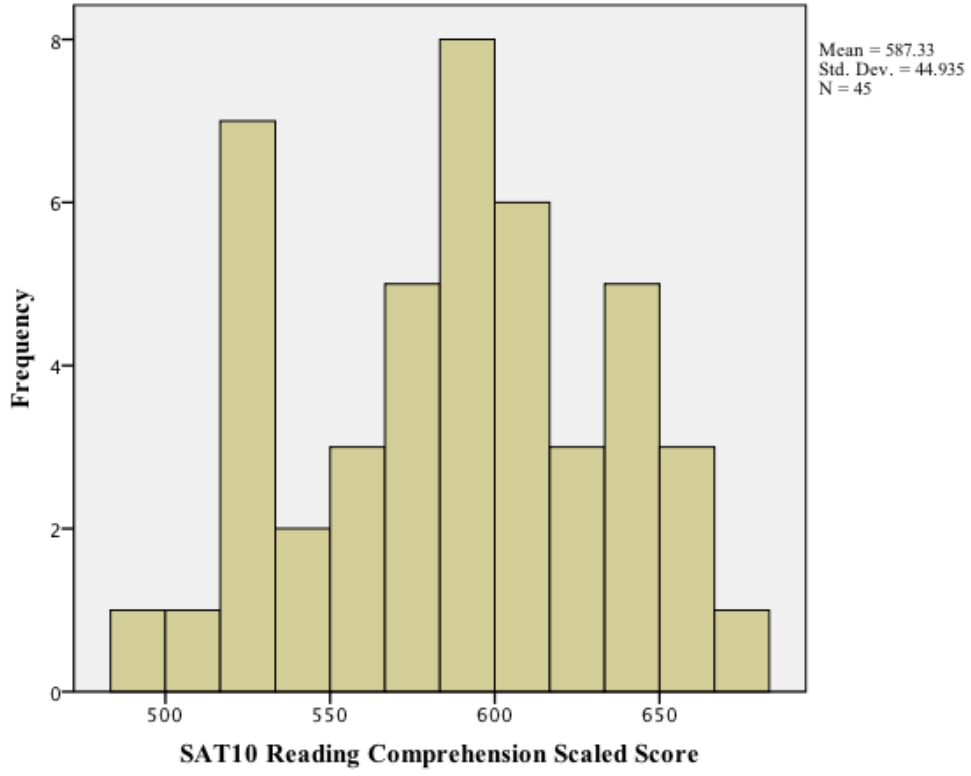




Figure 9  
*DIBELS ORF Scores for Non-Hispanic White Students Participating in Spring SAT 10 Reading Comprehension*

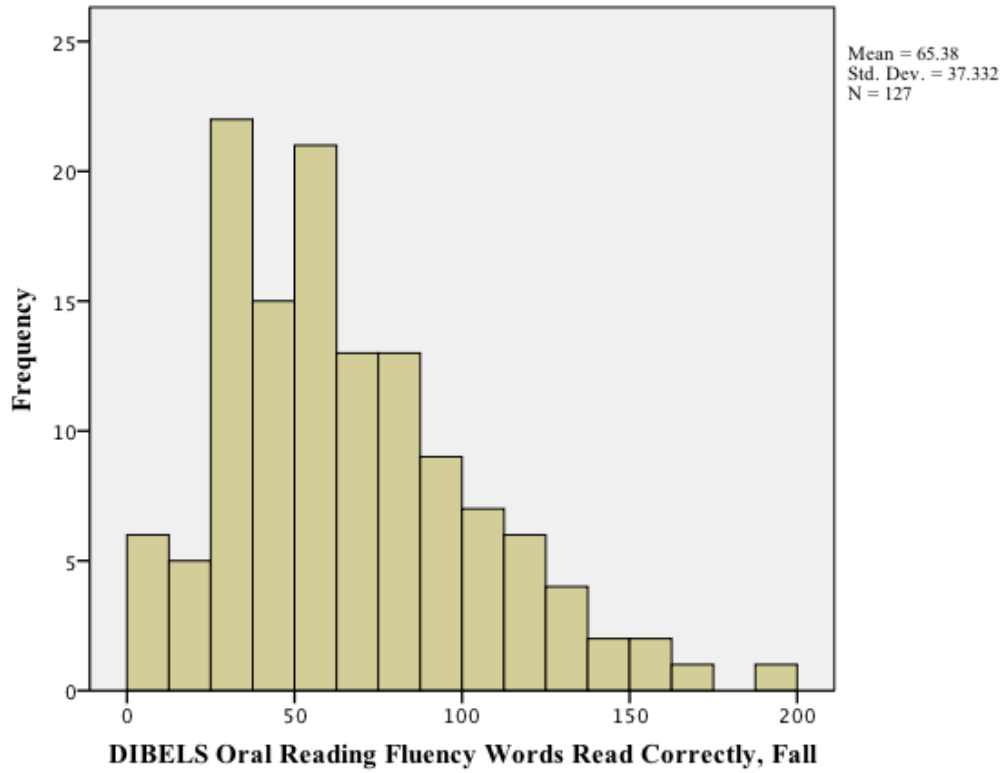


Figure 10  
*Spring SAT 10 Reading Comprehension Score for Non-Hispanic White Students*

