

The Ability of Oral Fluency to Predict Reading Comprehension
Among ELL Children Learning to Read

by

Joseph Ridge Millett

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2011 by the
Graduate Supervisory Committee:

Robert Atkinson, Chair
Jay Blanchard
Marilyn Thompson
James Christie

ARIZONA STATE UNIVERSITY

December 2011

ABSTRACT

The current study analyzed existing data, collected under a previous U.S. Department of Education Reading First grant, to investigate the strength of the relationship between scores on the first- through third-grade Dynamic Indicators of Basic Early Literacy Skills - Oral Reading Fluency (DIBELS-ORF) test and scores on a reading comprehension test (TerraNova-Reading) administered at the conclusion of second- and third-grade. Participants were sixty-five English Language Learners (ELLs) learning to read in a school district adjacent to the U.S.-Mexico border. DIBELS-ORF and TerraNova-Reading scores were provided by the school district, which administers the assessments in accordance with state and federal mandates to monitor early literacy skill development. Bivariate correlation results indicate moderate-to-strong positive correlations between DIBELS-ORF scores and TerraNova-Reading performance that strengthened between grades one and three. Results suggest that the concurrent relationship between oral reading fluency scores and performance on standardized and high-stakes measures of reading comprehension may be different among ELLs as compared to non-ELLs during first- and second-grade. However, by third-grade the correlations approximate those reported in previous non-ELL studies.

This study also examined whether the Peabody Picture Vocabulary Test (PPVT), a receptive vocabulary measure, could explain any additional variance on second- and third-grade TerraNova-Reading performance beyond that explained by the DIBELS-ORF. The PPVT was individually administered by researchers collecting data under a Reading First research grant prior to the current study.

Receptive vocabulary was found to be a strong predictor of reading comprehension among ELLs, and largely overshadowed the predictive ability of the DIBELS-ORF during first-grade. Results suggest that receptive vocabulary scores, used in conjunction with the DIBELS-ORF, may be useful for identifying beginning ELL readers who are at risk for third-grade reading failure as early as first-grade.

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
CHAPTER	
1 INTRODUCTION.....	1
Overview.....	1
Statement of the Problem	3
Purpose of the Study.....	5
Research Questions.....	6
2 REVIEW OF LITERATURE	7
Dynamic Indicators of Basic Early Literacy Skills (DIBELS)	7
Reading Fluency	9
DIBELS-ORF Correlations with Reading Comprehension Performance	14
English Language Learners (ELLs)	23
DIBELS-ORF Predicting Reading Comprehension Performance among ELLs.....	25
Relationship Between Vocabulary and Reading Comprehension ..	28
Summary	30
3 METHOD	33
Participants.....	33
Demographics	33

	Selection Criteria	33
	Procedure	33
	Instruments.....	34
	Dynamic Indicators of Basic Early Literacy Skills: Oral	
	Reading Fluency (DIBELS-ORF).....	34
	TerraNova-Reading, Second Edition	35
	Peabody Picture Vocabulary Test (PPVT)	36
	Analyses of Data.....	37
4	RESULTS	38
	Descriptive Statistics	38
	Correlation Analyses	39
	Multiple Regression Analyses.....	40
	DIBELS-ORF and PPVT Predicting Second-Grade	
	TerraNova Reading.....	41
	DIBELS-ORF and PPVT Predicting Third-Grade TerraNova-	
	Reading	42
5	DISCUSSION	45
	What is the Correlation Between Performance on the Second-Grade	
	TerraNova-Reading Test and DIBELS-ORF Scores from Grades	
	One and Two, Respectively?.....	46
	What is the Correlation Between Performance on the Third-Grade	
	TerraNova-Reading Test and DIBELS-ORF Scores from Grades	
	One, Two, and Three?	48

Will a Receptive Vocabulary Measure (PPVT-III) Explain any Additional Variance on TerraNova-Reading Performance Beyond that Explained by the DIBELS-ORF?.....	49
Limitations	53
Directions for Future Research.....	54
Conclusion	56
REFERENCES	59
APPENDIX	
A TABLES AND FIGURES.....	72

LIST OF TABLES

Table		Page
A1.	Assessment Calendar	73
A2.	Means, Standard Deviations, and Skew Statistics for Each Measure	73
A3.	Correlations among All Measures	74
A4.	Coefficients for Models Predicting Second-Grade TerraNova- Reading.....	74
A5.	Results for Models Predicting Second-Grade TerraNova-Reading	75
A6.	Coefficients for Models Predicting Third-Grade TerraNova- Reading.....	75
A7.	Results for Models Predicting Third-Grade TerraNova-Reading ...	76

LIST OF FIGURES

Figure		Page
A1.	Distribution of First-Grade DIBELS-ORF Scores.....	77
A2.	Distribution of second-Grade DIBELS-ORF Scores.....	78
A3.	Distribution of Third-Grade DIBELS-ORF Scores	79
A4.	Distribution of Second-Grade TerraNova-Reading Scores	80
A5.	Distribution of Third-Grade TerraNova-Reading Scores	81
A6.	Distribution of First-Grade PPVT Scores.....	82
A7.	Distribution of Second-Grade PPVT Scores	83
A8.	Distribution of Third-Grade PPVT Scores.....	84
A9.	Scatter plot Matrix Depicting the Relationship Between All Variables.....	85

Chapter 1

INTRODUCTION

Overview

In response to a congressional directive in 1997, the National Reading Panel (NRP) was formed to conduct a comprehensive review of the available research to determine the most effective, scientifically-based, methods for teaching children to read. The 2000 NRP results included a recommendation for the explicit instruction in the areas of phonemic awareness, phonics, vocabulary, text comprehension, and fluency. Development of these skills was characterized as essential for reading development. These basic early literacy skills form the foundation upon which the ultimate goals of reading proficiency and comprehension are developed (Carlisle, Beeman, Davis, & Spharim, 1999; Chaney, 1998; Mattingly, 1972; NRP Report, 2000; NICHD Early Child Care Research Network, 2005; Snow, Burns, & Griffin 1998; Snowling, 2005).

The publication of the 2000 NRP report, along with the passage of the 2001 United States No Child Left Behind Act (NCLB) and associated federal initiatives such as Reading First, have resulted in a dramatic increase of attention focused on assessing and monitoring early literacy skill development in young children. Although technically expired since September 30, 2007, NCLB has been temporarily extended until reauthorization occurs. Until then, NCLB requires that all schools receiving Title I funds make Adequate Yearly Progress (AYP). Meeting this requirement includes having a sufficient proportion of the students reach proficient reading levels each year as measured by standardized reading

achievement tests. For students in kindergarten through third grade, schools are required to implement scientifically-based reading programs to ensure that all students are reading at or above grade level by the end of third grade. As part of making AYP, schools are held accountable for the progress of all subgroups, including disadvantaged and at-risk students who traditionally have been left behind in past educational reform efforts (Deschenes, Cuban, & Tyack, 2001; Mahon, 2006; Wiley & Wright, 2004). Groups identified as at-risk for underachievement in reading include English language learners (ELLs). Spanish-speaking ELL students are the fastest growing school-aged group in the U.S. (Capps, Fix, Murray, Ost, Passel, & Herwanto, 2005; Goldenberg, 2010; National Clearinghouse for English Language Acquisition, 2009).

In order to make AYP each year, and to have all children reading at or above grade level by third grade, it is imperative to be able to determine which children are on pace developing age-appropriate early literacy skills, and which children are in need of intervention (Ericson & Juliebo, 1998; Good, Simmons, & Kame'enui, 2001; National Research Council, 1999). The Reading First initiative of the NCLB Act provided grant support for schools that use approved scientifically-based assessment tools as part of their programs to (a) monitor development of basic early literacy skills, (b) predict reading difficulties as early as possible, and (c) help ensure that all children learn to read by the third grade. A widely used assessment tool for monitoring basic early literacy skill development and predicting reading difficulties is the Dynamic Indicators of Basic Early Literacy Skills Oral Reading Fluency test (DIBELS-ORF; 6th Ed., Good &

Kaminski, 2002). The DIBELS-ORF specifically measures reading fluency, one of the five basic early literacy skills that have been identified as essential for successful reading achievement (National Reading Panel, 2000). The DIBELS-ORF is currently being administered to both native English speaking children and ELL children alike as a probe to inform decision making about educational needs. While validity studies exploring the relationship between the DIBELS-ORF and state mandated reading achievement tests for non-ELL children learning to read have become more plentiful over recent years, the same cannot be said for ELL children learning to read (Klingner & Edwards, 2006).

Statement of the Problem

With the NCLB mandate to close the achievement gap that exists between disadvantaged/at-risk and advantaged children, numerous concerns and questions still surround the issue of how to accurately monitor literacy skill development among children learning English. Stakeholders generally agree that it is important to be able to make informed decisions about the appropriate interventions that may be necessary to ensure that ELL children remain on track to be reading by third grade. The results of several studies in recent years have indicated that the DIBELS-ORF, in particular, can be predictive of third grade reading achievement test outcomes (Buck & Torgesen, 2003; Good, Simmons, & Kame'enui, 2001; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Schilling, Carlisle, Scott, & Zeng, 2007; Shapiro, Solari, & Petscher, 2008; Vander Meer, Lentz, & Stoller, 2005; Wood, 2006). The problem remains, however, that the available research has been largely limited to an English speaking population. To date, very little

empirical data exists indicating whether or not the same predictive relationship applies to ELLs (Baker, Baker, Katz, & Otterstedt, 2009; Riedel, 2007; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Wilson, 2005). None of the studies reviewed looked exclusively at ELLs. The majority of studies that included ELLs had only a small proportion as part of the sample, and most did not analyze ELLs separately

The school-aged ELL population enrolled in U.S. public schools has been growing at an accelerated rate over the past few decades (Boyle, Taylor, Hurlburt, & Soga, 2010; Goldenberg, 2010; National Clearinghouse for English Language Acquisition, 2009). A large proportion of ELLs are in the early elementary grades. Around 40 percent of the ELL population enrolled in public schools is concentrated between kindergarten and third-grade (Lui, Ortiz, Wilkinson, Robertson, & Kushner, 2008), the most critical period for literacy skill development (NRP Report, 2000). In addition to having a large ELL population, Arizona is an English immersion classroom state. For schools and school districts with high proportions of ELLs, these conditions greatly compound the difficulty of making AYP.

Given these conditions coupled with the current testing and accountability climate, casting light on this largely unexplored area of research is long overdue. It is critically important to examine the usefulness of DIBELS-ORF in predicting performance on the state mandated standardized reading achievement test being used with Spanish-speaking ELL children learning to read since (a) performance on a state mandated reading achievement test is what determines whether or not a

third grader is proficient in reading, and (b) the DIBELS-ORF is so widely used for monitoring basic literacy skill development during first through third grades (ELLs included).

Purpose of the Study

The widespread adoption of the DIBELS-ORF as a means of monitoring progress in oral reading fluency development seems to assume that performance on the DIBELS-ORF will be related to performance on subsequent reading achievement tests among ELLs and non-ELLs alike. Few studies have investigated the long term predictive utility of the DIBELS-ORF (i.e. across grade levels), and even less were found to include a significant cohort of ELLs. Given the widespread use of the DIBELS-ORF, it is imperative that both the short- and long-term diagnostic accuracy of this measure be understood for ELLs. For example, if it is possible to accurately identify ELL students in first grade who are at-risk for reading failure on the high stakes reading achievement test given in third grade, the chances for a timely and effective intervention are maximized. To address this issue, the current study will examine the correlations between first through third-grade DIBELS-ORF scores and performance on second- and third-grade reading achievement tests (measuring reading comprehension).

A few studies in the literature on the DIBELS-ORF suggest that the DIBELS-ORF may vary in predictive accuracy of third-grade reading comprehension at some levels, and/or for some students (Riedel, 2007; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Schilling, Carlisle, Scott, & Zeng, 2007; Wilson, 2005). A recent research study suggests that it may be useful to add

a vocabulary component to a predictive model using DIBELS-ORF to predict performance on a reading achievement test (Riedel, 2007). The current study will shed additional light on this preliminary finding by examining whether or not a receptive vocabulary measure would add any predictive utility for reading comprehension beyond that provided by the DIBELS-ORF.

Research Questions

1. What is the correlation between performance on the second-grade TerraNova-Reading test and DIBELS-ORF scores from grades one and two, respectively?
2. What is the correlation between performance on the third-grade TerraNova-Reading test and DIBELS-ORF scores from grades one, two, and three, respectively?
3. Will a receptive vocabulary measure (PPVT-III), administered yearly in first through third grade, explain any additional variance on second- and third-grade TerraNova-Reading performance, beyond that explained by the DIBELS-ORF?

Chapter 2

REVIEW OF LITERATURE

The current literature review will give a brief overview of the DIBELS-ORF, its origins, and what it is intended to measure. Reading fluency will then be defined, along with a discussion of its relevance to reading comprehension. A review of the available research investigating the relationship between the DIBELS-ORF and outcomes on standardized year-end reading achievement tests measuring reading comprehension will then be presented. The research discussed will focus on DIBELS-ORF and state mandated reading achievement correlational and predictive studies (both concurrent and subsequent criterions) between grades one and three. A discussion on using a vocabulary measure to add predictive utility to the DIBELS-ORF will be provided, along with a brief review of literature linking vocabulary to reading comprehension. Finally, after discussing the status of ELLs, both in the U.S. and in Arizona specifically, the available studies involving ELLs and the DIBELS-ORF will be reviewed.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

DIBELS is an assessment tool used to monitor early literacy skill growth. DIBELS is a battery of seven subtests (Word Use Fluency, Initial Sound Fluency, Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, Oral Reading Fluency, and Retell Fluency) designed to function as indicators of key early reading skills. It is widely adopted in the United States as a result of its inclusion in the Department of Education, NCLB Reading First grants. It remains unclear, however, how closely related all of the DIBELS

subtests are to reading comprehension. The subtest with the strongest empirically supported relationship to reading comprehension is the Oral Reading Fluency (ORF) subtest; (Buck & Torgesen, 2003; Good, Simmons, & Kame'enui, 2001; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Schilling, Carlisle, Scott, & Zeng, 2007; Shapiro, Solari, & Petscher, 2008; Vander Meer, Lentz, & Stoller, 2005; Wood, 2006).

DIBELS-ORF developers assert that the ORF “is a standardized set of passages and administration procedures designed to (a) identify children who may need additional instructional support, and (b) monitor progress toward instructional goals” (Good & Kaminski, 2002, p. 30). Development of the DIBELS battery was based on the Curriculum-Based Measurement (CBM) concept (Deno, 1985). The basic intent behind CBM is frequent monitoring of skills using a set of quick (usually one minute), simple, and inexpensive standardized probes that are based on the current curriculum, and sensitive to literacy skill growth (Deno, 2002; Wiley & Deno, 2005). The DIBELS-ORF is a type of CBM measure, and as such is intended to be a quick and efficient measure of reading fluency in terms of accurate decoding of connected text (Good & Kaminski, 2002).

Using calibrated reading passages for each grade level, scores on the DIBELS-ORF are simply the number of words read correctly per minute (WCPM). Benchmark goals for the DIBELS-ORF have been set for each grade level that are to be achieved by the spring of that year in order for a student's level of fluency to be considered on target. The low-risk benchmark goals are 40

WCPM for first grade, 90 WCPM for second grade, and 110 WCPM for third grade. Students who achieve benchmark level WCPM or higher are considered low-risk for reading failure. Students are considered at-risk and in need of extra instructional support if WCPM scores fall below 20, 70, and 80 at each grade, respectively (Good & Kaminski, 2002).

Even though comprehension rather than fluency is the ultimate goal of reading instruction, research maintains that fluency is closely associated to comprehension (Deno, 1985; Deno, Mirkin, Chiang, & Lowry, 1980; Fuchs, Fuchs, & Deno, 1982; Fuchs, Fuchs, & Maxwell, 1988; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Kuhn & Stahl, 2003; McGlinchey & Hixson, 2004; Stage & Jacobsen, 2001), and is a reflection of how automated the lower level reading processes are that are necessary for comprehension (Good, Simmons, & Kame'enui, 2001; Schilling, Carlisle, Scott, & Zeng, 2007). Thus it is important to define what fluency is and identify why it is important for reading comprehension.

Reading Fluency

To begin, it is important to make a distinction between oral reading fluency and silent reading fluency. While silent reading is a more common feature of everyday life, the NRP (2001) concluded that guided repeated oral reading, and not independent silent reading, facilitated the development of the following: word recognition, reading fluency and reading comprehension. The prominence of the DIBELS battery in Reading First programs is evidence of the emphasis that has been placed on oral reading over silent reading in recent years.

