

UNIVERSITY OF CALIFORNIA
RIVERSIDE

An Examination of the Predictive Validity of Early Literacy Measures for
Korean English Language Learners

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Jeanie Eunjoo Nam

December 2011

Dissertation Committee:

Dr. Mike Vanderwood, Chairperson

Dr. Sara Castro-Olivo

Dr. Greg Palardy

Copyright by
Jeanie Eunjo Nam
2011

The Dissertation of Jeanie Eunjoo Nam is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

I would like to express my deepest gratitude to my advisor and dissertation chairperson, Dr. Mike Vanderwood, for his continual support and guidance throughout this process. I would also like to thank my dissertation committee members Dr. Sara Castro-Olivo and Dr. Greg Palardy who have generously given their time and valuable feedback to better my work. Finally, I would also like to recognize and thank the University of California Linguistic Minority Research Institute (UC-LMRI) for providing me with a dissertation grant to conduct this study.

Dedication

Above all, I would like to thank my parents, Richard and Susan Nam. Thank you for your daily prayers, your many sacrifices, and for always encouraging me to do my best. I would not have made it this far without your amazing love and support.

ABSTRACT OF THE DISSERTATION

An Examination of the Predictive Validity of Early Literacy Measures for
Korean English Language Learners

by

Jeanie Eunjoo Nam

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, December 2011
Dr. Mike Vanderwood, Chairperson

The primary aim of this study was to examine the predictive validity of early literacy measures with Korean English language learners (ELLs) representing varying levels of English language proficiency. First-grade Korean ELLs ($N = 102$) were screened in the winter using measures of Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), and Word Identification Fluency (WIF). Spring reading criterion measures included Oral Reading Fluency (ORF) and the Woodcock Reading Mastery Tests (WRMT). Among the winter screening measures, WIF was found to be most correlated with end-of-the year reading outcomes for the larger sample, as well for students aggregated by language proficiency groups. While moderate to high correlations were also found for winter NWF with respect to spring outcomes, correlations were much smaller in magnitude for winter PSF. Results of hierarchical regression models, with PSF, NWF, and WIF entered respectively, also found that the addition of WIF resulted in

a large and significant change in R^2 with respect to both spring ORF and WRMT scores. Final models accounted for a total of 75.6% of the variance in spring ORF scores and 60.4% of the variance in WRMT scores. Furthermore, of the winter screening measures, only winter WIF was found to be significant at the $p < 0.05$ level in both of the final models. Finally, an examination of the diagnostic accuracy of PSF and NWF at-risk cutoffs established by the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) yielded negative predictive values at or above 85% for all tests with respect to spring outcomes; however, positive predictive values were significantly lower, ranging from 17% to 47%. Although past studies have found that many of the same assessments tools that have been used with native English speakers can be used with ELLs, the results of this study indicate that there is a need for the continued development of decision rules and appropriate assessment tools that are sensitive to the unique cultural and linguistic backgrounds of ELLs.

Table of Contents

Chapter 1: Introduction	1
The Importance of Early Identification and Intervention.....	3
Response to Intervention (RtI).....	5
Standards for Educational and Psychological Testing.....	8
Chapter 2: Review of Selected Literature	11
Cross-Linguistic Relationships.....	12
Cross-linguistic Transfer of Phonological Awareness.....	14
Predictors of Reading Achievement for ELLs.....	17
Phonological Awareness.....	18
Alphabetic Principle.....	20
Word Identification.....	25
Oral Reading Fluency.....	26
Diagnostic Accuracy.....	32
Language Proficiency Assessments.....	36
Research Questions.....	38
Chapter 3: Methods	39
Participants and Setting.....	39
Procedures.....	41
Hierarchical Regression Analyses.....	42
ROC Analysis.....	45
Materials.....	46

Chapter 4: Results	51
Descriptive Statistics for the Entire Sample.....	51
Correlational Analyses for the Entire Sample.....	53
Descriptive Statistics Aggregated by English Proficiency Level.....	55
Correlational Analyses Aggregated by English Proficiency Level.....	55
Hierarchical Linear Regression Analyses.....	59
Diagnostic Accuracy of DIBELS Cutoffs for Entire Sample.....	62
Diagnostic Accuracy of WIF.....	64
Diagnostic Accuracy of DIBELS Cutoffs for B/EI Group.....	65
Chapter 5: Discussion	67
Concurrent and Predictive Validity of Screening Tools.....	68
Summary of Regression Analyses.....	74
Relationship Between Fluency and Comprehension.....	76
Examination of Diagnostic Accuracy.....	76
Limitations.....	79
Potential Applications to Educational Practice.....	81
References	86