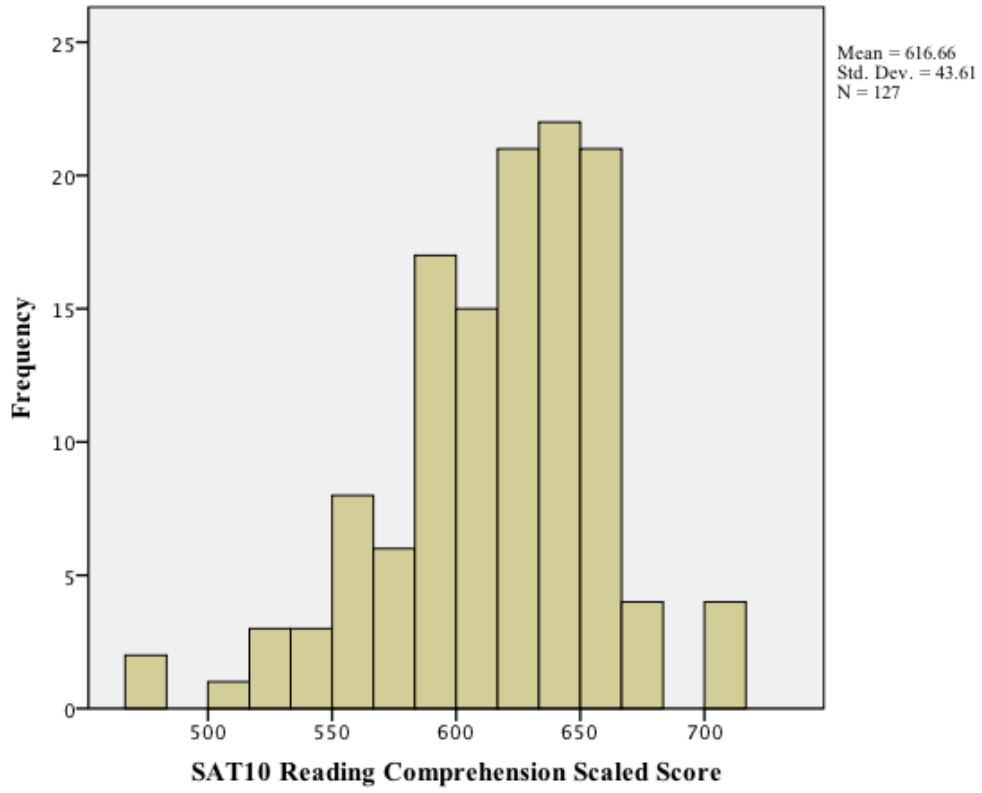


Figure 11  
*DIBELS ORF Scores for Non-Hispanic Minority Students Participating in Spring SAT 10 Reading Comprehension*