Research investigating the difference between oral reading fluency and silent reading fluency suggests that they are different constructs, and invoke different cognitive processes (Berent & Perfetti, 1995; de Jong, & Share, 2007; Frost, 1998; Poeppel, 2001). It appears that oral reading is not simply the same process as silent reading with the added component of verbal output. Oral reading tasks require a higher level of phonological processing, while silent reading may take a more direct lexical pathway (Berninger, Abbott, Trivedi, Olson, Gould, Hiramatsu, Holsinger, McShane, Murphy, Norton, Boyd, & Westhaggen, 2010; Bookheimer, Zeffiro, Blaxton, Gaillard, & Theodore, 1995; de Jong, & Share, 2007; Jennings, McIntosh, Kapur, Tulving, & Houle, 1997; Poeppel, 2001). However, among beginning readers, the more indirect, phonological processing route is heavily relied upon for both oral and independent (silent) reading (Morton, 1989; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Sprenger-Charolles & Casalis, 1995; Sprenger-Charolles, Siegel, & Bechenec 1998). The dependence a beginning reader has on phonological processing is evident by their preference for oral reading as they learn to more automatically link letters to sounds (Birch, 2002).

Fluency has been described as a neglected component of reading (Allington, 1983; NRP Report, 2000). Different definitions of oral reading fluency (referred to as fluency from here on) that emphasize aspects such as rate, expression, accuracy, stress, automaticity, pauses, pitch, etc., have been proposed over the past several decades (Chomsky, 1976; Dahl & Samuels, 1974; Dowhower, 1987; Herman, 1985; Laberge & Samuels, 1979; Schreiber, 1980).

The National Reading Panel (2000) asserts that a fluent reader “can read text with speed, accuracy, and proper expression” (chap. 3, p.5). However, others view such aspects as only symptomatic of fluency. For example, Samuels (2006) argues that aspects such as speed and expression are only indicators of fluency and not essential characteristics of fluency itself. Samuels (2007) defines fluency as simply, “the ability to decode and comprehend text at the same time” (p. 9). However, Riedel (2007) argued that fluency and comprehension should be viewed as separate constructs and that comprehension should not necessarily be a condition of fluency.

Although fluency definitions are the topic of much debate, most definitions of fluency are based on LaBerge and Samuels’ (1974) information processing, or automaticity, theory of reading. The theory asserts that the capacity of the working memory in cognitive processing is limited, especially when attending to more than one task at a time (e.g. decoding and comprehending). When reading is labored and slow, or rather, when too much attention is focused on the lower-level processes of reading (e.g. decoding), capacity for the higher-level process of comprehension will be limited (NRP Report, 2000; Perfetti, 1985, 2007). In order to maximize the amount of conscious attention allotted toward comprehension the reader needs to be able to recognize words with automaticity and minimize the amount of thinking about reading (decoding). Thus, automaticity becomes an important aspect of fluency that ultimately leads to comprehension. In other words, fluent reading is a strong indicator of how

automated the lower-level processes of reading are (Good, Simmons, & Kame'enui, 2001; Schilling, Carlisle, Scott, & Zeng, 2007).

Today, most of the research literature describes accuracy and rate as the two components that constitute what is meant by automaticity (Kuhn & Stahl, 2003). In fact, many researchers today define fluency in terms of accuracy and rate of reading (Dowhower, 1991; Fuchs & Fuchs, 1992; Hasbrouk & Tindal, 2006; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Torgesen, Rashotte, & Alexander, 2001). Simply put, accuracy refers to correctly decoding words and rate refers to the speed at which connected text is decoded. The speed of oral reading seems to have a positive relationship with reading comprehension (Daane, Campbell, Grigg, Goodman, & Oranje, 2005).

Skepticism has grown over equating fluency with speed of reading. There is growing criticism that widely used assessment tools, such as the DIBELS-ORF, approach fluency as simply a race to read words as quickly as possible (Mathson, Allington & Solic, 2005; Rasinski, 2004). In response, some researchers now include prosody in addition to accuracy and rate as the primary components of fluency (Hudson, Lane & Pullen, 2005; Kamhi, Allen & Catts, 2001; Kuhn & Stahl, 2003; Mathson, Allington & Solic, 2005; National Reading Panel Report, 2000). Prosody refers to appropriate use of intonation features such as pauses, pitch, and stress. Some researchers view prosodic reading, or reading with appropriate expression, as an important component of fluency because it is a good indication of comprehension (Dowhower, 1991; Rasinski & Zutell, 1990)

Precisely whether or not prosody is necessary for comprehension, or if it can be used as an index for determining levels of comprehension remains unclear (Dowhower, 1991; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Mathson, Allington & Solic, 2005; Schwanenfluegel, Hamilton, Huhn, Wisenbaker, & Stahl, 2004). A compelling argument for defining fluency as simply the rate of accurate decoding of connected text is that it makes the measurement of fluency much simpler (Hasbrouck & Tindal, 1992). For example, researchers in a study conducted for the National Assessment of Educational Progress (NAEP) defined fluency in terms of “phrasing, adherence to the author’s syntax, and expressiveness” (p. 37). In doing so, they found fluency much more difficult to measure (Daane, Campbell, Grigg, Goodman, & Oranje, 2005). The results of another study indicate that including prosody does not add much variance explained over and above what rate and accuracy is able to explain when predicting comprehension (Schwanenfluegel, Hamilton, Huhn, Wisenbaker, & Stahl, 2004).

Fluency, measured as simply the rate of accurate decoding of connected text, has been widely supported as an indicator of reading achievement. (Adams, 1990; Allington, 1983; Fuchs & Fuchs, 1992; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Hasbrouck & Tindal, 1992; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Samuels, Schermer, & Reinking, 1992; Torgesen, Rashotte, & Alexander, 2001), and has been shown to be strongly correlated with reading comprehension specifically (Breznitz, 1987; Daane, Campbell, Grigg, Goodman, & Oranje, 2005; Deno, 1985; Deno, Marston, Shinn, & Tindal, 1983; Dowhower, 1987; Fuchs, Fuchs, & Maxwell, 1988; Hosp & Fuchs, 2005; McGlinchy & Hixson 2004;

Rasinski, 1990; Tenenbaum & Wolking, 1989; Wanzek, Roberts, Linan-Thompson, Vaughn, & Murray, 2010).

It should be noted that not all studies examining reading fluency and reading comprehension use the DIBELS-ORF specifically to measure oral reading fluency. There are several CBM-type oral reading fluency measures currently being used with similar methods for administration. A few examples of alternate CBM oral reading fluency probes include the Gray Oral Reading Test (GORT), fourth edition (Weiderholt & Bryant, 2001), and the Test of Oral Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999). However, for consistency purposes, the current study will review only the research that explores the relationship between the DIBELS-ORF and standardized reading assessments measuring reading comprehension that are currently being used for AYP status and high-stakes decision making.

DIBELS-ORF Correlations with Reading Comprehension Performance

A frequent criticism of the DIBELS battery is that the subtests are not valid predictors of reading comprehension despite their wide spread use for predicting future reading difficulty (Goodman, 2006; Pearson, 2006; Pressley, Hilden, & Shankland, 2005). Given such criticism, and the fact that reading achievement is ultimately determined by state mandated assessments given each year, there is a need for continued research exploring how well the DIBELS-ORF predicts outcomes on year-end standardized and high stakes reading achievement tests measuring reading comprehension.

Most of the aforementioned research investigated the correlation between the DIBELS-ORF and reading comprehension tests when administered at approximately the same time (concurrent administrations), both in spring of the same year. Even though it is recommended that DIBELS administration begin in first grade, only a few studies were found to have investigated DIBELS-ORF before third grade (Baker, Smolkowski, Katz, Fien, Seeley, Kame'enui, & Beck, 2008; Riedel, 2007; Schilling, Carlisle, Scott, & Zeng, 2007). While establishing concurrent validity is important, establishing the predictive validity of the DIBELS-ORF across a significant time period (i.e. grade levels) appears more critical in establishing the DIBELS-ORF as a valuable tool to help identify students in need of reading intervention as early as possible. Early in the past decade, very few DIBELS-ORF predictive validity studies were published. Although within-grade predictive studies have become more numerous over the most recent decade, very few have examined the predictive utility of the DIBELS-ORF across grade levels. Studies currently found in the literature generally indicate a moderate to strong relationship between the DIBELS-ORF and reading comprehension as measured by various standardized reading achievement tests currently being used in different states throughout the U.S.

Two separate studies were conducted in Colorado that investigated the relationship between DIBELS-ORF and the Colorado Student Assessment Program (CSAP). CSAP is a statewide reading test measuring reading comprehension. In the early study, Shaw and Shaw (2002) reported some of the strongest correlations found between reading comprehension measures and the

DIBELS-ORF. Although the sample of third-graders was small ($n = 58$), correlation coefficients of .73 for both the fall and winter administrations and a correlation of .80 for the spring administration were found. In the subsequent study, Wood (2006) found a similar relationship albeit the participants were 281 third, fourth, and fifth-graders. Looking at how consistent the DIBELS-ORF and CSAP correlation would remain across the three years, correlations ($p < .001$) of .70 (third-grade), .67 (fourth-grade), and .75 (fifth-grade) were found. Wood described these results as contradictory to previous studies that suggest a weakening relationship between oral reading fluency and reading comprehension over time (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993; Fuchs, Fuchs, Hosp, & Jenkins, 2001).

Two years later, Shapiro, Solari, and Petscher (2008) conducted a study similar to Wood (2006) by also investigating the relationship between DIBELS-ORF and a statewide reading achievement test among third through fifth graders in Pennsylvania. The Pennsylvania System of School Assessment (PSSA) is the measure designed for educational accountability purposes in Pennsylvania, with the reading portion measuring comprehension. Using an ethnically diverse sample of 1000 students, the results were correlations similar to those reported by Wood (2006). Across the three years, concurrent DIBELS-ORF and PSSA correlations were .68 (third and fourth grades), and .75 (fifth-grade).

Pressley, Hilden, and Shankland (2005) conducted a concurrent validity study using DIBELS-ORF to predict grade equivalent scores on the reading comprehension portion of the TerraNova, both administered in the spring of third-

grade. Participants were 191 third-graders with 39 percent classified as minority. The correlation reported by Pressley and colleagues ($r = .44$) was the weakest found between the DIBELS-ORF and a standardized reading comprehension test. It was reported that less than 20 percent of variance in TerraNova scores was explained by the DIBELS-ORF. However, the authors suggested that the comprehensive design of the TerraNova might have played a contributing role to the weak relationship. It was explained that most studies have used state achievement tests that have a summative design geared toward minimal competency.

In Florida, two studies evaluated the relationship between the DIBELS-ORF and the reading comprehension portion of the Florida Comprehensive Assessment Test-Sunshine State Standards (FCAT-SSS). In the earlier study, Buck and Torgesen (2003) reported a concurrent correlation of $.70$ ($p < .001$) with a sample consisting of 1,102 third-graders (83 percent white). Later in 2008, using an extremely large sample ($n = 35,207$) Roehrig, Petscher, Nettles, Hudson, and Torgesen found the identical result ($r = .70$) within the third grade. However the more recent sample of participants was much more ethnically diverse (36 percent white, 36 percent African American, 23 percent Latino, 5 percent other), with 20 percent listed as ELL. Roehrig and colleagues also looked at the earlier third-grade fall and winter administrations of the DIBELS-ORF in predicting spring FCAT-SSS results, with correlations of $.67$ and $.68$ reported respectively. Further, DIBELS-ORF administrations were used to predict performance on the Stanford Achievement Test-Tenth Edition (SAT-10). The correlations were nearly identical

to those of the FCAT-SSS, with coefficients reported as .69 (fall), .68 (winter), and .70 (spring). DIBELS-ORF was also found to predict FCAT-SSS scores equally well across demographics of ethnicity, language, and socio-economic status.

Further, both studies evaluated how accurately the risk-level cutoffs of the DIBELS-ORF (at-risk and low-risk) predicted performance on the FCAT-SSS. Buck and Torgesen (2003) reported that while only nine percent of third graders categorized as low-risk on the DIBELS-ORF did not achieve adequate performance on the FCAT-SSS, nearly 20 percent of the at-risk students ended up meeting the FCAT-SSS reading criterion. Roehrig and colleagues (2008) also found the DIBELS-ORF to be an inaccurate predictor for a significant number of students, and suggested that recalibrating the DIBELS-ORF risk-level cut scores would yield a greater proportion of students correctly identified as at-risk or low-risk in regard to performance on the FCAT-SSS. Using recalibrated cutoffs, the number of students correctly identified as at-risk or low-risk by DIBELS-ORF was reported to improve by as much as 10 percent. Other researchers have documented the predictive inaccuracy across cutoff categories on the DIBELS-ORF (Riedel, 2007; Schilling, Carlisle, Scott, & Zeng, 2007; Wilson, 2005).

In the state of Oregon, Schilling, Carlisle, Scott, and Zeng, (2007) investigated both the concurrent and predictive utility of the DIBELS-ORF using a very large sample made up of 2,588 first-graders, 2,437 second-graders, and 2,527 third-graders. In the spring of each year students, grades one through three, take the Iowa Test of Basic Skills (ITBS), a standardized test with a reading

comprehension component. Spring concurrent correlations ($p < .001$) between the reading comprehension subtest of the ITBS and the DIBELS-ORF were reported as .74, .75, and .63 for first, second, and third grades respectively. Using hierarchical regression analyses to determine the amount of variance in the ITBS reading total explained by DIBELS-ORF combined with other DIBELS subtests, (Phoneme Segmentation Fluency, Nonsense Word Fluency, and Word Use Fluency) the DIBELS-ORF by far made the largest contribution in each model. Across the three grades, the amount of variance accounted for by the DIBELS-ORF was 48 percent (first-grade), 56 percent (second-grade), and 45 percent (third-grade). An interesting note is that these results seem to support previous studies suggesting a weakening relationship between reading fluency and reading comprehension measures as reading ability improves (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993; Fuchs, Fuchs, Hosp, & Jenkins, 2001).

The accuracy of DIBELS at-risk and low-risk categories in predicting ITBS performance was also evaluated. It was determined that the DIBELS-ORF was reasonably accurate at identifying students in the second and third grades who would be reading below grade level on the ITBS at the end of the year. Among students scoring in the at-risk category in the fall, 80 percent of second-graders and 76 percent of third-graders scored below the twenty-fifth percentile on the ITBS in spring. However, the ITBS performance of students scoring in the low- and some-risk categories was not as accurately predicted. From the fall administration of the DIBELS-ORF, 32 percent of second-graders and 37 percent of third-graders identified as low-risk, and around 70 percent of second- and

third-graders identified as some-risk, were not reading at grade level on the ITBS in the spring.

In the state of Michigan, a study similar to Roehrig et al. (2008) was conducted that also used the DIBELS-ORF to predict achievement on the SAT-10 (Baker, Smolkowski, Katz, Fien, Seeley, Kame'enui, & Beck, 2008). A large sample was used ($n = 9,600$) that consisted of Latino, Asian, American Indian, and Hawaiian Pacific Islander students. Among Michigan schools, the SAT-10 is administered to first and second-graders before moving on to the Oregon Statewide Reading Assessment (OSRA) in third-grade. The OSRA primarily measures reading comprehension. The SAT-10 consists of subtests measuring word reading, sentence reading, and reading comprehension. The study did not provide disaggregated SAT-10 results to look at reading comprehension exclusively.

First-grade SAT-10 correlations of .72 and .82, respective to first grade winter and spring DIBELS-ORF administrations, were reported. For second-grade, SAT-10 and DIBELS-ORF correlations were .72 (fall), .79 (winter), and .80 (spring). For third-grade, DIBELS-ORF concurrent correlations with the OSRA were .65 (fall), .68 (winter), and .67 (spring). Setting this study apart from most other DIBELS-ORF studies was that it also looked at the predictive validity of DIBELS-ORF across grades. For the second-grade SAT-10, correlations with DIBELS-ORF assessments from winter and spring of first-grade were .63, and .72. Correlations between DIBELS-ORF assessments from second-grade and the OSRA were .58 (fall), .63 (winter), and .63 (spring).

Riedel (2007) conducted a study consisting of 1,518 first-graders from the Memphis City Schools district in Tennessee. It is worth noting that this study included a small portion of ELLs. African Americans made up 92 percent of the sample. The first grade winter and spring administrations of the DIBELS-ORF were used to predict outcomes on reading comprehension achievement tests administered at the end of grades 1 and 2. The test used at the end of first-grade is the Group Reading Assessment and Diagnostic Evaluation (GRA+DE). The GRA+DE is a standardized test of overall reading ability that assesses vocabulary along with comprehension. The reading comprehension test used at the end of second-grade was the TerraNova. Among the monolingual children, DIBELS-ORF correlations with the GRA+DE were .59 (winter) and .67 (spring). The correlations with the TerraNova were .49 (winter) and .54 (spring).