List of Tables

Table 1: Predictive Accuracy Table.....	97
Table 2: Winter and Spring DIBELS 6 th Edition Benchmarks for 1 st Grade.....	98
Table 3: Descriptive Statistics of Measures (N = 102).....	99
Table 4: Results of Winter and Spring Screening (N = 102).....	100
Table 5: Intercorrelations of Measures for Entire Sample (N = 102).....	101
Table 6: Descriptive Statistics of Measures for B/EI (N = 27), Intermediate (N = 42), and EA/A Groups (N = 33).....	102
Table 7: Intercorrelations of Measures for B/EI Group (N = 27).....	103
Table 8: Intercorrelations of Measures for Intermediate Group (N = 42).....	104
Table 9: Intercorrelations of Measures for EA/A Group (N = 33).....	105
Table 10: Hierarchical Multiple Regression Analyses Predicting Spring ORF and WRMT Performance From Winter Early Literacy Screening Measures (N = 102).....	106
Table 11: Predictive Accuracy of PSF With Respect to ORF for B/EI (N= 27).....	107
Table 12: Predictive Accuracy of NWF With Respect to ORF for B/EI (N= 27).....	108
Table 13: Predictive Accuracy of PSF With Respect to WRMT for B/EI (N= 27).....	109
Table 14: Predictive Accuracy of NWF With Respect to WRMT for B/EI (N= 27).....	110
Table 15. Optimal PSF and NWF DIBELS cutoffs to Predict "Below Expectations" Performance on Spring ORF and WRMT.....	111

List of Figures

Figure 1: Examples of Korean words with their English pronunciation.....	112
Figure 2: The 14 consonants and 10 vowels in Hangul.....	113
Figure 3: ROC Curve: Winter PSF DIBELS Benchmarks to Predict "Below Expectations" Performance on Spring ORF (N = 102).....	114
Figure 4: ROC Curve: Winter NWF DIBELS Benchmarks to Predict "Below Expectations" Performance on Spring ORF (N = 102).....	115
Figure 5: ROC Curve: Winter WIF Cutoffs to Predict "Below Expectations" Performance on Spring ORF (N = 102).....	116
Figure 6: ROC Curve: Winter PSF DIBELS Benchmarks to Predict "Below Expectations" Performance on Spring WRMT (N = 102).....	117
Figure 7: ROC Curve: Winter NWF DIBELS Benchmarks to Predict "Below Expectations" Performance on Spring WRMT (N = 102)	118
Figure 8: ROC Curve: Winter WIF Cutoffs to Predict "Below Expectations" Performance on Spring WRMT (N = 102).....	119

Chapter 1: Introduction

One of the most critical challenges facing schools today is the rapidly growing population of English language learners (ELLs). In 1995-1996, there were a total of 3.2 million ELLs enrolled in public schools. Most recent statistics indicate that there are an estimated 5.3 million ELLs, comprising 10% of the total public school enrollment (National Clearinghouse for English Language Acquisition, 2011). From the 1997-1998 to 2008-2009 school year, the number of ELLs has increased by 51%, while the general population of students has grown by 7%. English learners are a heterogeneous group of students representing diverse cultures, educational backgrounds, socio-economic levels, English proficiency levels, and more than 350 native languages (Hopstock & Stephenson, 2003). According to Kindler (2002), Spanish and Asian languages are the most common languages spoken by ELLs, with 79% of students claiming Spanish as their native language, followed by Vietnamese (2.0%), Hmong (1.6%), Chinese (1.0%), Korean (1.0%), and other languages (15.4%).

In the Los Angeles Unified School District, the second largest school district in the country, approximately 31.2% of the students are ELLs (California Department of Education, Educational Demographics Office, 2011). While the majority of these students speak Spanish in their homes, Korean represents the second most common language spoken by students in this district. Data from the 2000 US census indicated that the Asian population, and more specifically, the Korean population, had grown tremendously over the previous decade (Barnes & Bennett, 2002). In 1990, there were 799,000 Koreans in the US. By 2000, this number rose to 1,077,000, representing an

increase of 35%. California currently has the largest population of Koreans in the US, with approximately one-third of US Koreans residing in this state (Korean American Coalition-Census Information Center, 2003).