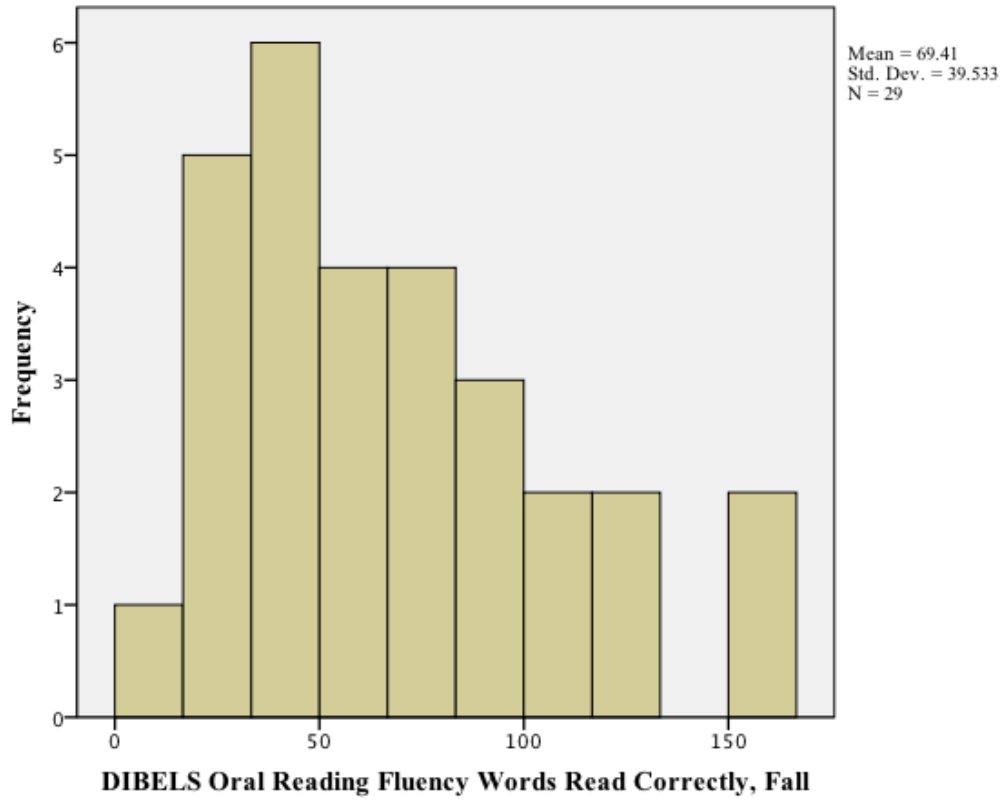
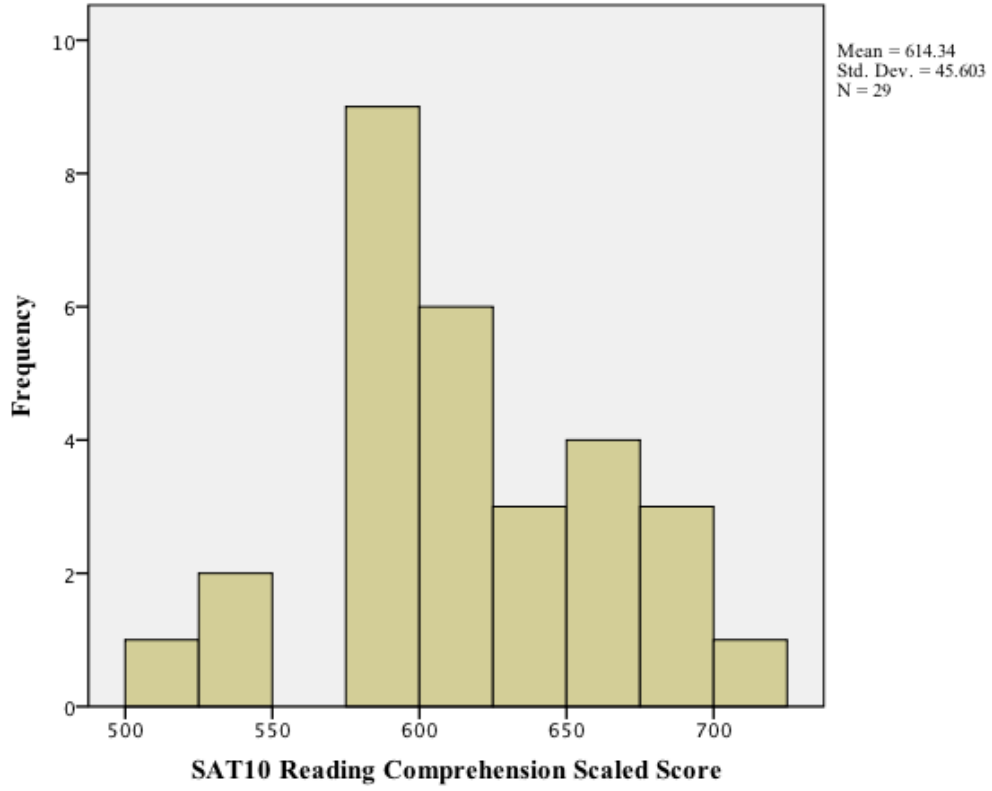


Figure 12  
*Spring SAT 10 Reading Comprehension Scores for Non-Hispanic Minority Students*