Stepwise logistic regression analyses using various DIBELS measures to predict performance on the reading comprehension tests were also conducted. It was reported that upon being added into the predictive model, no DIBELS subtest other than ORF contribute a practically significant amount of variance explained in the dependent measures (The other DIBELS subtests used were LNF, PSF, and NWF). DIBELS-ORF predicted status performance (satisfactory or not satisfactory) on the reading tests given at the end of first and second-grade with 80 percent, and 72 percent accuracy, respectively. The addition of any other subtest, or combination of subtests, to the ORF in the predictive model, increased the predictive accuracy by only one percent or less.

Finally, the two most recent studies were Pearce and Gayle (2009), and Wanzek, Roberts, and Linan-Thompson (2010). In the former, DIBELS-ORF predicted performance on the reading comprehension subtest of the Dakota State Test of Educational Proficiency (DStep). The sample was made up of American Indian ($n = 115$) and White ($n = 428$) students. Overall, the correlation was .63 with the DIBELS-ORF accounting for about 40 percent of the variance of the DStep ($p < .001$). Wanzek and colleagues looked at the predictive validity of DIBELS-ORF across grade levels. Using a sample ($n = 461$) with 66 percent of the participants categorized as Hispanic, scores from first, second, and third grades were each used to predict reading comprehension on both the SAT-10 and the Texas Assessment of Knowledge and Skills (TAKS) taken in third grade. DIBELS-ORF correlations were .64 (first), .68 (second), and .69 (third) for the SAT-10; and .51 (first), .57 (second), and .60 (third) for the TAKS.

In sum, the amount of research examining the predictive utility of the DIBELS-ORF has increased over the past ten years. The studies reviewed for the current study generally indicate a moderate-to-strong relationship between the DIBELS-ORF and various standardized reading achievement tests that are currently being used in different states throughout the U.S. to measure reading comprehension. While the samples of some of the studies reviewed included ELLs (in varying proportions), it remains unclear if the same relationship between the DIBELS-ORF and reading comprehension exists when looking exclusively at ELLs.

English Language Learners (ELLs)

According to the National Research Council (1998), ELLs are “students who come from language backgrounds other than English and whose proficiency is not developed enough where they can profit fully from English-only instruction” (August & Hakuta, 1997, p. 15). Unlike other demographic categories such as gender and race, the classification of ELL is fluid. There are different levels of proficiency within the category of ELL. In Arizona, ELL students are currently assessed with a recently revised version of the Arizona English Language Learner Assessment (AZELLA), commonly referred to as AZ2, which measures English proficiency in listening, speaking, reading, and writing. Children who score below the proficient level on the AZELLA are classified as ELL. The proficiency levels currently used are: Pre-emergent, Emergent, Basic, Intermediate, and proficient. This means that English language skills among children classified as ELL in Arizona can vary dramatically from student to student. ELLs are reassessed every year to determine if reclassification of English proficiency level is necessary. Similar categorizations are used in other states, however the process used to identify ELLs, and associated English proficiency levels, varies from state to state.

In the years before NCLB, the needs of ELL children were largely ignored in educational reform initiatives (Snow, Burns, & Griffin, 1998; Stephenson, Johnson, Jorgensen, & Young, 2003). Today, ELLs are the fastest growing school-aged group in the country. Estimates indicate that between the years 1990 and 2001, elementary through secondary ELL enrollment increased by over 100

percent (Kindler, 2002). By 2008 there were over five million ELL children enrolled in U.S. public schools (National Clearinghouse for English Language Acquisition, 2010). Around 11 percent of the total enrollment of elementary level (kindergarten through sixth-grade) students in the United States is ELL (Boyle, Taylor, Hurlburt, & Soga, 2010; Kindler, 2002). Estimates indicate that the ELL numbers will continue to rise at a remarkable rate over the coming years. Some researchers have suggested that by the year 2030, ELLs will comprise around 40 percent of students in public U.S. elementary and secondary schools (Thomas & Collier, 2002)

An analysis of standardized reading test scores across the nation reveals that by the time they reach fourth-grade, ELLs are still lagging behind their non-ELL counterparts in reading achievement-level performance. Seventy-three percent of ELL fourth-graders scored below the basic reading level on standardized reading assessments (Perie, Grigg, & Donahue, 2005).

Arizona has one of the highest ELL concentrations in the nation. Over 14 percent of Arizona's total enrollment is non-proficient in the language of instruction. The vast majority (81 percent) of Arizona ELLs are Spanish speaking (Arizona Auditor General Report, 2008). Since the year 2000, Arizona law mandates that all public school instruction be conducted in English (Arizona Revised Statutes §15-752, 2000). The problem is that non-proficiency in the language of instruction is now identified a risk factor and has been found to be a contributor to poor outcomes on reading assessments (Abedi & Dietel, 2004; Lesaux & Siegel, 2003; Snow, Burns, & Griffin, 1998).

Such findings emphasize the need to identify, as early as possible, the ELLs who are not properly developing basic early literacy skills. To the point, the DIBELS-ORF is an assessment tool currently being used, with ELLs and non-ELLs alike, to monitor development of reading fluency. However, as validity of the DIBELS-ORF as a predictor of standardized reading assessment outcomes becomes more and more established among native English speaking children learning to read, it remains unclear whether it applies to ELL children as well.

DIBELS-ORF Predicting Reading Comprehension Performance among ELLs

Studies investigating assessments currently used to monitor progress among ELLs before third-grade are rare. Only a few studies were found that investigated the use of DIBELS-ORF specifically among ELLs learning to read (Baker, Baker, Katz, & Otterstedt, 2009; Riedel, 2007; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Wilson, 2005). Most studies that include ELLs have very small ELL proportions in the sample, and are usually not disaggregated to look at ELLs and non-ELLs separately. None of the studies reviewed for the present study looked exclusively at the DIBELS-ORF predicting performance on standardized reading achievement tests among ELLs (between first- and third-grade). There were four studies found that merit mentioning.

In Arizona, Wilson (2005) looked at the spring concurrent correlations between the DIBELS-ORF and the Arizona Instrument to Measure Standards (AIMS) among 241 third graders. The AIMS is an achievement test measuring grade-level reading standards with an emphasis on comprehension. An overall

correlation of .74 between the DIBELS-ORF and the reading portion of the AIMS was reported. The author disaggregated the sample into demographic subgroups (i.e. ethnicity, gender, ELL status) to look for any differences. The group for which one of the highest correlations between the AIMS and the DIBELS-ORF were found was the ELL group ($n = 65$). Second only to the classification of Hispanic, the correlation for ELLs was .78 ($p < .01$). Interestingly, the group with the lowest correlation of .66 ($p < .01$, $n = 175$) was the non-ELL group.

As indicated previously, the Roehrig et al. (2008) study, with 20 percent of the sample classified as ELL, found language status (ELL or non-ELL) not to be a significant factor when predicting FCAT-SSS performance with the DIBELS-ORF. In contrast, Baker, Baker, Katz, and Otterstedt (2009) reported a significant difference in the DIBELS-ORF and SAT-10 correlations between ELLs and non-ELLs during first-grade. Concurrent correlations between the spring administrations of the DIBELS-ORF and the SAT-10 were .66 (ELL), and .75 (non-ELL).

Finally, the Riedel (2007) study is of particular importance to recent DIBELS-ORF research and to the current study. Of the 1,518 participants, only 59 were ELL first-graders. Results for the ELL and non-ELL students were looked at separately and compared. As mentioned previously, the first grade winter and spring administrations of the DIBELS-ORF were used to predict performance on the GRA+DE at the end of first-grade. Winter DIBELS-ORF correlations with the GRA+DE were .59 (non-ELL), and .72 (ELL). For the concurrent spring administration the correlations were .67 (non-ELL), and .80

(ELL). These results are in agreement with the Wilson (2005) study, and suggest that the DIBELS-ORF may be a better predictor for children learning English.

An important aspect of the Riedel (2007) study was a further investigation into why the DIBELS-ORF was not an accurate predictor of reading achievement test performance for some students. Approximately 15% of the sample that had satisfactory ORF scores in spring of first grade ended up doing poorly on the reading achievement test. Further analysis suggested that receptive vocabulary may be a significant factor. Students who did not achieve satisfactory performance on the reading achievement test scored an average of more than 20 normal curve equivalent points lower on the vocabulary subtest of the GRA+DE than students who achieved satisfactory performance. In addition, some students with poor DIBELS-ORF scores went on to perform well on the GRA+DE reading comprehension test. Such students had strong scores on the vocabulary subtest.

The relationship between vocabulary and oral reading fluency is of particular interest when it comes to predicting reading comprehension among ELLs. Samuels (2007) noted that students for which DIBELS-ORF is a poor predictor of comprehension are very often ELLs. ELL students can very quickly develop excellent decoding skills as they learn English, however English vocabulary knowledge develops slower (Manis, Lindsey, & Bailey, 2004). Thus ELLs are more often misidentified by the DIBELS-ORF as being at low-risk for reading failure, because even if decoding and pronunciation are smooth, comprehension may be minimal (Samuels, 2007; Wiley, & Deno, 2005).

Relationship between Vocabulary and Reading Comprehension

The report of the National Reading Panel (2000) identified vocabulary as one of five critical components for reading achievement. Indeed, an enduring finding in reading research is an established relationship between vocabulary knowledge and reading comprehension (Baumann, Kame‘enui, & Ash, 2003; Becker, 1977; Campbell, Bell, & Keith, 2001; Davis, 1942; Muter, Hulme, Snowling, & Stevenson, 2004; Nation & Snowling, 2004; NICHD Early Child Care Research Network, 2005; National Reading Panel, 2000; Snow, Tabors, Nicholson, & Kurland, 1995; Snow, Burns, & Griffin, 1998; Snowling, 2005; Stanovich, 1986; Storch & Whitehurst, 2002). Conceptually speaking, a strong relationship between reading comprehension and vocabulary is a logical assumption. Simply put, a large vocabulary is needed in order to comprehend written or spoken language. The larger the vocabulary a reader has, the more meaning that can be extracted from both contextual clues within the text, as well as from the individual words themselves. Young readers with a small vocabulary size often struggle to achieve comprehension (Hart & Risley, 2003).

Two major categories of vocabulary knowledge are receptive and expressive vocabulary. Receptive vocabulary refers to the words that a person is able to recognize when spoken or read. Expressive vocabulary consists of the words that a person is able to use when speaking or writing. Receptive vocabulary is a reflection of vocabulary breadth, while the expressive vocabulary is a better indication of vocabulary depth, or the depth of semantic understanding within the lexicon (Ouellette, 2006). The number of words to which an individual is able to

assign at least some meaning, when they are either read or heard, is typically larger than the number of words that are known well enough to use when speaking or writing (Kamil & Hiebert, 2005). When it comes to measuring vocabulary, “knowing a word is not an all-or-nothing proposition” (Beck & McKeown, 1991, p. 791). Rather, knowledge of word meaning can be viewed as a continuum that ranges in degree of understanding. In other words, instead of only having either a deep understanding, or absolutely no knowledge of a word’s meaning, a reader will often have a general sense, or a contextualized understanding of a word (Beck & McKeown, 1991; Henriksen, 1999; Rosenthal & Ehri, 2010).

The results of previous research indicate a moderate-to-strong relationship between receptive vocabulary and reading comprehension among ELLs and non-ELLs alike (Cotton & Crewther, 2009; Kim & Petscher, 2010; Lesaux & Kieffer, 2010; Llach & Gallego, 2009; Muter, Hulme, Snowling, & Stevenson, 2004; Snow, Tabors, Nicholson, and Kurland, 1995; Swansen, Rosston, Gerber, & Solari, 2008; Tannenbaum, Torgesen, & Wagner, 2006; Webb, 2009). However the strength of the relationship varies depending on factors such as age (Kim & Petscher, 2010; Tannenbaum, Torgesen, & Wagner, 2006; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997). For example Kim and Petscher (2010) observed that the relationship between receptive vocabulary and reading comprehension strengthened significantly from year to year. For first, second, and third-grade, the correlations between scores on the PPVT and the SAT-10 were

.40, .50, and .58, respectively. These results are supported by similar previous findings (Cotton & Crewther, 2009).

The Peabody Picture Vocabulary Test is a receptive vocabulary measure that has been commonly used for more than half a century. The PPVT-III (Dunn & Dunn, 1997) is the third edition of The PPVT, and has often been used in studies that involve ELLs (Uchikoshi, 2006). In recent years the PPVT has frequently been used as a predictor of reading comprehension. Among the studies found, the majority of concurrent correlations between the PPVT and various reading comprehension measures for second and third-grade, ranged between .45 and .58 (Cotton & Crewther, 2009; Crosson & Lasaux, 2010; Cutting & Scarborough, 2006; Kim & Petscher, 2010; Lervag & Aukrust, 2010; Lesaux & Kieffer, 2010; Swansen, Rosston, Gerber, & Solari, 2008; Tannenbaum, Torgesen, & Wagner, 2006).

Such findings warrant the further investigation of receptive vocabulary as a predictor of reading comprehension among ELLs. Moreover, these findings justify the goal of determining if it can explain any variance on reading comprehension measures beyond that explained by DIBELS-ORF.

Summary

Overall, the available literature has supported the DIBELS-ORF as a moderate-to-strong predictor of performance on standardized reading achievement tests measuring reading comprehension among English-speaking monolinguals. Very little research exists that has looked directly at that predictive relationship when it comes to children learning English. Some of the available

research suggests the relationship is weaker among ELLs (Baker, Baker, Katz, & Otterstedt, 2009), some assert that there is no difference (Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008), while still others indicate a stronger relationship may exist among ELLs compared to non-ELLs (Riedel, 2007; Wilson, 2005).

It is not clear why there are conflicting results in regard to the strength of the relationship between reading comprehension and the DIBELS-ORF among ELLs in comparison to non-ELLs. A possible source of variation could be differing English proficiency levels of the ELLs within the samples. None of the studies found for the current literature review took into account the impact that varying English proficiency levels among their ELL participants would have on correlations between the DIBELS-ORF and reading achievement tests. Most likely, the ELL participants were not homogenous in terms of English proficiency, and varied significantly.

Some non-ELL studies have observed that whereas the DIBELS-ORF is generally useful at predicting satisfactory or unsatisfactory performance on end of year reading achievement tests, there often remains a portion of students for which the DIBELS-ORF misidentifies as either at-risk, or low-risk with regard to performance on high-stakes reading achievement tests (Buck & Torgesen, 2003; Riedel, 2007; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Schilling, Carlisle, Scott, & Zeng, 2007; Wilson, 2005). Research emphasizes the need to continue to improve current approaches in identifying children learning to read who are at risk for reading failure.

Riedel (2007) identified vocabulary as a potentially important factor toward a better understanding of the relationship between reading fluency measures and reading comprehension measures. In Riedel's study, students with satisfactory DIBELS-ORF scores, but poor reading comprehension performance, had lower vocabulary scores than students with satisfactory DIBELS-ORF scores. Conversely, students with poor DIBELS-ORF scores who performed well on the reading comprehension test had high vocabulary scores. Interestingly, vocabulary is one of the five basic early literacy skills that have been identified as essential for successful reading achievement (National Reading Panel, 2000). The findings of Riedel (2007) suggest a need for further investigation into the potential contribution vocabulary can make in explaining variance in performance on reading achievement tests beyond that explained by the DIBELS-ORF.

Considering the widespread use of the DIBELS-ORF as an indicator of future performance on reading achievement tests, identifying a way to make the DIBELS-ORF a more accurate predictor of future high-stakes reading outcomes could have useful and practical implications for educators. If a significant amount of additional variance in reading achievement outcomes is contributed by receptive vocabulary scores, it would support the usefulness of considering receptive vocabulary skills alongside DIBELS-ORF performance when identifying whether or not children are at-risk for reading failure and in need of intervention.

Chapter 3

METHOD

Participants

Demographics. Participants were sixty-five Hispanic, Spanish-speaking, kindergarten children from a school district adjacent to the U.S.–Mexico border. This particular district consisted of a high percentage of ELL children immersed in a predominately Spanish-speaking community. Although Arizona state law mandates English-only instruction, it is likely that the classroom instruction is the only consistent English exposure that children living in this particular community have on a daily basis. All participating children were eligible for the Federal Free and Reduced Lunch Program.

Selection Criteria. The sixty-five participants were all of the kindergarteners in that district that performed at the “emergent” (no English) level of the Stanford English Language Proficiency Test [SELP], (Harcourt Assessment, 2004) upon entering kindergarten in 2005. The SELP was administered by the school district. None of the participating children had any preschool prior to kindergarten, and none were classified as special needs.

Procedure

In compliance with NCLB (2001), and state and federal mandates for schools receiving Reading First grant money, the school district directed the administration of the DIBELS-ORF, and the TerraNova-Reading test. The DIBELS-ORF is normally administered first in the spring of first-grade, and then each fall, winter, and spring of second and third grades. For consistency, only

scores from the spring administrations of the DIBELS-ORF were used for this study. The TerraNova is administered each spring starting in second grade. The Reading portion of the TerraNova (TerraNova-Reading) administrations from grades two and three were used for the current study. The Peabody Picture Vocabulary Test—Third Edition, Form IIIA (PPVT; Dunn & Dunn, 1997) was administered by the principal investigator for research conducted under a U.S. Department of Education Reading First grant, prior to the current study. The lead researcher monitored all administration sessions to ensure that all administration protocols were followed. The PPVT was administered in winter (January) of each year. Table A1 is an assessment calendar indicating when each assessment was administered at each grade level.