A critical factor for academic success is a solid foundation in literacy in the early grades, but learning to read can be a challenging process for many students. According to Good, Simmons, and Smith (1998), an estimated 20% of all students experience difficulties in learning to read. Acquiring the skills and knowledge to read in a second language can be especially difficult for ELLs (Snow, Burns, & Griffin, 1998). Despite recent efforts to close the achievement gap between ELLs and native English speakers, including the passage of federal legislation, such as the No Child Left Behind (NCLB) Act in 2002, and implementation of federally funded reading programs (e.g., Reading First), the latest results from the National Assessment of Educational Progress (NAEP) indicate that ELLs are continuing to perform significantly below native English speakers in reading achievement (National Center for Education Statistics, 2009). Assessment results from 2007 showed that 70% of fourth-grade ELLs performed below basic levels in reading achievement, while only 30% performed at or above basic. More recent statistics from 2009 showed very little changes in scores, with 71% of fourth-grade ELLs performing below basic levels (National Center for Education Statistics). When examining the reading performance of all students, including native English speakers, 33% of students performed below basic levels in 2007 and 2009.

The Importance of Early Identification and Intervention

It is important to identify students at-risk for poor reading outcomes and to implement evidence-based interventions in the early primary grades. Children who experience success in acquiring early literacy skills rapidly grow in their vocabulary knowledge and reading abilities. In contrast, children who encounter early reading difficulties have fewer experiences with reading and reduced motivation, leading them to fall behind their peers as they progress through school (Stanovich, 1986). In a longitudinal study of 54 children from first- to fourth-grade, Juel (1988) found that students who do not learn to read adequately by first-grade have a difficult time becoming successful readers in the later grades. If a student was a poor reader by the end of first-grade, there was an 88% probability that the child would remain a poor reader at the end of fourth-grade. If the child was an average reader in first-grade, there was an 87% probability that the child would remain an average reading in fourth-grade.

Fortunately, past studies have also found that many poor readers, including ELLs with low levels of English proficiency, can be brought to at least average levels of performance if they are provided with supplemental, high-quality intervention during the early stages of reading development (Gersten et al., 2007; Healy, Vanderwood, & Edelston, 2005; Lesaux & Siegel, 2003; Vaughn et al., 2006). Vaughn and colleagues provided an early literacy intervention to first-grade, Spanish-speaking ELLs who were at-risk for reading difficulties. The intervention was provided in English, which was the language of students' core reading instruction. Intervention groups of 3 to 5 students met daily, for 50 min per session. The intervention provided explicit and systematic

instruction in phonological awareness (PA), phonics, vocabulary, comprehension, spelling, and writing. Following 7 months of intervention, the intervention students significantly outperformed control students on multiple measures of English literacy skills including, letter knowledge, PA, word reading, reading comprehension, and spelling ($d = .76 - 1.24$). In particular, students in the intervention group demonstrated significantly greater gains in reading comprehension than students in the control group, as measured by the Woodcock Language Proficiency Battery Passage Comprehension subtest ($d = 1.08$). Prior to the intervention, at-risk students were more than 1 *SD* below the normative sample in reading comprehension. At the end of the intervention, these students performed within the normal range in reading comprehension, whereas control group students exhibited relatively little gains. On measures for which statistical significance was not achieved, effect sizes were still 0.40 or greater and favored the intervention group.

Lesaux and Siegel (2003) examined the effects of an early literacy intervention for students who entered kindergarten with little or no exposure to English. The sample included students from 33 different language backgrounds, but the predominant native languages spoken by students in the group were Cantonese, Mandarin, Korean, Spanish, Polish, and Farsi. Regardless of language status or ability, the students were provided with small-group PA instruction in kindergarten followed by phonics instruction in first-grade. Measures taken in second-grade indicated that the reading skills of ELLs were comparable to non-ELLs. By second-grade, many ELLs had caught up to, and in some cases, outperformed their native English peers on various reading and spelling tasks.

Follow-up assessments conducted in fourth-grade with 824 of the 1,238 children who were initially part of the study yielded similar results (Lesaux, Rupp, & Siegel, 2007). ELLs who were initially at-risk in kindergarten continued to perform similar to or better than native English speakers on fourth-grade measures of word reading, pseudoword reading, spelling, and comprehension. In addition, both ELLs and non-ELLs exhibited comparable developmental trajectories in word reading from kindergarten through fourth-grade. The results of this study provide strong evidence that a model of early identification and intervention is critical to support the literacy development of ELLs who may be at-risk for reading problems.