## REFERENCES CITED

- Andreassen, R. & Braten, I. (2010). Examining the prediction of reading comprehension on different multiple choice tests. *Journal of Research in Reading, 33*(3), 263-283).
- Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J., Kame'enui, E. J., et al. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review, 37*(1), 18-37.
- Chard, D. J., & Dickson, S. V. (1999). Phonological awareness: Instructional and assessment guidelines. *Intervention in School and Clinic, 34*(5), 261-270.
- Cutting, L.E., & Scarborough, H.S. (2006). Prediction of reading comprehension; Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277-299.
- Darling-Hammond, L. (2010). *The flat world and education; how America's commitment to equity will determine our future*. New York, NY: Teachers College Press.
- Education Enterprise Steering Committee, (2010). *New federal race and ethnicity reporting assistance manual*. Retrieved from Oregon Department of Education website: [www.ode.state.or.us/search/page/?=3718](http://www.ode.state.or.us/search/page/?=3718)
- Fien, H., Park, Y., Baker, S., Smith, J., Stoolmiller, M., & Kame'enui, E. (2010). An examination of the relation of nonsense word fluency initial status and gains to reading outcomes for beginning readers. *School Psychology Review 39*(4), 631-653.
- Fletcher, J.M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10*(3), 323-330.
- Fuchs, L., Fuchs, D., Hosp, M., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading 5*(3), 239-256.
- Goffreda, C., Diperna, J. (2010). An empirical review of psychometric evidence for the dynamic indicators of basic early literacy skills. *School Psychology Review, 30*(3), 463-483.
- Goffreda, C., Diperna, J., & Perderson, J. (2009). Preventive screening for early readers: predictive validity of the dynamic indicators of basic early literacy skills (DIBELS). *Psychology in the Schools, 46*(6), 539-552.

- Good, R.H., III, & Kaminski, R. A. (Eds.) (2002). *Dynamic Indicators of Basic Early Literacy Skills (6<sup>th</sup> ed.)*. Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from: <http://dibels.uoregon.edu/>.
- Good, R.H., III, Laimon, D., Kaminiski, R.A., & Smith, S. (2002). Initial Sound Fluency. In R.H. Good & R.A. Kaminski (Eds.), *Dynamic Indicators of Basic Literacy Skills (6<sup>th</sup> ed.)*. Eugene, OR: Institute for the Development of Educational Achievement. Available: <http://dibels/uoregon.edu?>.
- Good, R.H., Simmons, D.C., & Kame'enui, E.J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading* 5(3), 257-288.
- Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *The Journal of Special Education*, 34(2), 90-103.
- Hagan-Burke, S., Burke, M., and Crowder, C. (2006). The convergent validity of the Dynamic Indicators of Basic Early Literacy skills and the Test of Word Reading Efficiency for the beginning of first grade. *Assessment for Effective Instruction*, 31(4), 1-15.
- Hintze, J. M., Callahan, J. E. I. I. I., Matthews, W. J., Williams, S. A. S., & Tobin, K. G. (2002). Oral reading fluency and prediction of reading comprehension in African American and Caucasian elementary school children. *School Psychology Review*, 31(4) 540-553.
- Jamgochian, E., Park, B.J., Nese, J.F.T., Lai, C., Sáez, L., Anderson, D., Alonzo, J. Tindal, G. (2010). Technical adequacy of the easyCBM grade 2 reading measures (Technical Report 1004) Eugene, Oregon: University of Oregon, Behavioral Research and Teaching.
- Johnson, E., Jenkins, J., Petscher, Y., Catts, H. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice*, 24(4), 174-185.
- Klein, J. R., & Jimerson, S. R. (2005). Examining ethnic, gender, language, and socioeconomic bias in oral reading fluency scores among Caucasian and Hispanic students. *School Psychology Quarterly*, 20, 23–50.
- Keenan, J., Betjemann, R.S., & Olson, R. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12, 281-300.

- Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement of reading. *School Psychology Quarterly, 14*, 327-342.
- McPherson, E. (2010). Beyond the nation's capital: Minority students' stumbling on the tracks after Hobson. *Education and Urban Society, 42*, 795-816.
- Pearce, L. R., & Gayle, R. (2009). Oral reading fluency as a predictor of reading comprehension with American Indian and White elementary students. *School Psychology Review, 38*(3), 419-427.
- Pearson, P.D., & Hamm, D.N. (2005). The assessment of reading comprehension: A review of practices – past, present, and future. In S.G. Paris and S.A. Stahl (Eds.), *Children's reading comprehension and assessment* (p.13-69). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Riedel, B. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first grade students. *Reading Research Quarterly 42*(4), 546-567.
- Schilling, S., Carlisle, J., Scott, S., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *The Elementary School Journal 107*(5), 429-448.
- Speece, D.L., Mills, C., Ritchey, K.D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *Journal of Special Education 36*(4), 223-233.
- Speece, D. L., & Ritchey, K. D. (2005). A longitudinal study of the development of oral reading fluency in young children at risk for reading failure. *Journal of Learning Disabilities, 38*(5), 387-399.
- Valencia, S., Smith, A., Reece, A., Li, M., Wixson, K., & Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly 45*(3), 270-291.