Instruments

Dynamic Indicators of Basic Early Literacy Skills: Oral Reading

Fluency (DIBELS-ORF). The Dynamic Indicators of Basic Early Literacy Skills: Oral Reading Fluency subtest (DIBELS-ORF), 6th Edition (DIBELS; Good & Kaminski, 2002) is an individually administered, standardized test of oral fluency with connected text. The sixth edition of the Dynamic Measurement Group's Technical Report (2008) provides an account of DIBELS-ORF concurrent and criterion-related validity with reading comprehension scores for elementary students that range between .60 and .90; alternate form reliability scores ranging between .89 and .94; and test-retest reliability estimates ranging from .92 to .97. Good and Kaminski (2002) explained the DIBELS-ORF administration procedures as follows:

Student performance is measured by having students read a passage aloud for one minute. Words omitted, substituted, and hesitations of more than three seconds are scored as errors. Words self-corrected within three seconds are scored as accurate. The number of correct words per minute from the passage is the oral reading fluency rate. (Good & Kaminski, 2002, p. 30)

Three different reading passages are used for the DIBELS-ORF administration. From the scores of the three administrations, the median score is recorded as the fluency score. There are DIBELS-ORF Benchmark goals for each grade level. The low-risk benchmark goals are 40 WCPM for first grade, 90 WCPM for second grade, and 110 WCPM for third grade. Students who achieve benchmark level WCPM or higher are considered low-risk for reading failure. Students with WCPM scores that fall below 20, 70, and 80 at grades one, two, and three, respectively, are considered at-risk and in need of extra instructional support if (Good & Kaminski, 2002).

TerraNova-Reading, Second Edition. The TerraNova is a nationally norm-referenced assessment that is “designed and developed to provide achievement test scores that are valid for most types of educational decision making” (CTB/McGraw Hill, 2001, p. 31). The TerraNova (second edition) was part of Arizona’s state mandated assessment program during the time of data collection for the current study. Only the reading portion (TerraNova-Reading), measuring reading comprehension, was used for the purposes of this study. The TerraNova-Reading Preliminary Technical Bulletin describes the test content as

being aligned with current instructional practices, and using authentic reading selections. Test results are provided in the following forms: local percentiles, national percentiles, national stanines, grade equivalents, normal curve equivalents, and a range of scaled scores that span all grade levels tested. The instrument generates a set of scores that are categorized by four performance levels: Advanced, Proficient, Basic, or Below Basic. Complete TerraNova battery internal consistency coefficients range from .67 to .91 (CTB/McGraw-Hill, 2003).

Peabody Picture Vocabulary Test (PPVT). The Peabody Picture Vocabulary Test—Third Edition, Form IIIA (PPVT; Dunn & Dunn, 1997) is a measure of receptive vocabulary. The PPVT has been nationally standardized. Split-half reliability coefficients included in the examiner manual ranged from .86 to .97 (Dunn & Dunn, 1997).

For PPVT administration, the child is shown a page with four pictures on it and asked to indicate which of the four represents the correct meaning of a word that is spoken orally by the test administrator. For example, the word ball is orally presented while simultaneously showing the child a page with pictures of a car, a baseball, a flower, and a hammer. The child is instructed to point to the picture that correctly represents the meaning of the spoken word. The PPVT progressively increases in difficulty until the child fails to correctly identify 8 words within a given set of 12. The total number of words correctly identified determines the score. The raw score is converted into a standard score that is based on age. The PPVT consists of 204 total items and is suitable for ages two-and-a-half years to adult.

Analyses of Data

Both descriptive and inferential analyses were utilized in this study. Means, standard deviations, and correlations were examined among all study variables. First, bivariate correlations were computed to estimate the strength of the relationship between the first through third grade DIBELS-ORF and the criterion TerraNova-Reading scores. Next, bivariate correlations between first-through third-grade DIBELS-ORF scores and first- through third-grade PPVT-III scores were conducted. To control for Type I error across the multiple bivariate correlations, the Holm (1979) sequential step-down application of the Bonferroni method was computed.

A multiple regression model was estimated to determine the amount of unique variance in second- and third-grade TerraNova-Reading scores that is explained by an oral reading fluency measure and a receptive vocabulary measure at each grade level (first- through third-grade). The predictor variables were entered into the regression analysis sequentially according to the research hypothesis that PPVT scores contribute to the prediction of TerraNova-Reading scores over and above DIBELS-ORF scores. In addition, F tests and t statistics were examined to evaluate the statistical significance of the model and individual predictors, respectively, at each grade level. Both the adjusted R^2 and R^2 change were examined as estimates of overall model effect size to compare the relative influence of each predictor in the model.

Chapter 4

RESULTS

Results provided in this chapter address the following questions regarding young Spanish-speaking ELLs learning to read: (1) What is the correlation between performance on the second-grade TerraNova-Reading test and DIBELS-ORF scores from grades one and two, respectively? (2) What is the correlation between performance on the third-grade TerraNova-Reading test and DIBELS-ORF scores from grades one, two, and three, respectively? and (3) Will the PPVT explain any additional variance on second- and third-grade TerraNova performance, beyond that explained by the DIBELS-ORF, at each grade level? Results from descriptive statistics, bivariate correlation analyses, and regression analyses will be presented in this chapter.

Descriptive Statistics

Means and standard deviations for all measures at each grade level are shown in Table A2. The mean for the DIBELS-ORF scores indicate that the participants, as a group, met the benchmark at each grade level. Distributions are depicted by a histogram for each measure (see Figures A1 through A8). The score distributions for most of the measures are unimodal and reasonably normal. The first-grade DIBELS-ORF distribution of scores is the most skewed, showing a moderate positive skew (see Figure A1 and Table A2). Histograms of score distributions and skew statistics also suggest slightly skewed distributions among the rest of the measures, though they are not markedly non-normal (see Table A2 for skew statistics). The first-grade PPVT scores suggest a bimodal distribution

(see Figure A6), showing a slight clustering of vocabulary scores at two levels during first-grade.

A scatterplot matrix is also provided to depict bivariate linear relationships among all pairs of variables (see Figure A9). The scatterplots suggest a positive linear relationship between the DIBELS-ORF scores (grades one through three) and the dependent variables (second- and third-grade TerraNova-Reading scores). However, upon close inspection of the first-grade DIBELS-ORF scores against the dependent measures, the wide scatter of the data points, along with the flatter slope of the regression line, suggest that the linear relationship may be weak. The scatterplots depicting second- and third-grade DIBELS-ORF scores suggest a much stronger linear relationship with the dependent measures.

Correlation Analyses

Correlations among all measures were calculated and are presented in a correlation matrix (see Table A3). Looking at the strength of the relationship between DIBELS-ORF scores and performance on the TerraNova-Reading tests, correlations range from .30 (first-grade DIBELS-ORF with second-grade TerraNova-Reading) to .68 (third-grade DIBELS-ORF with third-grade TerraNova). The first-grade DIBELS-ORF has a higher correlation with the third-grade TerraNova-Reading ($r = .40, p < .01$) than it does with the second-grade TerraNova-Reading ($r = .30, p < .05$). Similarly, the second-grade DIBELS-ORF has a stronger relationship with the third-grade TerraNova-Reading ($r = .66, p < .01$) than it does with the concurrent second-grade TerraNova-Reading ($r = .48,$

$p < .01$). The concurrent correlation between the third-grade DIBELS-ORF and the third-grade TerraNova-Reading was .68 ($p < .001$). With regard to identifying the DIBELS-ORF administrations with the strongest relationship with TerraNova-Reading performance in second- and third-grade, the concurrent administrations yielded the strongest correlations. However, the moderate-to-strong relationships (Cohen, 1988) between the TerraNova-Reading administrations and DIBELS-ORF scores from preceding years are significant (see Table A3). For example, the second-grade DIBELS-ORF is nearly as strong of a predictor of the third-grade TerraNova-Reading ($r = .66, p < .01$) as the concurrent third-grade DIBELS-ORF is ($r = .68, p < .01$).

Multiple Regression Analyses

The next question in the current study was to determine whether receptive vocabulary scores would explain any variance on second- and third-grade TerraNova-Reading performance beyond that explained by DIBELS-ORF scores, and thus provide additional utility for predicting reading comprehension. Linear regression models consisted of DIBELS-ORF and PPVT scores at each grade level as predictors of performance on the TerraNova-Reading test administered at grades two and three. The predictor variables were entered into the regression analysis sequentially according to the research hypothesis that PPVT scores contribute to the prediction of TerraNova-Reading scores over and above DIBELS-ORF scores.

To ensure that the assumptions for multiple linear regression were met, an a priori visual inspection of scatterplots of the residuals was conducted and it

revealed no patterns in the plotted errors. Correlations between the predictor variables (DIBELS-ORF, and PPVT) were also examined to check for potential multicollinearity issues (see Table A3). Only the correlations that involved the third-grade DIBELS-ORF were significant, and the highest was .37 ($p < .001$), thus it was determined that multicollinearity was not a major concern.

DIBELS-ORF and PPVT predicting second-grade TerraNova-Reading. A multiple regression analysis was conducted, entering the first-grade DIBELS-ORF and PPVT scores as predictors of second-grade TerraNova-Reading scores (see Tables A4 and A5 for regression results). The DIBELS-ORF alone accounted for a small, but significant seven percent of the variance in second-grade TerraNova-Reading scores, $R^2_{adj.} = .074$, $F(1, 52) = 5.24$, $MSE = .926$, $p < .05$. Alone the first-grade PPVT accounted for nearly 52 percent of the explained variance in the second-grade TerraNova-Reading scores, $R^2_{adj.} = .519$, $F(1, 52) = 58.10$, $MSE = .481$, $p < .001$. Upon adding PPVT scores to DIBELS-ORF scores in the regression model, R^2 increased by 48 percent, $\Delta F(1, 51) = 55.91$, $p < .001$, for a total of 55 percent of the explained variance accounted for by the full model, $R^2_{adj.} = .550$, $F(2, 51) = 33.34$, $MSE = .450$, $p < .001$. Beta weights were statistically significant for both predictors in the full model, as indicated by t -statistics (see Table A4). These results indicate the significant predictive utility of the PPVT beyond that accounted for by DIBELS-ORF in the first grade.

Next, the same procedure was followed with second-grade DIBELS-ORF and PPVT scores entered sequentially as predictors of second-grade TerraNova-

Reading scores in a multiple regression analysis (see Tables A4 and A5 for model results). Alone, the second-grade DIBELS-ORF accounted for 22 percent of the variance in the TerraNova-Reading scores, $R^2_{\text{adj.}} = .217$, $F(1, 51) = 15.44$, $MSE = .783$, $p < .001$. The second-grade PPVT, alone, accounted for 47 percent of the explained variance, $R^2_{\text{adj.}} = .474$, $F(1, 51) = 47.95$, $p < .001$. Adding PPVT scores to the DIBELS-ORF scores in the regression model, increased R^2 by 35 percent, $\Delta F(1, 50) = 42.50$, $p < .001$.

Again, the second-grade PPVT seems to account for a large proportion of the explained variance in second-grade TerraNova-Reading scores after controlling for the second-grade DIBELS-ORF. However the proportion of variance explained by the second-grade DIBELS-ORF, while still smaller, is statistically significant and larger than for the previous model in which first-grade scores were used to predict second-grade TerraNova-Reading. Beta weights were statistically significant for both predictors in the full model, as indicated by t -statistics (see Table A4). The full second-grade model explained about 57 percent of the explained variance in second-grade TerraNova-Reading scores, $R^2_{\text{adj.}} = .568$, $F(1, 50) = 35.25$, $MSE = .432$, $p < .001$.

DIBELS-ORF and PPVT predicting third-grade TerraNova-Reading.

Next, third-grade TerraNova-Reading scores were predicted by first- through third-grade DIBELS-ORF and PPVT scores entered sequentially in three respective multiple regression analyses (see Tables A6 and A7 for regression results). Using first-grade scores, the DIBELS-ORF accounted for about 14 percent of the variance explained in the third-grade TerraNova-Reading scores,

$R^2_{\text{adj.}} = .139$, $F(1, 48) = 8.89$, $MSE = .861$, $p < .01$. Alone, the first-grade PPVT accounted for 32 percent of the explained variance in third-grade TerraNova-Reading scores, $R^2_{\text{adj.}} = .321$, $F(1, 48) = 24.15$, $MSE = .679$, $p < .001$. The first-grade PPVT scores explained an additional 27 percent of the variance in TerraNova-Reading scores beyond that explained by the first-grade DIBELS-ORF, $\Delta F(1, 47) = 22.91$, $p < .001$. Supporting the results of the correlation analysis, the results of the regression analysis suggest the first-grade DIBELS-ORF to be a stronger predictor of third-grade reading comprehension than it is of second-grade reading comprehension. The full model accounted for over 40 percent of the explained variance in third-grade TerraNova-Reading scores, $R^2_{\text{adj.}} = .409$, $F(2, 47) = 17.93$, $MSE = .591$, $p < .001$. Beta weights were statistically significant for both predictors, as indicated by t -statistics (see Table A6).

In comparison to models with first-grade scores as predictors, the predictive utility of the second-grade DIBELS-ORF was much higher, $R^2_{\text{adj.}} = .429$, $F(1, 48) = 37.74$, $MSE = .571$, $p < .001$, with approximately 43 percent of the variance in the third-grade TerraNova-Reading explained by the DIBELS-ORF. Alone, the second-grade PPVT accounted for 35 percent of the explained variance, $R^2_{\text{adj.}} = .351$, $F(1, 48) = 27.54$, $p < .001$. PPVT scores explained an additional 20 percent of the variance beyond that explained by the second-grade DIBELS-ORF, $\Delta F(1, 47) = 26.87$, $p < .001$, with 63 percent of the variance in TerraNova-Reading scores explained by the full model, $R^2_{\text{adj.}} = .629$, $F(2, 47) = 42.48$, $MSE = .371$, $p < .001$. Beta weights were statistically significant for both predictors in the full model, as indicated by t -statistics (see Table A6).

In the final regression analysis, third-grade DIBELS-ORF scores accounted for 45 percent of the explained variance in third-grade TerraNova-Reading scores, $R^2_{\text{adj.}} = .453$, $F(1, 47) = 40.71$, $MSE = .547$, $p < .001$. The third-grade PPVT alone accounted for 35 percent of the explained variance, $R^2_{\text{adj.}} = .348$, $F(1, 47) = 26.612$, $MSE = .652$, $p < .001$. In the full model, third-grade PPVT scores accounted for an additional 15 percent of the variance beyond that explained by the third-grade DIBELS-ORF, $\Delta F(1, 46) = 18.15$, $p < .001$, with just under 60 percent of the variance in third-grade TerraNova-Reading scores explained by the full model, $R^2_{\text{adj.}} = .599$, $F(2, 46) = 36.86$, $MSE = .401$, $p < .001$. This is slightly smaller than the amount explained by second-grade DIBELS-ORF and PPVT scores. Beta weights were statistically significant for both predictors in the full model, as indicated by t -statistics (see Table A6).

Chapter 5

DISCUSSION

Recent U.S. federal-state education initiatives have brought about a heightened focus on assessing and monitoring early literacy skill development in young children. Across the U.S., reading proficiency is determined by performance on state-mandated reading achievement tests that include measures of reading comprehension. As a part of scientifically-based reading programs in many states and school districts, the DIBELS-ORF is currently being administered to both monolingual English-speaking children and ELL children alike as a probe for monitoring development toward achieving reading comprehension. While studies exploring the usefulness of the DIBELS-ORF in predicting performance on state mandated reading achievement tests among mono-lingual English-speaking children have become abundant over recent years, it is still unclear whether or not the results are representative of the ELL population (Klingner & Edwards, 2006).

While a growing research base supports a strong predictive relationship between the DIBELS-ORF and measures of reading comprehension (Buck & Torgesen, 2003; Good, Simmons, & Kame'enui, 2001; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Schilling, Carlisle, Scott, & Zeng, 2007; Shapiro, Solari, & Petscher, 2008; Vander Meer, Lentz, & Stoller, 2005; Wood, 2006), a modicum of research suggests that the DIBELS-ORF may not be a very accurate predictor of reading comprehension outcomes for some children (Riedel, 2007; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Schilling, Carlisle, Scott,

& Zeng, 2007; Wilson, 2005). One recent study suggested that adding vocabulary scores to a predictive model, along with DIBELS-ORF scores, might add a significant amount of explained variance in reading comprehension scores (Riedel, 2007).