Response to Intervention (RtI)

Given the benefits of early identification and intervention, the National Research Council (2002) has recommended that states adopt a universal screening and multi-tiered intervention strategy in general education as a means of identifying at-risk children and providing early intervention. Response to Intervention (RtI) is a tiered model with a focus on prevention, evidence-based instruction, and early identification. With this model, the needs of students are matched to high-quality instruction and learning rate is monitored over time in order to make important educational decisions.

In Tier 1, all students are provided with high quality, scientifically-based literacy instruction in the general education classroom. Reports by federally commissioned groups such as The National Literacy Panel on Language-Minority Children and Youth (August & Shanahan, 2006) and the Institute of Education Sciences (Gersten et al., 2007), have concluded that literacy instruction for ELLs must include explicit instruction

in five critical areas: PA, alphabetic knowledge, fluency, vocabulary, and comprehension. Universal screening is also conducted 3 times a year in order to monitor students' progress and to identify students who may be in need of supplemental instruction in Tier 2. In order to determine whether a student may be at-risk and in need of secondary-level reading intervention, schools have typically relied on measuring skills related to literacy development (e.g., PA, letter knowledge) and applying a specific cutoff to classify students into at-risk and not-at-risk groups.

The purpose of Tier 2 is to provide targeted intervention that enables at-risk students to “catch up” to their peers. Intervention is provided in small-group settings and progress is monitored regularly to ensure that students are benefiting from the intervention. Students who do not respond to intervention in Tier 2 may then be considered for more intensive and long-term intervention or referral to special education in Tier 3. Decisions about the intensity and duration of the intervention are based on students' learning rates and levels of performance.

In order to correctly identify at-risk students at Tier 1, valid and reliable screening tools must be identified and appropriate benchmarks must be developed for differentiating children who are at-risk from those who are not at-risk. High rates of false positives may result in resources being allocated to children who do not truly need remedial instruction. Of greater concern, however, is a high rate of false negatives. The consequences of failing to identify children who are at-risk far outweigh the consequences of over-identification, such that if these children are not identified, they will fail to receive the additional support that is imperative for their academic success.

With the growing population of ELLs, there is clearly a need to determine how to best serve this diverse population of students. The adoption of a data-based, problem-solving approach, such as RtI, may assist ELLs who are at-risk for reading disabilities and help prevent these students from falling behind their peers. However, one major challenge with RtI has been the development of screening tools and decision rules for identifying ELLs who are most in need of Tier 2 and Tier 3 services.

English learners who struggle with reading may be provided with general education intervention services within Tier 2 and may also be considered for a referral to special education, which is regarded as Tier 3 in many schools. Approximately 66% of ELLs identified as qualifying for special education services are served in programs for students with learning disabilities (LD; Zehler, Fleischman, Hopstock, & Stephenson, 2003), of which about 56% have difficulties in the area of reading (USDOE & NICHD, 2003). One major challenge with assessing ELLs for consideration of special education services is that low levels of English proficiency may sometimes be misinterpreted as a learning disability or an intelligence deficit (Langdon, 1989). In fact, past studies have found that ELLs who are less proficient in English are more likely to be placed in special education programs than their more proficient English-speaking peers (Abedi, 2006; Artiles, Rueda, Salazar, & Higuera, 2005). These results indicate that appropriate screening tools are needed to determine whether students' academic difficulties reflect a learning disability, limited English proficiency, or lack of access to high quality instruction matched to the students' needs (Klingner, Artiles, & Barletta, 2006). Without an accurate classification model that is sensitive to the diverse linguistic and cultural

backgrounds of ELLs, these students may be inappropriately identified and misplaced in special education.

Reading problems need to be detected early and empirically validated reading instruction must be provided to assist ELLs in developing English literacy skills. Although there is a substantial amount of research supporting the use of early literacy screening measures to address the literacy problems of native English speakers (e.g., Vellutino, Scanlon, & Tanzman, 1998; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993) and a growing literature base supporting its use with Spanish-speaking ELLs (e.g., Baker & Good, 1995; Fien et al., 2008; Vanderwood, Linklater, & Healy, 2008), there have been no studies that have examined the validity of these measures with Korean ELLs. The primary aim of this study will be to examine whether early literacy measures that are used to screen native English speakers in Tier 1 are valid indicators of reading proficiency for Korean ELLs.

Standards for Educational and Psychological Testing

There are several factors that must be considered when conducting assessments with individuals from diverse cultural and linguistic backgrounds. As outlined in the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 1999), using assessment tools and instructional strategies with ELLs based on studies that have been conducted with native English speakers is not consistent with recommended practices. The *Standards* state that:

Test norms based on native speakers of English either should not be used with individuals whose first language is not English or such individuals' test results should be interpreted as reflecting in part current level of English proficiency, rather than ability, potential, aptitude, or personality characteristics or symptomatology. (p. 91)

Several of the large-scale assessments that are currently used with ELLs were developed to assess native English speakers. Consequently, many of these assessments have failed to include ELLs in their norming samples (Butler & Stevens, 2001; Davidson, 1994). It is imperative that ELLs are included in the development, validation, and norming of assessments, especially when data from these measures are used to make important educational decisions. Failure to include ELLs in these samples may potentially lead to invalid interpretation of data and inappropriate test score use (Holmes & Duron, 2000).

Students' language background must also be carefully considered when selecting, administering, and interpreting test performance. According to the *Standards*, any test that employs language is, in part, a measure of students' language skills. Early literacy screening measures are typically conducted during the early elementary grades, which is when many ELLs are still at the very early stages of English language acquisition. Unfortunately, using tests with individuals who have not sufficiently acquired proficiency in the language of the test may seriously compromise the reliability and validity of these assessments.

Recent statistics clearly indicate that ELLs represent a culturally and linguistically diverse group of students. Consequently, we cannot assume that what works with one group of ELLs will work for all subgroup populations of ELLs. According to the *Standards*, “test developers should collect for each linguistic subgroup studied the same form of validity evidence collected for the examinee population as a whole” (AERA, APA, & NCME, 1999, p. 97). In addition, the *Standards* recommend that separate, parallel analyses of data be conducted using samples from different linguistic subgroups. These standards call for an examination of different subgroup populations of ELLs when conducting assessment validation studies. In order to determine whether ELLs are at-risk for reading difficulties and in need of secondary-level reading interventions in Tier 2 or 3, decision rules and appropriate assessment tools must be specifically developed for different subgroup populations of ELLs.

Chapter 2: Review of Selected Literature

Learning how to speak, read, and write in English is not a simple skill for children to acquire (Cummins, 1984; McLaughlin, 1992). English learners need to learn the complex rules and structure of a new language, which involves understanding the grammar, vocabulary, and pronunciation of the language. In addition to developing English literacy skills, students must also acquire the academic knowledge (e.g., terminology and writing conventions) associated with each content area (National Clearinghouse for English Language Acquisition, 2007).

The rate of second language acquisition is a complex process and may be influenced by a multitude of factors including, students' instructional environment, age of arrival, prior educational history, and native language skills (Cummins, 1984). ELLs represent a very diverse group of students. While some students have had previous schooling and are able to read and write in their native language, others have had little or no formal schooling and have limited native language literacy skills (Lesaux, 2006). As a result, the rate at which these students acquire proficiency in English may vary substantially. According to Cummins, English learners can develop social proficiency, or basic interpersonal conversation skills (BICS), within the context of everyday living and without formal instruction. Although BICS can be acquired within 2 to 3 years, it is not sufficient to meet the cognitive and linguistic demands of the classroom. The acquisition of cognitive academic language proficiency (CALP) has been found to be reliant on formal schooling, taking about 5 to 7 years to develop. Consequently, while English learners may be able to quickly acquire the linguistic skills to interact with their peers, it

may take significantly longer for them to develop the academic language proficiency needed for school success.

Cross-Linguistic Relationships

While English and Korean are both alphabetic languages, such that graphemes in both languages correspond to phonemes, there are fundamental differences between the two languages. One significant difference is that Korean orthography has a nonlinear spatial layout, similar to Chinese. The symbols are arranged from left-to-right and top-to-bottom to form square blocks that correspond to single syllables (Perfetti & Dunlap, 2008; Perfetti & Liu, 2005; Taylor, 1980; see Figure 1). For example, the word for market, /si-jang/, is written 시장, instead of a linear string of five letters, s | i | j | a | n | g. Each syllabic block represents a CV, CVC, or CVCC syllable (Taylor & Taylor, 1995). The location of the symbol within the syllable block also corresponds to the position of the sound within the spoken syllable (Cho & McBride-Chang, 2005). For example, the first sound in the syllable is always in the left or top position of the syllable-block. Since phonemes and syllables are both important phonological units represented in Korean, Hangul has been referred to as an alpha-syllabary script (Wang, Park, Lee, & 2006).