Against this background, the current study sought to answer the following questions: (1) What is the correlation between performance on the second-grade TerraNova-Reading test and DIBELS-ORF scores from grades one and two, respectively? (2) What is the correlation between performance on the third-grade TerraNova-Reading test and DIBELS-ORF scores from grades one, two, and three, respectively? and (3) Will the PPVT explain any additional variance on second- and third-grade TerraNova-Reading performance, beyond that explained by the DIBELS-ORF, at each grade level?

Results indicated statistically significant positive correlations between TerraNova-Reading and the DIBELS-ORF at each grade level. The general trend of the correlations suggests that the relationship becomes stronger with each successive grade. This represents a departure from earlier studies suggesting a weakening relationship between reading fluency and reading comprehension measures as reading ability improves (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Schilling, Carlisle, Scott, & Zeng, 2007). However the samples in previous studies consisted mostly of English-speaking monolinguals.

What is the correlation between performance on the second-grade TerraNova-Reading test and DIBELS-ORF scores from grades one and two,

respectively? To answer question one, bivariate correlations were calculated to evaluate the strength of the relationship between DIBELS-ORF scores from first- and second-grade and reading comprehension scores as measured by the second-grade TerraNova-Reading achievement test. A correlation of .30 was found between the first-grade DIBELS-ORF and the second-grade TerraNova-Reading. As previously noted, Riedel (2007), and Baker et al. (2009) predicted second-grade reading comprehension outcomes using first-grade DIBELS-ORF scores with samples of English-speaking monolinguals. The reported correlations with first-grade DIBELS-ORF were .54 (TerraNova) and .72 (SAT-10), respectively. Although there is a substantial spread between the correlation values reported in the non-ELL studies, the notion of a weaker relationship existing between DIBELS-ORF and reading comprehension among ELLs, compared to non-ELLs at this grade level is supported.

Not surprisingly, results of the current study indicate that second-grade TerraNova-Reading scores have a stronger relationship with the second-grade DIBELS-ORF (.48) than with the first-grade DIBELS-ORF (.30). However, the concurrent second-grade correlation of .48 is much lower than the respective concurrent second-grade correlations of .75 and .80 reported by Schilling et al. (2007) and Baker et al. (2008) in studies consisting mostly of non-ELLs. Thus, results of the present study further suggest that the relationship between DIBELS-ORF scores and reading comprehension outcomes may be weaker among ELLs as compared to non-ELLs before third-grade. These results also have theoretical support from Baker, Baker, Katz, and Otterstedt (2009).

What is the correlation between performance on the third-grade TerraNova-Reading test and DIBELS-ORF scores from grades one, two, and three? Looking at first-grade DIBELS-ORF predicting third-grade TerraNova-Reading, the present study found a correlation of .40. This is a very interesting finding, suggesting that the first-grade DIBELS-ORF is actually a better predictor of third-grade reading comprehension than it is of second-grade reading comprehension. However, the correlation of .40 is much weaker than correlations reported by Wanzek and colleagues (2010) in a similar comparison (first-grade DIBELS-ORF predicting third-grade reading comprehension scores) among non-ELLs. Correlations of .64 (SAT-10), and .51 (TAKS) were reported.

The correlation between second-grade DIBELS-ORF and third-grade TerraNova-Reading was .66, a correlation stronger than the concurrent second-grade relationship, and nearly as strong as the concurrent third-grade relationship. This finding may have some useful implications for early identification of second-grade ELL students who are at-risk for reading failure the next year (third-grade). In two non-ELL studies, Wanzek et al. (2010), and Baker et al. (2008) reported similar second-grade DIBELS-ORF relationships with the third-grade SAT-10 (.68), and the third-grade OSRA (.63), respectively.

Although a weaker relationship between DIBELS-ORF scores and reading comprehension outcomes seems evident among ELLs as compared to non-ELLs during grade one and possibly grade two (concurrent second grade correlation), the concurrent third-grade correlation of .68 found in the current study suggests that by third-grade the relationship is very similar between the two groups. Most

concurrent third-grade correlations reported in previous studies have ranged between .65 and .70 (Buck & Torgesen, 2003; Pearce & Gayle, 2009; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Schilling, Carlisle, Scott, & Zeng, 2007; Wanzek, Roberts, & Linan-Thompson, 2010).

As discussed in the review of literature, there is not a clear explanation for conflicting results in regard to the strength of the relationship between reading comprehension and the DIBELS-ORF among ELLs in comparison to non-ELLs. A possible source of variation could be differing English proficiency levels of the ELLs within the samples. Another contributing factor could be a difference in how the relationship changes over time among ELLs compared to non-ELLs. Results from the current study suggest a weak relationship exists for ELLs during grade one and possibly grade two, but then becomes comparatively stronger by grade three.

Will a receptive vocabulary measure (PPVT-III) explain any additional variance on TerraNova-Reading performance beyond that explained by the DIBELS-ORF? Linear multiple regression analyses were conducted to determine whether or not a receptive vocabulary measure (PPVT) would explain any additional variance on second- and third-grade TerraNova-Reading performance, beyond that explained by the DIBELS-ORF, at each grade level. The full first-grade model, predicting second-grade TerraNova-Reading scores, accounted for 55 percent of the explained variance. The PPVT was, by far, the stronger factor in the analysis. Alone, PPVT scores accounted for about 50 percent of the explained variance in the TerraNova-Reading, while the DIBELS-

ORF accounted for only 7 percent. These results, coupled with the first-grade correlation results, emphasize the need for caution when using first-grade DIBELS-ORF results as a basis for determining whether or not ELL students are on track for reading proficiency by third-grade.

In the full second-grade model, DIBELS-ORF and PPVT scores together accounted for 57 percent of the explained variance in second-grade TerraNova-Reading scores. This is almost exactly the same amount of variance explained by the first-grade model. However, this time the DIBELS-ORF, alone, accounted for about 22 percent, while the PPVT accounted for 47 percent alone. It is evident that while the predictive utility of the DIBELS-ORF has gained some ground by second-grade, the PPVT is still dominating the predictive relationship with the TerraNova-Reading. However, it is interesting that the second-grade PPVT is a little less predictive of second-grade TerraNova-Reading than the first-grade PPVT.

For the prediction of third-grade TerraNova-Reading scores, the full first-grade model accounted for 41 percent of the explained variance. Alone, the first-grade DIBELS-ORF accounted for 14 percent, and the first-grade PPVT accounted for 32 percent. This finding has useful implications. The PPVT, as early as first-grade, lends a relatively large amount of predictive utility toward the accurate identification of ELL students who are at-risk for reading failure two years later.

The full second-grade model proved to be the strongest of all. Together the second-grade DIBELS-ORF and second-grade PPVT accounted for 63 percent

of the explained variance in the third-grade TerraNova-Reading. Interestingly, it appears that by second-grade, the DIBELS-ORF and PPVT switch places in terms of dominance in the predictive relationship (predicting third-grade TerraNova-Reading). Alone the second-grade DIBELS-ORF accounted for 43 percent of the explained variance, while the second-grade PPVT is around 35 percent.

Whereas the second-grade model appears to be the strongest predictor of third-grade TerraNova-Reading performance, not surprisingly the full third-grade model was also strong, accounting for 60 percent of the explained variance in the third-grade TerraNova-Reading scores. In this final model, the third-grade DIBELS-ORF, alone, accounted for over 45 percent of the explained variance, while the third-grade PPVT seemed to level off at 35 percent (same as in the second-grade model).

The results of the current study provide support for the findings of Riedel (2007), and emphasize the importance of vocabulary knowledge in predicting reading comprehension achievement, especially during grades one and two. Students for which DIBELS-ORF may be a poor predictor of comprehension are very often ELLs (Samuels, 2007). While learning English, ELL students often develop excellent decoding skills; however comprehension may be minimal because English vocabulary knowledge develops more slowly (Manis, Lindsey, & Bailey, 2004). In this way ELLs are likely to be misidentified by the DIBELS-ORF as being at low-risk for reading failure (Samuels, 2007; Wiley, & Deno, 2005).

It is interesting to note the strength of the relationship between the DIBELS-ORF and the PPVT, respectively, with the TerraNova-Reading, varied across grade levels. Beginning in first-grade, the PPVT accounted for a large proportion of the variance explained in both the second- and third-grade TerraNova-Reading, while the first-grade DIBELS-ORF was found to be a relatively weak predictor. However, over the subsequent two years, the DIBELS-ORF scores become a stronger predictor of reading comprehension while, inversely, the PPVT became a weaker predictor. By third-grade, DIBELS-ORF scores were a stronger predictor of TerraNova-Reading performance than the PPVT.

The observed change in the relative predictive ability of the DIBELS-ORF and the PPVT across grade levels may have important implications for more accurately determining, before third-grade, whether or not students are at risk for reading failure. For example, although the mean for the first-grade DIBELS-ORF scores in the current study exceeded the recommended benchmark, the first-grade DIBELS-ORF was found to be a relatively weak predictor of both second- and third-grade TerraNova-Reading scores. This suggests that there may be a significant danger of incorrectly determining risk status for reading failure among first-grade ELL students when it is based solely on DIBELS-ORF performance. In other words, reaching benchmark on the DIBELS-ORF in first-grade may not necessarily be a good indication that a student is on track for successful reading comprehension performance by third-grade. Results of the current study suggest that the implementation of a vocabulary measure (e.g. PPVT), in conjunction with

administration of the DIBELS-ORF during first-grade, could significantly boost the overall accuracy or reading comprehension risk status projections. To the point, strategic implementation of a vocabulary measure may catch the ELL first-graders determined to be reading with automaticity (as measured by the DIBELS-ORF), but at an unsatisfactory level of comprehension.

Limitations

While the current study makes a contribution to the research on the ability of DIBELS-ORF to predict reading comprehension among ELL children learning to read, there are several caveats. The sample of participants for this study was drawn from one small school district near the U.S.-Mexico border. Additionally, the limited size of the current study was due in part to the restriction of including only young ELLs who scored at the “emergent” level on the Stanford English Language Proficiency test (SELP). Hence, the results from such a small sample size in one geographical location may not reflect the true relationship between the variables represented in the general ELL population. Further research that extends to other geographical locations and considers variations in socio-economic status, and different English proficiency levels, is necessary to help generalize the findings. Further, the results of the current study are limited to Spanish-speaking ELLs, and may not generalize to ELLs who speak other languages or who have different cultural backgrounds.

The interpretations of the results in this study are, of course, limited to the individual assessments used to measure the constructs. The TerraNova-Reading test is only one of many tests currently being used to measure reading

comprehension in the U.S. The predictive accuracy of the DIBELS-ORF still needs to be examined using a variety of other reading comprehension measures, within the ELL population, to see if similar relationships emerge. Also, the PPVT is only one measure of receptive vocabulary. Results of the current study may not generalize to other measures of vocabulary knowledge. Finally, the DIBELS-ORF scores used for the present study were limited to the spring administrations of the DIBELS-ORF during grades one, two, and three. The DIBELS-ORF is also administered during the fall and winter of the school year (first-grade excluded), and is often administered on into grades four, five, and six.

Directions for Future Research

There are several areas that need further consideration in future research. Starting with the need for more ELL/DIBELS-ORF research in general, DIBELS-ORF predictive studies have primarily been focused on monolingual English-speaking students. Very little empirical data exists indicating whether or not the results of those studies are reflective of the ELL population (Baker, Baker, Katz, & Otterstedt, 2009; Riedel, 2007; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Wilson, 2005). Given the current testing and accountability climate in the U.S., and that the school-aged ELL population enrolled in U.S. public schools is growing at an accelerated rate (Kindler, 2002; National Clearinghouse for English Language Acquisition, 2009), it is critically important for future research to continue examining the usefulness of DIBELS-ORF in predicting performance on state mandated standardized reading achievement tests among ELLs.

Of the existing studies examining the usefulness of DIBELS-ORF to predict reading comprehension performance, the focus has primarily been on concurrent third grade relationships. While the current study examined first-through third-grade, more research is needed that examines DIBELS-ORF performance in grades both prior and subsequent to third-grade (among ELLs and non-ELLs alike). Also, the current study examined only the relationship of the spring DIBELS-ORF administrations with reading comprehension performance. Further research is needed to evaluate the relationship with the fall and winter administrations of the DIBELS-ORF as well.

It is important to keep in mind that ELLs are not a homogenous group in terms of level of English proficiency. The amount of exposure to the English language may be an important contributing factor in how quickly ELLs develop early literacy skills in English and gain reading success. For example, there may be a difference in emphasis placed on development of individual early literacy skills (e.g. phonics, vocabulary, text comprehension, and fluency), that is related to level of English language proficiency. While the participants of the current study all performed at the emergent level of an English proficiency test upon entering kindergarten, future research should include an analysis on the level of English language proficiency and its effect on how predictive the DIBELS-ORF will be of reading comprehension.

Of course, not all ELLs are Spanish-speaking ELLs. Thus, another question to consider for future research is: Do the results of the current study generalize to ELLs who are not Spanish-speaking ELLs? Studies examining the

relationship between the DIBELS-ORF and reading comprehension performance with an array of different types of ELLs are needed to see if similar relationships are found.

Another direction for future research is further examination of the role of vocabulary in predicting reading comprehension performance among ELLs. More research is needed to better understand how these constructs are related and to determine if implementation of measures of vocabulary knowledge, in conjunction with the DIBELS-ORF, is a viable direction to go for more accurate predictions of reading achievement among ELLs learning to read. The results of the current study are based on a small sample of ELLs, thus replication using larger samples of ELLs from various regions of the U.S. is important to determine how representative the results of the current study are of the ELL population in general.

Conclusion

The results of the present study found positive, moderate-to-strong relationships between oral reading fluency measure (DIBELS-ORF) and reading comprehension. These findings suggest that the relationship between oral reading fluency and performance on standardized and high-stakes measures of reading comprehension may be different among ELLs as compared to non-ELLs during first-grade and second-grade. Among the first- through third-grade administrations of the DIBELS-ORF, the strongest predictors of second- and third-grade reading comprehension outcomes were the concurrent administrations. The weakest predictor was the first-grade administration. However, by second-

grade the DIBELS-ORF is a moderately strong predictor of third-grade reading comprehension among ELLs. The relationship between the DIBELS-ORF and reading comprehension in third-grade approximates the correlations reported in previous studies that represent the non-ELL population. However, the weaker first- and second-grade concurrent correlations found indicate that the relationships reported in previous non-ELL studies may not be representative of ELLs.

In the current study, receptive vocabulary was not only the stronger predictor of reading comprehension among ELLs, but largely overshadowed the predictive ability of the DIBELS-ORF during first-grade (and even on into second-grade). Further, considering the large amount of variance in reading comprehension that it accounts for, vocabulary may be an important factor toward understanding why the DIBELS-ORF is not always an accurate predictor of reading comprehension performance for some children (often ELLs). While further research is necessary, the results of the present study suggest that using a measure of vocabulary (such as the PPVT), in conjunction with the DIBELS-ORF, substantially increases the accuracy of predicting reading comprehension performance (as measured by state mandated reading assessments), and may be a useful tool to help educators accurately identify beginning readers who are at risk for third-grade reading failure as early as first-grade. Measures of vocabulary knowledge may also help to identify those beginning ELL readers who decode with accuracy and speed on the DIBELS-ORF measure, but apparently comprehend very little. More research is necessary to further explore how these

constructs are related using larger samples of ELLs from various regions of the U.S.

In sum, as a result of recent U.S. federal-state education initiatives, children learning to read are being tested early and often to monitor early literacy skill development on the road to reading comprehension achievement. It is important to implement practical and valid measures that accurately predict reading achievement, as measured by state mandated tests, as early as possible. Much research is still necessary in order to better understand the predictive utility of such measures among ELLs. There is some question as to whether the DIBELS-ORF is as predictive of reading comprehension achievement among ELLs as it is among non-ELLs. Results of the current study suggest that the relationship may be comparatively weaker among ELLs early on before third-grade. However a receptive vocabulary measure (PPVT) accounts for a large and significant amount of explainable variance in reading comprehension (TerraNova-Reading) beyond that explained by the DIBELS-ORF. More research is necessary to determine if implementation of vocabulary knowledge measures, in conjunction with the DIBELS-ORF, would be a useful direction to go for more accurate predictions of reading achievement among ELLs learning to read.