Korean has often been characterized as an orthographically transparent or shallow language such that there is a one-to-one correspondence between graphemes and phonemes (Cho & McBride-Chang, 2005; Kang & Simpson, 1996). There are 14 consonants and 10 vowels in the Korean alphabet, also known as Hangul (see Figure 2). English, in contrast, has a deep orthographic structure, such that a letter in English can represent a variety of sounds and a single sound can be represented by several different

letters (Genesee, Geva, Dressler & Kamil, 2006). For example, the sound /f/ can be represented in English by each of the following graphemes: “f” as in *fun*, “ph” as in *phone*, and “gh” as in *cough*. The single letter “a” can also represent several different vowel sounds in words such as “*about, fan, car, hate, fall, and bare* (Taylor, 1980).

Although several of the phonemes represented in Korean are similar to those in English (e.g., /m/ and /n/), Korean lacks the sounds /f/, /v/, initial /l/, and the /th/ sounds as in *think* and *this* (Taylor & Taylor, 1995).

Children may use different strategies when learning to read orthographies with varying levels of transparency (Ziegler & Goswami, 2005). For example, the regularity and consistency in which graphemes correspond to phonemes in shallow orthographies (e.g., Greek, Italian, Spanish, German) may assist children with learning the letter-sound correspondences of the language, which in turn can be used to facilitate word recognition (Wimmer, Landerl, Linortner, & Hummer, 1991). Additionally, studies have found that grapheme-phoneme recoding skills are rapidly acquired by children learning to read languages with shallow orthographies, such that accurate decoding of words in these languages may often approach ceiling levels within the first year of schooling (Goswami, Ziegler, & Richardson, 2005). In contrast, learners of languages with deep orthographies (e.g., English, Danish, French) may develop accurate and fluent decoding skills at a much slower rate.

According to Katz and Frost’s (1992) orthographic depth hypothesis (ODH), the transparency of the first- (L1) and second-language (L2) may impact L2 reading acquisition. A cross-linguistic study by D’Angiulli, Siegel, and Serra (2001) examined

the reading abilities of 81 English-speaking children whose L1 was Italian. Strong correlations were found between various English and Italian reading tasks. The authors asserted that exposure to a language with more predictable grapheme–phoneme correspondences, such as Italian, may enhance English phonological and literacy skills.

Cross-linguistic Transfer of Phonological Awareness. Phonological awareness (PA) is the understanding that spoken words and syllables can be divided into smaller components, or phonemes. Although early PA skills, such as rhyming, identifying the beginning and ending sounds of words, and blending can be seen in preschool children as young as 2 or 3 years old, more advanced PA is typically evident between kindergarten and first-grade (Chafouleas, Lewandowski, Smith, & Blachman, 1997; Daly, Chafouleas, Skinner, 2005).

The terms “phonological awareness” and “phonemic awareness” are often used interchangeably in the literature, however, these terms refer to distinctive skills. Phonological awareness can be conceptualized as occurring at the level of the syllable, onset-rime, or phoneme (Goswami, 2000). Syllable awareness refers to the ability to segment words into syllables (e.g., pencil, into pen-cil). Onset-rime awareness is the ability to detect that a syllable is made up of the onset, which corresponds to any phonemes before the vowel, and the rime, which refers to the vowel sound and the phonemes that follow (e.g., cat, into c-at). Finally, phonemic awareness refers to the ability to detect the smallest sounds that make up words (e.g., fan, into f-a-n). Therefore, phonemic awareness is a subcomponent of phonological awareness that focuses on recognizing and manipulating individual sounds in words.

Although PA is a general skill that is involved in learning to read alphabetic languages, different levels of PA may be more or less relevant for reading, depending on the language that is considered (Durgunoglu & Öney, 1999). For example, a study with native speakers of Greek indicated that syllable and phoneme level awareness made significant and independent contributions to early reading acquisition (Aidinis & Nunes, 2001). Høien, Lundberg, Stanovich and Bjaalid (1995) found that phoneme, syllable, and onset-rime tasks contributed significantly to predicting performance on a word recognition tasks for Norwegian first-graders. However, phoneme awareness explained far more variance in word recognition tasks than syllable and onset-rime awareness. Native English speakers tend to segment the syllable into the onset and rime (Stahl & Murray, 1994), and onset-rime awareness has been found to contribute to the development of English literacy skills (Goswami & Mead, 1992). The saliency of the rime unit in English may be attributed to the high percentage of rime neighbors in English (Ziegler & Goswami, 2005). For example, there are many more words that share the same sounds in the rime unit for the word *cat* (e.g., bat, pat, that, fat, etc.) than in the body (e.g., cap, can, etc.).

Given the high salience of syllables in the Korean script relative to other languages, it is not surprising that both phoneme-and syllable-level awareness have been found to play a role in learning to read in Korean (Cho & McBride-Chang, 2005). Studies with beginning, native Korean speakers have also found that the body-coda unit (e.g., CV-C, the Korean word for bear, kom, into ko-m) may be a more accessible sub-syllabic structure than the onset-rime (e.g., C-VC, kom, into k-om; Yi, 1998). Increased

sensitivity to the body-coda unit among Korean speakers may be due to the greater proportion of body than rime neighbors in Korean phonology. Additionally, children's body-coda awareness has been found to be a better predictor of Korean pseudoword reading and spelling than their onset-rime awareness (Kim, 2007).

According to Cummins's (1979) linguistic interdependence hypothesis, there is a significant relationship between children's literacy development in their L1 and L2, such that children who manifest deficits in their L1 are more likely to develop similar difficulties in their L2. In contrast, children who have strong literacy skills in their L1 will be able to apply these skills to acquiring literacy in their L2. Despite having considerable differences in orthography and phonological structures, studies have found evidence of cross-linguistic transfer of PA with Korean speakers. Kim (2008) examined the cross-language transfer of Korean PA skills with a sample of Korean-English bilingual kindergarten students in the US. Korean PA was assessed using measures of syllable, onset-rime, body-coda, and phonemic awareness. English PA was measured using the blending, matching, and segmenting subtests of the Comprehensive Test of Phonological Processing (CTOPP). Results of the study indicated that children's overall Korean PA and English PA were positively correlated ($r = .84$). In addition, Korean PA was significantly correlated with English sight word reading ($r = .67$), pseudoword reading ($r = .63$), and word reading ($r = .55$). Individual measures of Korean PA, including Korean onset-rime awareness, body-coda awareness, and phonemic awareness were also significantly correlated with English PA ($r = .77, .52, \text{ and } .86$, respectively). According to Kim:

As normally developing bilingual children develop their representations of phonological structures by accumulating phonetic similarities from lexical information in two languages, their phonological representations may develop cross-linguistically as a language general ability. While their phonological representations may be anchored in their L1 initially, and at a certain point of bilingual proficiency, L2 phonological characteristics transfer to their L1 phonological representations, resulting in bidirectional influence between L1 and L2 phonology. (p. 13)

Studies with Spanish-speaking students have also found evidence supporting the cross-linguistic transfer of PA across languages. For example, Quiroga, Lemos-Britton, Mostafapour, Abbott, & Berninger (2002) examined the factors that predict English word reading for Spanish-speaking ELLs. Measures of Spanish PA were significantly correlated with English PA ($r = .63$), and both Spanish and English PA measures were found to predict English word reading for first-grade students. Furthermore, children who were able to read words accurately in Spanish were also better at reading words in English, indicating that both PA and word reading ability can transfer across languages.

Predictors of Reading Achievement for ELLs

A growing body of research has demonstrated that many of the same early literacy screening tools and reading performance indicators that are used with native English speakers can be used to identify ELLs who may be at-risk for poor reading outcomes and in need of supplemental reading instruction (Baker & Good, 1995; Gersten et al., 2007; Geva, Yaghoub-Zadeh, & Schuster, 2000; Haager & Windmueller, 2001;

Lesaux & Siegel, 2003; Lesaux et al, 2007). The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are a set of standardized, grade-based early literacy measures that can be used for screening purposes (Good & Kaminski, 2002). These measures assess critical early literacy domains as discussed in the report by the National Reading Panel (2000) including, PA, phonics, oral reading fluency, and comprehension. Whereas state mandated assessments are usually administered to students once per year and typically take several days to administer, these early literacy screening tools can be administered quickly (1-3 min), providing teachers with immediate data that can be used to identify students who are not showing adequate progress.