REFERENCES

- Abedi, J., & Dietel, R. (2004). Challenges in the no child left behind act for english-language learners. *Phi Delta Kappan*, 85, 782-785.
- Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Allington, R. (1983). Fluency: The neglected reading goal. *The Reading Teacher*, 36, 556-561.
- August, D., & Hakuta, K. (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.
- Arizona Revised Statutes, Title 15 (Education), Section 3.1 (English Language Education for Children in Public Schools), 751-756.01.
- Baker, D., Baker, S., Katz, R., & Otterstedt, J. (2009). *The English reading performance of English learners in Oregon Reading First across three years of implementation (2003-2006)*. Technical Report: The Oregon Reading First Center. Retrieved January 23, 2011, from http://oregonreadingfirst.uoregon.edu/downloads/other/three_year_ell_report.pdf
- Baker, S., Smolkowski, K., Katz, R., Fien, H., Seeley, J., Kame'enui, E. & Beck, C.T. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review*, 37(1), 18-37.
- Baumann, J., Kame'enui, E., & Ash, G. (2003). Research on vocabulary instruction: Voltaire redux. In J. Flood, D. Lapp, J. R. Squire, & J. M. Jensen (Eds.), *Handbook on research on teaching the English language arts* (2nd ed., pp. 752-785). Mahwah, NJ: Erlbaum.
- Beck, I., & McKeown, M. (1991). Conditions of vocabulary acquisition. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of Reading Research*, Volume II (pp. 789-814). New York: Longman.
- Becker, W. (1977). Teaching reading and language to the disadvantaged – What we have learned from field research. *Harvard Educational Review*, 47, 518-543.
- Berent, I., & Perfetti, C. (1995). A rose is a REEZ: The two-cycles model of phonology assembly in reading English. *Psychological Review*, 102, 146-184.

- Berninger, V., Abbott, R., Trivedi, P., Olson, E., Gould, L., Hiramatsu, S., Holsinger, M., McShane, M., Murphy, H., Norton, J., Boyd, A., Westhaggen, S. (2010). Applying the multiple dimensions of reading fluency to assessment and instruction. *Journal of Psychoeducational Assessment, 28*, 3–18.
- Birch, B. (2002). *English L2 reading: Getting to the bottom*. Mahwah, NJ: Erlbaum.
- Bookheimer, S., Zeffiro, T., Blaxton, T., Gaillard, W., and Theodore, W. (1995). Regional cerebral blood flow during object naming and word reading. *Human Brain Map, 3*, 93–106.
- Boyle, A., Taylor, J., Hurlburt, S., & Soga, K. (2010). Title III accountability: Behind the numbers. U. S. Department of Education. Retrieved from <http://www2.ed.gov/rschstat/eval/title-iii/behindnumbers.pdf>
- Breznitz, Z. (1987). Increasing first graders' reading accuracy and comprehension by accelerating their reading rates. *Journal of Educational Psychology, 79*, 236–242.
- Buck, J., & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* (FCRR Tech. Rep. No. 1). Tallahassee: Florida Center for Reading Research. Available: www.fcrr.org/TechnicalReports/TechnicalReport1.pdf
- Campbell, J., Bell, S., Keith, L. (2001). Concurrent validity of the Peabody Picture Vocabulary Test-Third Edition as an intelligence and achievement screener for low SES African American children. *Assessment, 8*(1), 85-94.
- Capps, R., Fix, M., Murray, J., Ost, J., Passel, J., & Herwantoro, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Washington, DC: Urban Institute.
- Carlisle, J., Beeman, M., Davis, L., & Spharim, G. (1999). Relationship of meta-linguistic capabilities and reading achievement for children who are becoming bilingual. *Applied Psycholinguistics, 20*, 459-478.
- Chaney, C. (1998). Preschool language and meta-linguistic skills are links to reading success. *Applied Psycholinguistics, 19*(3), 433-46.
- Chomsky, C. (1976). After decoding: What? *Language Arts, 53* (3), 288-296.

- Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum.
- Committee on the Prevention of Reading Difficulties in Young Children, the Commission on Behavioral and Social Sciences and Education, and the National Research Council. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Cotton, S., & Crewther, S. (2009). The relationship between reading and intelligence in primary school aged children: implications for definitional models of dyslexia. *The Open Education Journal*, 2, 42-50.
- Cutting, L., & Scarborough, H. (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(3), 277–299.
- Crosson, A., & Lesaux, N. (2010). Revisiting assumptions about the relationship of fluent reading to comprehension: Spanish-speakers’ text-reading fluency in English. *Reading and Writing*, 23(5), 475-494.
- CTB/McGraw-Hill. (2001). *TerraNova technical report*. Monterey, CA: Author.
- CTB/McGraw-Hill. (2003). *TerraNova second edition: California Achievement Tests technical report*. Monterey, CA: Author.
- Daane, M., Campbell, J., Grigg, W., Goodman, M., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading* (NCES 2006-469). U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office.
- Dahl, P., & Samuels, S. (1974). *A mastery based experimental program for teaching poor readers high speed word recognition skills*. Unpublished manuscript, University of Minnesota, Minneapolis.
- Davis, F. (1942). Two new measures of reading ability. *Journal of Educational Psychology*, 33, 365–372.
- Davenport, D. (2008). Baseline study of Arizona’s English language learner programs and data Fiscal Year 2007. Office of the Auditor General. State of Arizona.
- de Jong, P. F., & Share, D. L. (2007). Orthographic learning during oral and silent reading. *Scientific Studies of Reading*, 11, 55–71.

- Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219–232.
- Deno, S. (2002). Problem solving as “best practice.” In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 37-55). Bethesda, MD: The National Association of School Psychologists.
- Deno, S., Mirkin, P., Chiang, B., & Lowry, L. (1980). *Relationships among simple measures of reading and performance on standardized achievement tests* (Res. Rep. No. 20). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.
- Deno, S., Marston, D., Shinn, M., & Tindal, G. (1983). Oral reading fluency: A simple datum for scaling reading disability. *Topics in Learning and Learning Disabilities, 2*(4), 53–59.
- Deschenes, S., Cuban, L., & Tyack, D. B. (2001). Mismatch: Historical perspectives on schools and students who don't fit them. *Teachers College Record, 103*(4), 525-547.
- Dowhower, S. (1987). Effect of repeated reading on second grade transitional readers' fluency and comprehension. *Reading Research Quarterly, 22*, 389–406.
- Dowhower, S. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory Into Practice, 30*, 158-164.
- Dunn, L., & Dunn, L. (1997). Peabody Picture Vocabulary Test-Third Edition. Circle Pines: American Guidance Service.
- Dynamic Measurement Group (2008). *DIBELS 6th Edition Technical Adequacy Information* (Tech. Rep. No. 6). Eugene, OR: Retrieved from <http://dibels.org/pubs.html>
- Ericson, L. & Juliebo, M. (1998). *The phonological awareness handbook for kindergarten and primary teachers*. Newark, DE: International Reading Association.
- Fuchs, L., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*(1), 45–58.
- Fuchs, L., Fuchs, D., & Deno, S. (1982). Reliability and validity of curriculum-based informal reading inventories. *Reading Research Quarterly, 18*, 6–26.

- Fuchs, L., Fuchs, D., Hamlett, C., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth should we expect? *School Psychology Review, 22*, 27–48.
- Fuchs, L., Fuchs, D., Hosp, M., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239–259.
- Fuchs, L., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*(2), 20–28.
- Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123*, 71–99.
- Good, R., & Kaminski, R. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th Ed.). Eugene, OR: Institute for the Development of Education Achievement. Retrieved from <http://www.dibels.uoregon.edu/>
- Good, R., & Kaminski, R. (Eds.). (2002). DIBELS administration and scoring guide. Retrieved from <http://www.dibels.uoregon.edu/>
- Good, R., Simmons, D., & Kame'enui, E. (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*, 257–288.
- Goodman, K. (2006). A critical review of DIBELS. In K. Goodman (Ed.), *The truth about DIBELS: What it is, what it does* (pp. 1–39). Portsmouth, NH: Heinemann.
- Harcourt Assessment. (2004). *Stanford English Language Proficiency Test*. San Antonio: Harcourt Assessment.
- Hart, B., & Risley, T. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator, 27*(1), 4–9.
- Hasbrouck, J., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children, 24*(3), 41–44.
- Hasbrouck, J., & Tindal, G. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59*, 636–644.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*, 303–317.

- Herman, P. (1985). The effect of repeated readings on reading rate, speech pauses, and word recognition accuracy. *Reading Research Quarterly*, 20, 553–564.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70.
- Hosp, M., & Fuchs, L. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review*, 34, 9–26.
- Hudson, R., Lane, H., & Pullen, P. (2005). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher*, 58(8), 702-714.
- Jenkins, J., Fuchs, L., van den Broek, P., Espin, C., & Deno, S. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, 95, 719–729.
- Jennings, J., McIntosh, A., Kapur, S., Tulving, E., & Houle, S. (1997). Cognitive subtractions may not add up: The interaction between semantic processing and response mode. *NeuroImage* 5, 229–239.
- Kamhi, A., Allen, M., & Catts, H. (2001). The role of the SLP in improving decoding skills. *Seminars in Speech and Language*, 22, 175-185.
- Kamil, M., & Hiebert, E. (2005). Teaching and learning vocabulary: Perspectives and persistent issues. In E.H. Hiebert & M.L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 1-23). Mahwah, NJ: Erlbaum.
- Kim, Y., & Petscher, Y. (2010). Relations of emergent literacy skill development with conventional literacy skill development in Korean. *Reading and Writing*, 23, 57-84. doi: 10.1007/s11145-010-9240-4
- Kindler, A. (2002). “*Survey of the States’ Limited English Proficient Students and Available Educational Programs and Services, 2000-2001 Summary Report*,” National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs, 2002.
- Klingner, J., & Edwards, P. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly*, 41, 108-117.

- Klingner, J., & Harry, B. (2006). The special education referral and decision-making process for English language learners: Child study team meetings and staffings. *Teachers College Record*, *108*, 2247-2281.
- Kuhn, M., & Stahl, S. (2003). Fluency: a review of developmental and remedial Practices. *Journal of Educational Psychology*, *95*(1), 3-21.
- LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293-323.
- LaBerge, D., & Samuels, S. (1979). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293-323.
- Lervag, A. & Aukrust, V. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry*. doi:10.1111/j.1469-7610.2009.02185.x.
- Lesaux, N., & Kieffer, M. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal*, *47*, 596-632.
- Lesaux, N., & Siegel, L. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology*, *39*, 1005-1019.
- Llach, M., & Gallego, M. (2009). Examining the Relationship between Receptive Vocabulary Size and Written Skills of Primary School Learners. *Atlantis*, *31*, 129-147. Retrieved from http://www.atlantisjournal.org/ARCHIVE/31.1/2009Agustin_Terrazas.pdf
- Mahon, E. (2006). High-stakes testing and English language learners: Questions of validity. *Bi-lingual Research Journal*, *30*, 479-497.
- Manis, F., Lindsey, K., & Bailey, C. (2004). Development of reading in grades K-2 in Spanish-speaking English-language learners. *Learning Disabilities Research & Practice*, *19*, 214-224.
- Mathson, D., Allington, R., & Solic, K. (2005). Hijacking fluency and instructionally informative assessments. In T. Rasinski, C. Blachowitz, & K. Lems (Eds.), *Fluency instruction: Research-based best practices* (pp. 106-119). New York: Guilford.
- Mattingly, I. (1972). Reading, the linguistic process, and linguistic awareness. In J. Kavanagh & I. Mattingly (Eds.) *Language by ear and by eye: The*

- relationship between speech and reading* (pp. 133-147). Cambridge, MA; MIT Press.
- McGlinchey, M., & Hixson, M. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33*(2), 193-203
- Morton, J. (1989). An information-processing account of reading acquisition. In A. Galaburda (Ed.), *From reading to neurons* (pp. 43–66). Cambridge, MA: MIT Press.
- Muter, V., Hulme, C., Snowling, M., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundation of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*, 665–681.
- Nation, K., & Snowling, M. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*, 342–356.
- National Clearinghouse for English Language Acquisition (2009). *The Growing Numbers of English Learner Students, 1998/99-2008/09*. Retrieved from http://www.ncela.gwu.edu/files/uploads/9/growingLEP_0809.pdf
- National Clearinghouse for English Language Acquisition FAQ (2010). *How many school-aged Limited English Proficient (LEP) students are there in the U.S?* Retrieved from <http://www.ncela.gwu.edu/faqs/view/4>
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office. Retrieved from <http://www.nichd.nih.gov/publications/nrp/report.cfm>
- National Research Council (1999). *Starting out right: A guide to promoting children's reading success*. M. Burns, P. Griffin, & C. Snow (Ed.) National Reading Council. Washington, DC: National Academy Press .
- NICHD Early Child Care Research Network (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology, 41*, 428-441.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

- Ouellette, G. (2006). What's meaning got to do with it?: The role of vocabulary in word reading and comprehension. *Journal of Educational Psychology*, 98(3), 554-566.
- Pearce, L., 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. (2006). Foreword. In K. Goodman (Ed.), *The truth about DIBELS: What it is, what it does* (pp. v–xix). Portsmouth, NH: Heinemann.
- Perfetti, C. (1985). Reading ability. New York, NY: Oxford University Press.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357–383.
- Perie, M., Grigg, W., & Donahue, P. (2005). The nation's report card: reading 2005. U.S. Department of Education, National Center for Education Statistics. Washington, D.C.: U.S. Government Printing Office.
- Poepfel, D. (2001). New approaches to the neural basis of speech sound processing: Introduction to special section on brain and speech. *Cognitive Science*, 25, 659–661.
- Pressley, M., Hilden, K., & Shankland, R. (2005). *An evaluation of end-grade-3 Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Speed reading without comprehension, predicting little* (Tech. Rep.). East Lansing, MI: Michigan State University, Literacy Achievement Research Center.
- Rasinski, T. (1990). Investigating measures of reading fluency. *Educational Research Quarterly*, 14(3), 37–44.
- Rasinski, T. (2004). Creating fluent readers. *Educational Leadership*, 61(6), 46-51.
- Rasinsky, T., & Zutell, J. (1990). Making a place for fluency instruction in the regular reading curriculum. *Reading Research and Instruction*, 29(2), 85-91.
- Riedel, B. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly*, 42(4), 556-567.

- Roehrig, A., Petscher, Y., Nettles, S., Hudson, R., & Torgeson, J. (2008). Accuracy of DIBELS Oral Reading Fluency measure for predicting third grade reading comprehension outcomes. *Journal of Psychology, 46*, 343-366.
- Rosenthal, J., & Ehri, C. (2010). Pronouncing new words aloud during the silent reading of text enhances fifth graders' memory for vocabulary words and their spellings. *Reading and Writing, 23*(5), 137-152.
- Samuels, S. (2007). The DIBELS test: Is speed of barking at print what we mean by reading fluency? *The Reading Research Quarterly, 42*(4), 563.
- Samuels, S. (2006). Reading fluency: Its past, present, and future. In T. Rasinski, C. Blachowitz & K. Lems (Eds.), *Fluency instruction: Research-based best practices* (PP. 7-20). New York: Guilford.
- Samuels, S., Schermer, N., & Reinking, D. (1992). Reading fluency: Techniques for making decoding automatic. In S. Samuels & A. Farstrup (Eds.), *What research has to say about reading instruction* (2nd ed., pp. 124-144). Newark, DE: International Reading Association.
- Schilling, S., Carlisle, J., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *The Elementary School Journal, 107*, 429-448.
- Schreiber, P. (1980). On the acquisition of reading fluency. *Journal of Reading Behavior, 12*, 177-186.
- Schwanenflugel, P., Hamilton, A., Kuhn, M., Wisenbaker, J., & Stahl, S. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of educational Psychology, 96*, 119-129.
- Seidenberg, M., Waters, G., Barnes, M., & Tanenhaus, M. (1984). When does spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*, 383-404.
- Shapiro, E., Solari, E., & Petscher, Y. (2008). Use of a measure of reading comprehension to enhance prediction on the state high stakes assessment. *Learning Individual Differences, 18*, 316-328.
- Shaw, R., & Shaw, D. (2002). *DIBELS Oral Reading Fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP)* (Technical Report). Eugene, OR: University of Oregon.

- Snow, G., Burns, M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Snow, C., Tabors, P., Nicholson, P., & Kurland, B., (1995). SHELL: Oral language and early literacy in kindergarten and first grade children. *Journal of Research in Childhood Education*, 10, 37–47.
- Snowling, M. (2005). Literacy outcomes for children with oral language impairments: Development interactions between language skills and learning to read. In H. Catts & A. Kamhi (Eds.), *The connections between language and reading disabilities* (pp.55-75). Mahwah, NJ: Erlbaum.
- Sprenger-Charolles, L., & Casalis, S. (1995). Reading and spelling acquisition in French first graders: Longitudinal evidence. *Reading and Writing: An Interdisciplinary Journal*, 7, 39-63.
- Sprenger-Charolles, L., Siegel, L., & Bechennec, D. (1998). Phonological Mediation and semantic and orthographic factors in silent reading in french. *Scientific Studies of Reading*, 2(1), 3-29.
- Stage, S., & Jacobsen, M. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*, 30, 407–419.
- Stanovich, K. (1986). Matthew Effects in Reading: Some consequences of individual Differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360- 407.
- Stephenson, A., Johnson, D., Jorgensen, M., & Young, M. (2003). *Assessing English language proficiency: Using valid results to optimize instruction*. Harcourt Technical Report (pp. 1-22). San Antonio, TX: Harcourt Assessment, Inc.
- Storch, S., & Whitehurst, G. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, 38, 934-947.
- Swanson, H., Rosston, K., Gerber, M. & Solari, E. (2008). Influence of oral language and phonological processing on children’s bilingual reading. *Journal of School Psychology*, 46, 413–429. doi:10.1016/ j.jsp.2007.07.002.
- Tannenbaum, K., Torgesen, J., & Wagner, R. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, 10(4), 381–398.

- Tenenbaum, H., & Wolking, W. (1989). Effects of oral reading rate on intraverbal responding. *The Analysis of Verbal Behavior*, 7, 83–89.
- Thomas, W., & Collier, V. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Santa Cruz, CA and Washington, DC: Center for Research on Education, Diversity & Excellence. Retrieved from: http://www.crede.ucsc.edu/research/llaa/1.1_final.html
- Torgesen, J., Rashotte, C., & Alexander, A. (2001). Principles of fluency instruction in reading: Relationships with established empirical outcomes. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain* (pp. 333–355). Parkton, MD: York Press.
- Torgesen, J., Wagner, R., & Rashotte, C. (1999). *Test of word reading efficiency (TOWRE)*. Austin, TX: Pro-Ed.
- Torgesen, J., Wagner, R., & Rashotte, C., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second- to fifth-grade children. *Scientific Studies of Reading*, 1, 161-185.
- Uchikoshi, Y. (2006). English vocabulary development in bilingual kindergartners: What are the best predictors? *Bilingualism: Language and Cognition*, 9, 33–49. doi:10.1017/S1366728905002361.
- Vander Meer, C., Lentz, F., & Stollar, S. (2005). *The relationship between oral reading fluency and Ohio proficiency testing in reading* (Technical Report). Eugene, OR: University of Oregon.
- Wanzek, J., Roberts, G., Linan-Thompson, S., Vaughn, S., Woodruss, A. & Murray, C. (2010). Differences in the relationship of oral reading fluency and high-stakes measures of reading comprehension. *Assessment for Effective Intervention*, 35(2), 67-77.
- Webb, S. (2009). The effects of pre-learning vocabulary on reading comprehension and writing. *The Canadian Modern Language Review*, 65, 441–470.
- Weiderholt, J., & Bryant, B. (2001). *Gray Oral Reading Test, Fourth Edition (GORT-4)*. Austin, TX: PRO-Ed.
- Wiley, H., & Deno, S. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education*, 26, 1-13.

- Wiley, T., & Wright, W. (2004). Against the undertow: The politics of language instruction in the United States. *Educational Policy*, 18(1), 142–168.
- Wilson, J. (2005). *The relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) oral reading fluency to performance on Arizona Instrument to Measure Standards (AIMS)* (Technical Report). Tempe, AZ: Assessment and Evaluation Department, Tempe School District No.3.
- Wood, D. (2006). Modeling the relationship between Oral Reading Fluency and performance on a statewide reading test. *Educational Assessment*, 11(2), 85-104.

APPENDIX A
TABLES AND FIGURES

Table A1

Assessment Calendar

Construct	Assessment	Grade 1	Grade 2	Grade 3
Oral reading fluency	DIBELS-ORF	spring	spring	spring
Receptive vocabulary	PPVT-III	winter	winter	winter
Reading comprehension	TerraNova		spring	spring

Table A2

Means, Standard Deviations, and Skew Statistics for Each Measure

Assessment	N	M	SD	Skew	
				Statistic	SE
First-grade					
ORF	63	61.70	30.31	.532	.302
PPVT	62	56.81	16.18	-.045	.304
Second-grade					
ORF	55	94.13	34.36	.117	.322
PPVT	53	68.49	17.84	.003	.327
TerraNova	55	591.93	26.70	.249	.322
Third-grade					
ORF	50	110.88	36.14	.247	.337
PPVT	51	88.27	17.80	-.088	.333
TerraNova	50	608.56	34.10	.016	.337

Table A3

Correlations among All Measures

	2.	3.	4.	5.	6.	7.	8.
1. ORF first-grade	.74**	.62**	.15	.18	.23	.30*	.40**
2. ORF second-grade	1	.89**	.25	.25	.28	.48**	.66**
3. ORF third-grade		1	.37**	.33*	.35*	.58**	.68**
4. PPVT first-grade			1	.87**	.71**	.73**	.58**
5. PPVT second-grade				1	.81**	.70**	.60**
6. PPVT third-grade					1	.68**	.60**
7. TerraNova second-grade						1	.75**
8. TerraNova third-grade							1

** Correlation is significant at the .01 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

Table A4

Coefficients for Multiple Regression Analyses Predicting Second-Grade TerraNova-Reading

Variable	<i>B</i>	<i>SE B</i>	β	<i>t</i>	<i>p</i>
First-grade					
ORF 1	.176	.082	.200	2.14	.037
PPVT 1	1.150	.154	.697	7.48	<.001
Second-grade					
ORF 2	.254	.073	.328	3.48	.001
PPVT 2	.918	.141	.614	6.52	<.001

Table A5

Results for Multiple Regression Models Predicting Second-Grade TerraNova-Reading

Model	<i>R</i>	<i>R</i> ²	Adj. <i>R</i> ²	Change statistics			
				ΔR^2	ΔF	<i>df</i> ^a	Sig. <i>F</i> change
First-grade							
ORF 1	.303	.092	.074	.092	5.24	1,52	.026
ORF 1 + PPVT 1	.753	.567	.550	.475	55.91	1,51	<.001
Second-grade							
ORF 2	.482	.232	.217	.232	15.44	1,51	<.001
ORF 2 + PPVT 2	.765	.585	.568	.353	42.50	1,50	<.001

^a Sample sizes varied between grades and measures due to attrition, thus producing different degrees of freedom.

Table A6

Coefficients for Multiple Regression Analyses Predicting Third-Grade TerraNova-Reading

Variable	<i>B</i>	<i>SE B</i>	β	<i>t</i>	<i>p</i>
First-grade					
ORF 1	.356	.125	.317	2.85	.006
PPVT 1	1.119	.234	.532	4.79	<.001
Second-grade					
ORF 2	.541	3.063	.546	6.07	<.001
PPVT 2	.890	.172	.466	5.18	<.001
Third-grade					
ORF 3	.507	.092	.538	5.52	<.001
PPVT 3	.794	.186	.415	4.26	<.001

Table A7

Results for Multiple Regression Models Predicting Third-Grade TerraNova-Reading

Model	<i>R</i>	<i>R</i> ²	Adj. <i>R</i> ²	Change statistics			
				ΔR^2	ΔF	<i>df</i> ^a	Sig. <i>F</i> change
First-grade							
ORF 1	.395	.156	.139	.156	8.89	1,48	.005
ORF 1 + PPVT 1	.658	.433	.409	.277	22.91	1,47	< .001
Second-grade							
ORF 2	.663	.440	.429	.440	37.74	1,48	< .001
ORF 2 + PPVT 2	.802	.644	.629	.204	26.88	1,47	< .001
Third-grade							
ORF 3	.681	.464	.453	.464	40.71	1,47	< .001
ORF 3 + PPVT 3	.785	.616	.599	.152	18.15	1,46	< .001

^a Sample sizes varied between grades and measures due to attrition, thus producing different degrees of freedom.

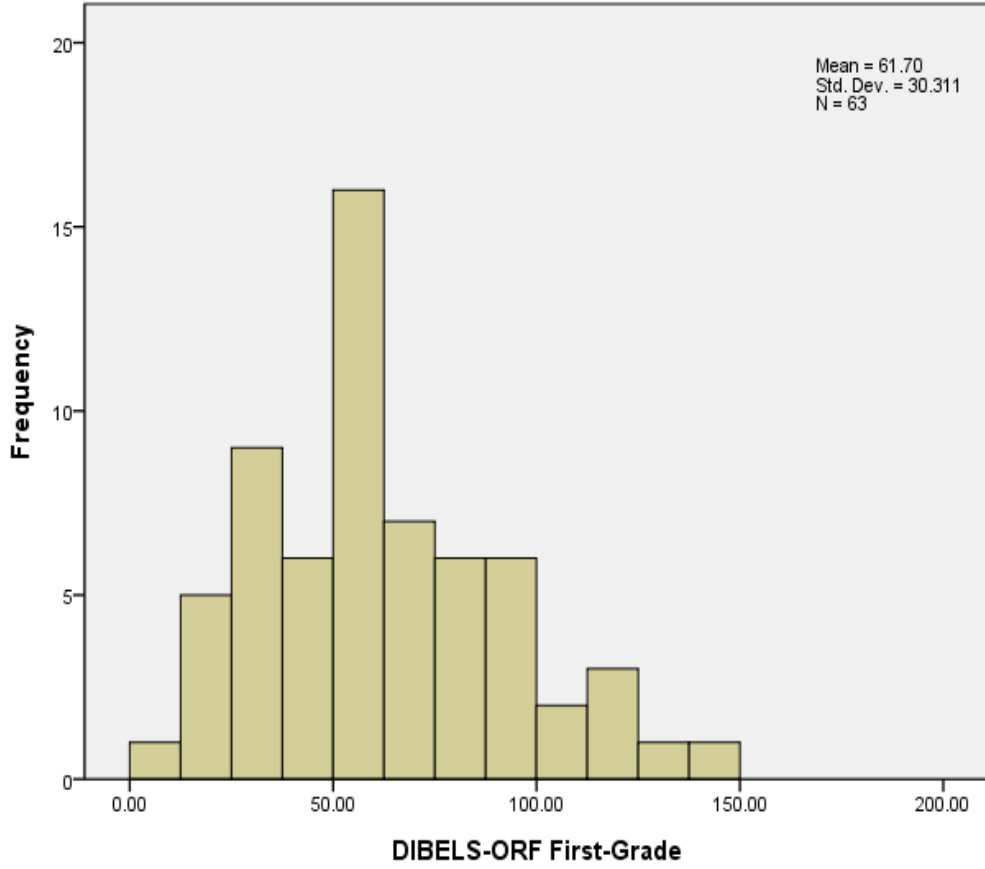


Figure A1. Distribution of first-grade DIBELS-ORF scores.

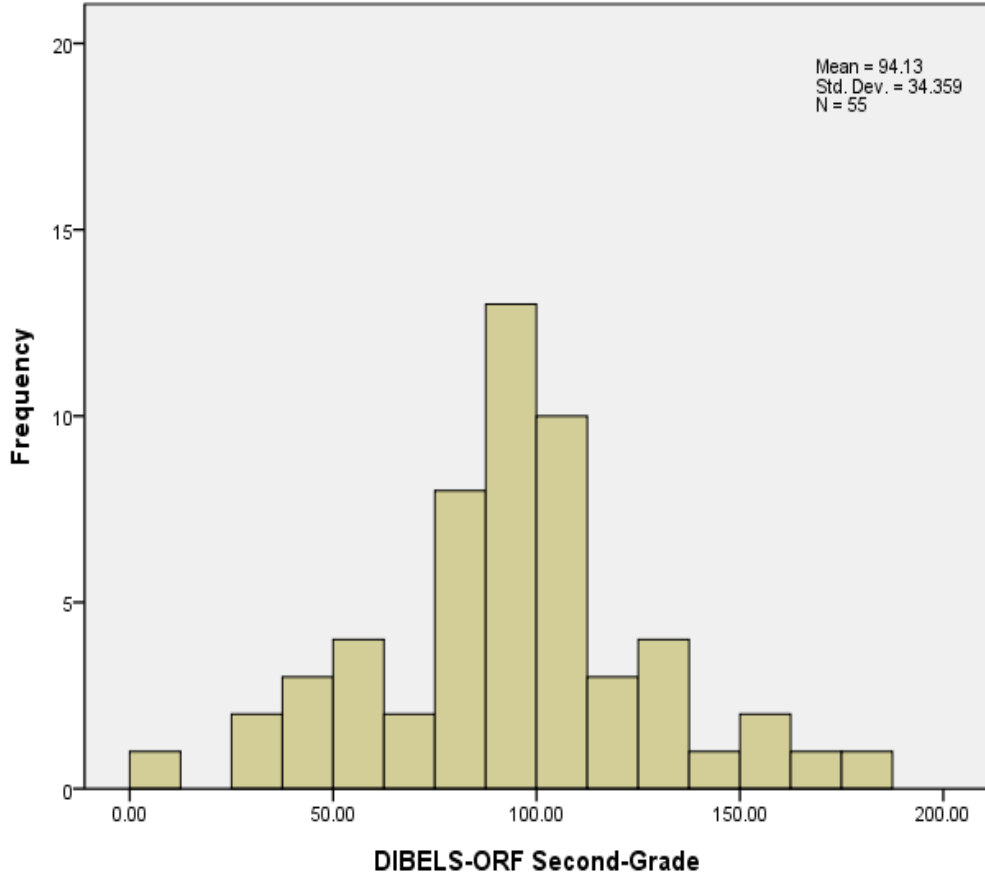


Figure A2. Distribution of second-grade DIBELS-ORF scores.

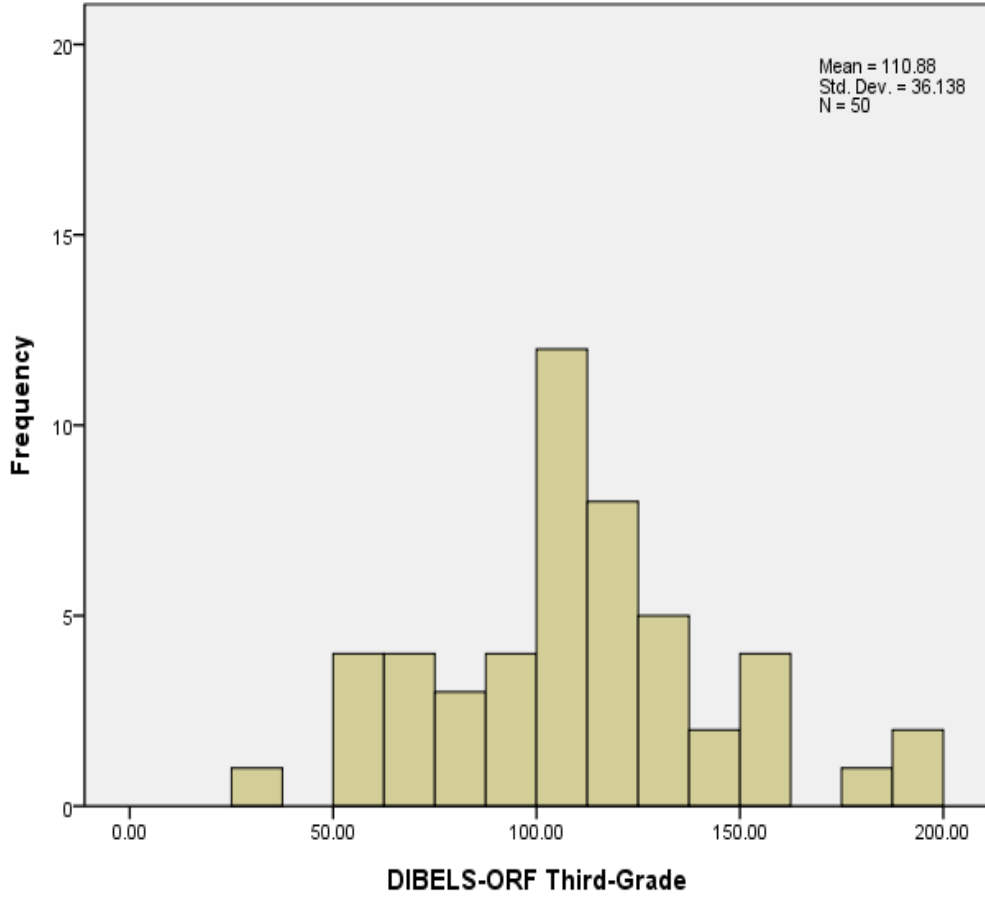


Figure A3. Distribution of third-grade DIBELS-ORF scores.

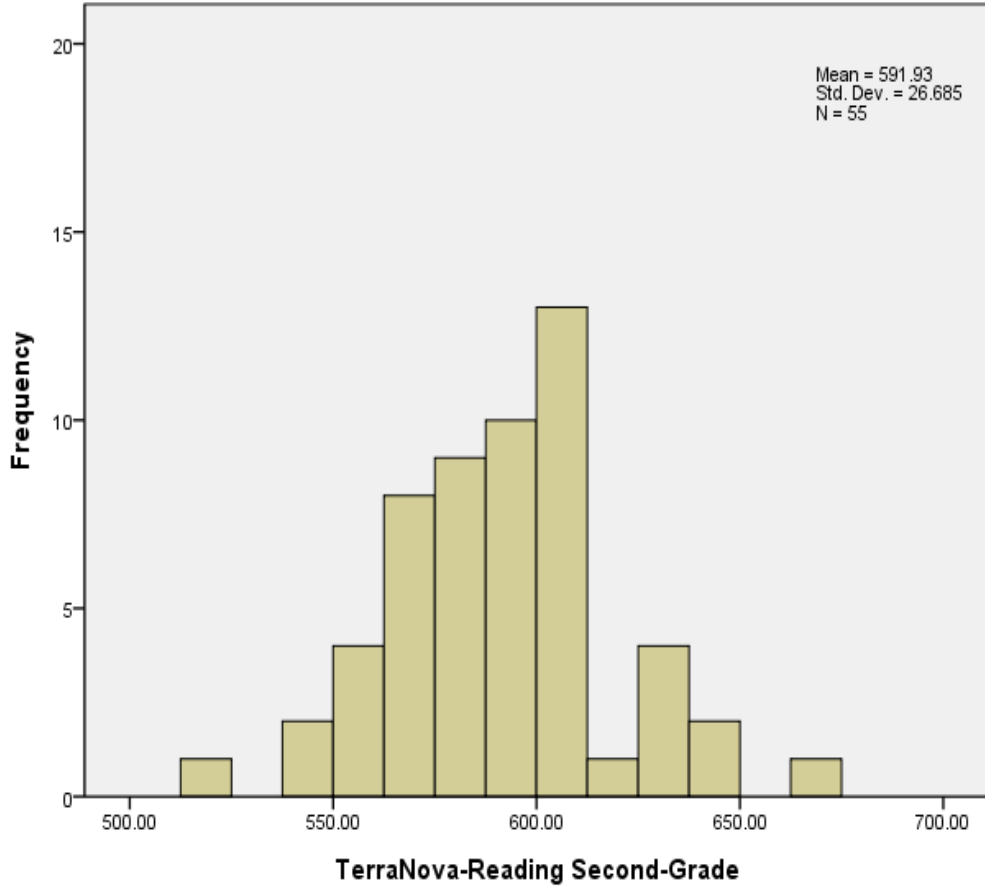


Figure A4. Distribution of second-grade TerraNova-Reading scores.

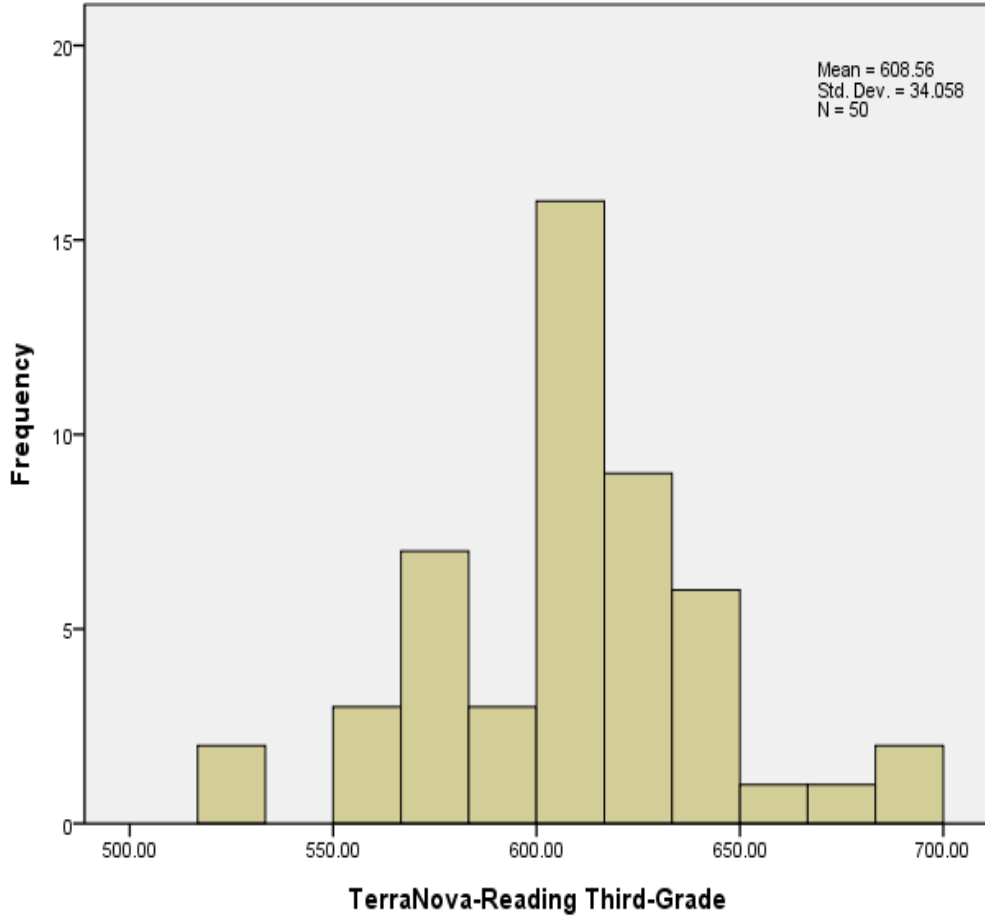


Figure A5. Distribution of third-grade TerraNova-Reading scores.

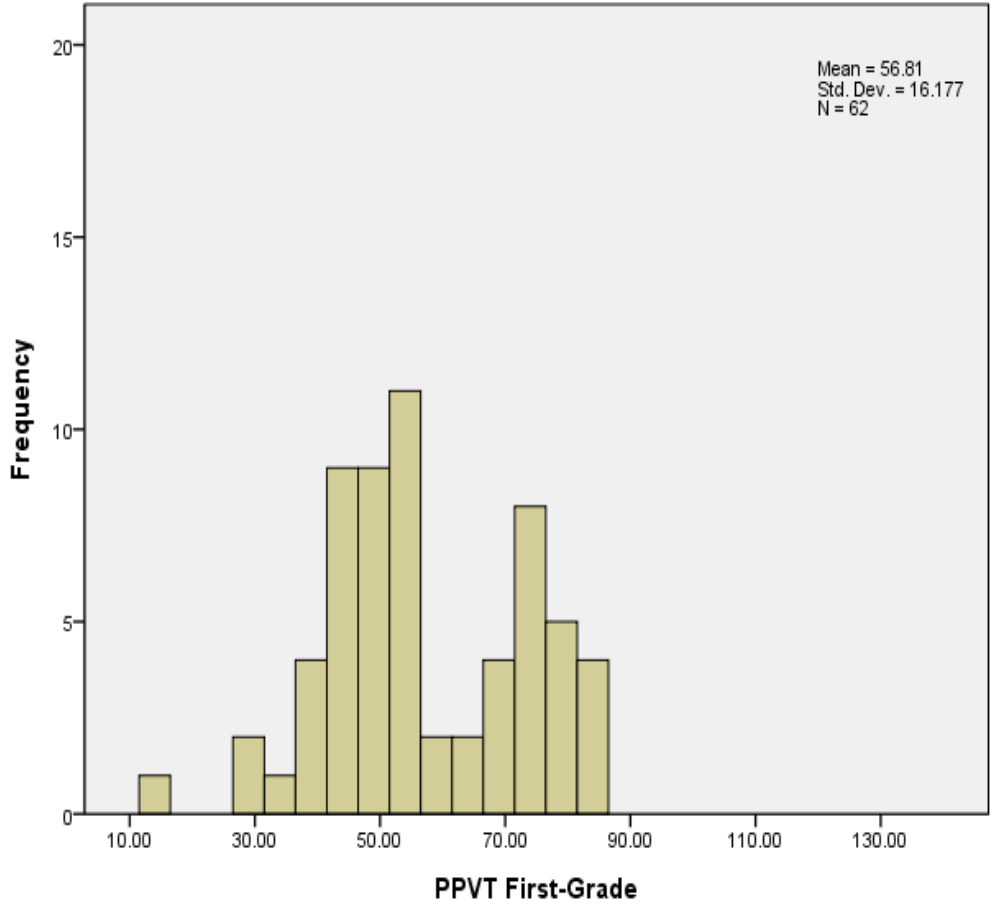


Figure A6. Distribution of first-grade PPVT scores.

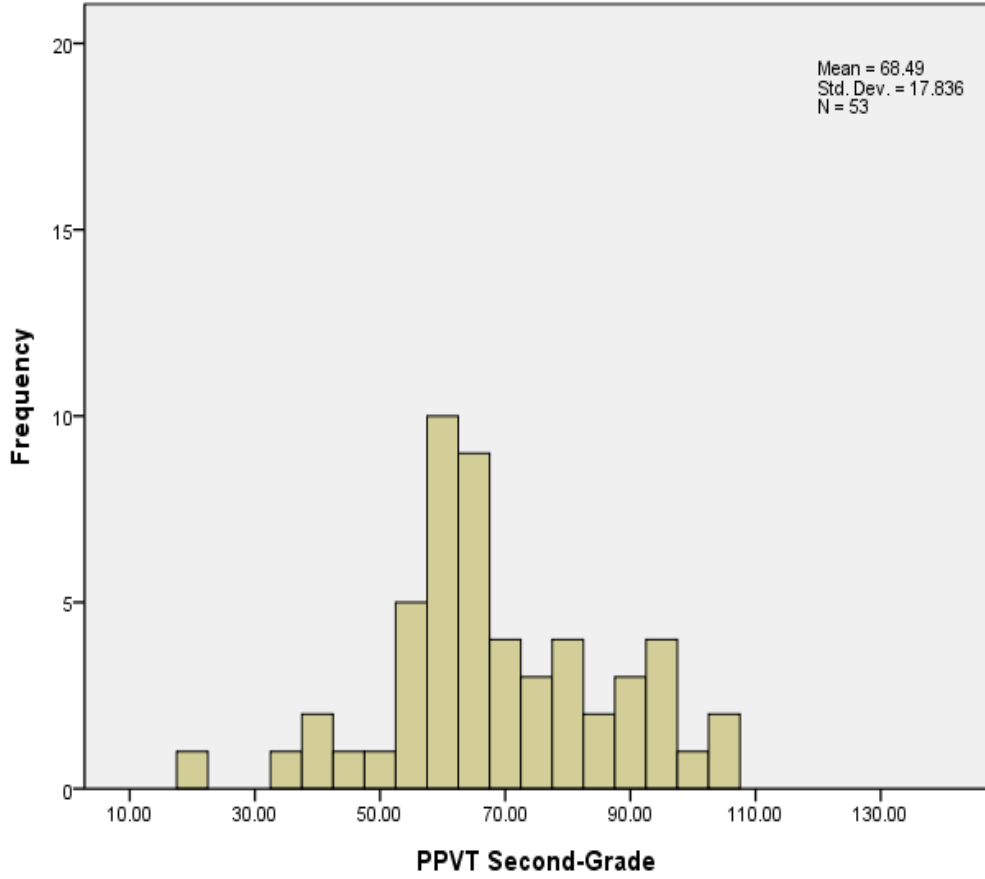


Figure A7. Distribution of second-grade PPVT scores.

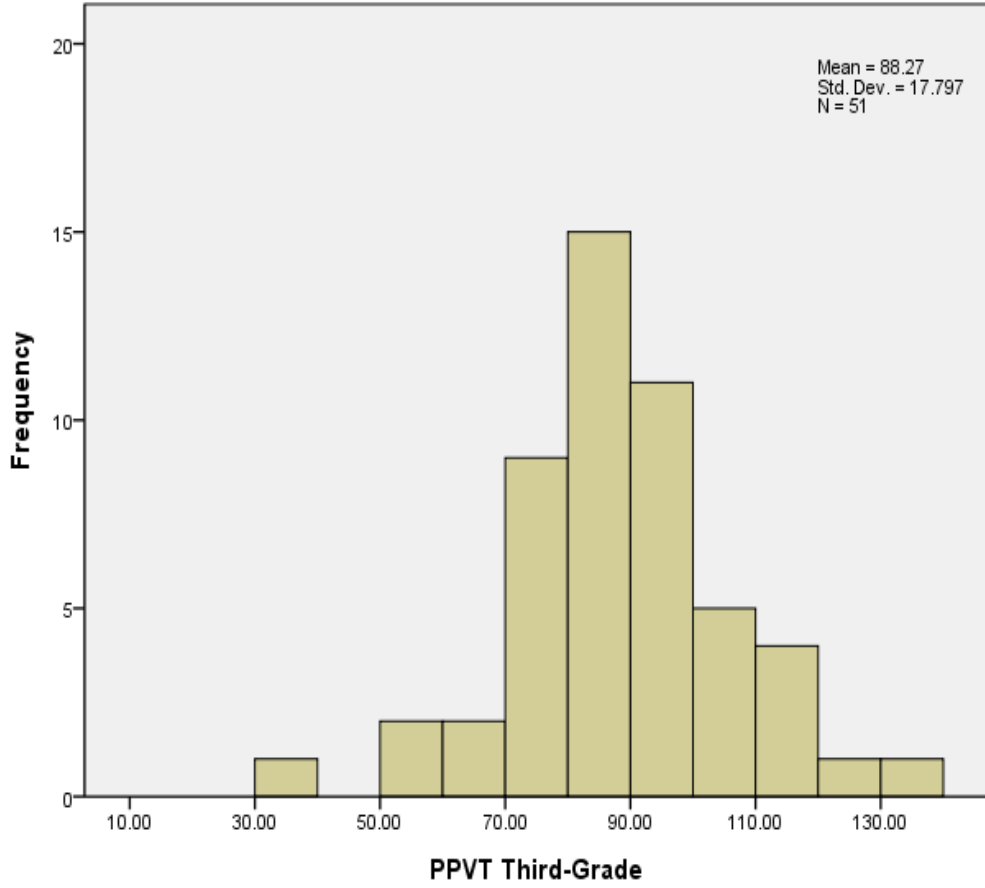


Figure A8. Distribution of third-grade PPVT scores.

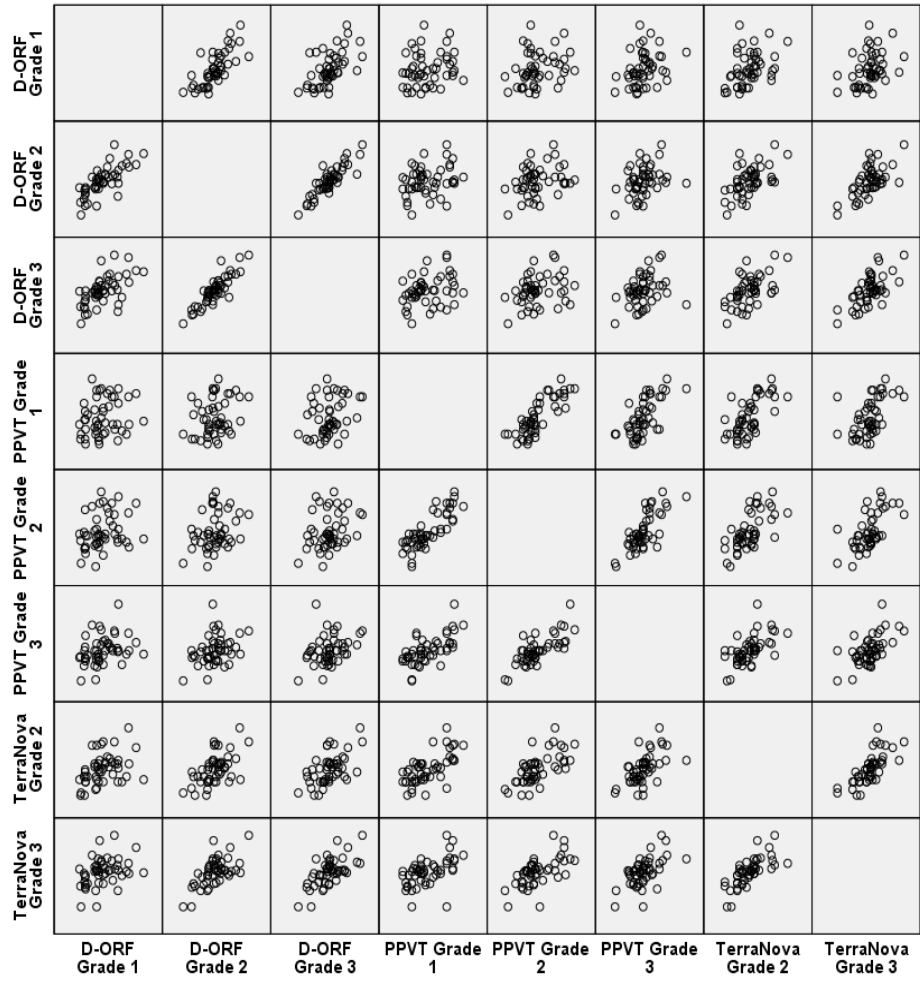


Figure A9. Scatterplot matrix depicting relationships between all variables.