Healy, Vanderwood, and Edelston (2005) screened first-grade students in a district comprising over 90% ELLs using DIBELS measures of PA and phonics. Based on the results of the initial screening, a Tier 2 intervention focused on PA, phonics, and vocabulary instruction was provided 2 times per week, for 16 weeks, at 30 min per session. By the end of the intervention period, all but 2 of the 15 intervention students reached the established cutoff scores on the DIBELS measures and were exited from the intervention. These results suggest that early literacy assessment tools, such as DIBELS, that are used to identify native English speakers who struggle with reading, may also assist in the identification of ELLs who may need additional support in reading.

Phonological Awareness. According to Stanovich (1986), children who are at-risk for reading failure can be identified as early as kindergarten using measures of phonological skills. A deficit in PA has been found to be the primary cause of most reading disabilities (Vellutino & Scanlon, 1987). In fact, as many as 90% of children and

adults with reading difficulties display significant deficits in PA (Blachman,1994). There is also considerable evidence that an understanding of these phonological units is a critical predictor of reading acquisition for both native English speakers (Stanovich, Cunningham & Cramer, 1984) and ELLs (Gersten et. al., 2007). While it is very common for schools to overlook or delay addressing reading difficulties until students' oral language proficiency is further developed, measures of PA have been found to be more reliable predictors of reading abilities than oral language proficiency (Durgunoglu, Nagy, & Hancin-Bhatt, 1993; Geva et al., 2000). Furthermore, regardless of English language proficiency level, English learners who receive a balanced early reading program with explicit and intensive, small-group PA instruction have been found to make significant long-term gains in reading ability (Healy et al., 2005; Lesaux & Siegel, 2002; Vaughn et al., 2006).

Given the cross-language transfer of PA across languages, as well as the prevalence of studies that have shown PA to be predictive of reading ability for ELLs, there is significant reason to believe that assessments of PA would assist educators in identifying students who may be at-risk for reading difficulties. While studies have found that children's ability to segment, isolate, and manipulate phonological units in spoken words is predictive of English reading development for native English speakers and ELLs (Cho & McBride-Chang, 2005; Kim, 2007; Quiroga, 2002), there have been no published studies to date that have examined whether a commonly used measure of PA, DIBELS Phoneme Segmentation Fluency (PSF), is also predictive of future reading outcomes for Korean ELLs.

Alphabetic Principle. The alphabetic principle, or the knowledge of how letters correspond to sounds, is a critical prerequisite to proficient reading, enabling beginning readers to sound out word segments and blend these sounds to form words (National Reading Panel, 2000). One measure of alphabeticity, pseudoword reading, has been found to be strongly correlated with real-word reading and comprehension for both native English speakers and ELLs (Burke & Hagan-Burke, 2007; Fien et al., 2008; Swanson, Trainin, Necochea, & Hammill, 2003; Vanderwood et al., 2008). A potential difficulty that may arise when students are asked to read real words is that it may not always be clear whether students are recalling words from memorization or reading words by applying their understanding of phonics rules. The use of pseudoword reading measures does not allow for students to use rote memory of whole words and specifically isolates how well they are able to apply their understanding of letter-sound correspondences

One commonly used measure of pseudoword reading is Nonsense Word Fluency (NWF). With this measure, students are presented with a page of pseudowords (e.g., *sim*), and asked to either produce the individual letter sounds constituting the pseudoword or to read the whole pseudoword. Scores are determined by counting the number of correctly produced letter sounds in 1 min (Good & Kaminski, 2002). Studies that have examined the validity of NWF have found this measure to be strongly related to overall reading achievement. For example, Good, Simmons, and Kame'enui (2001) found NWF in the winter of first-grade to be significantly correlated with reading fluency scores in the spring of first grade ($r = .78$) with a sample native English speakers. In addition,

results indicated that NWF accounted for 60% of the variance in reading fluency outcomes.

Burke and Hagan-Burke (2007) examined the concurrent validity of NWF with the Phonetic Decoding Efficiency and Sight Word Efficiency subtests of the Test of Word Reading Efficiency (TOWRE). The Phonetic Decoding Efficiency measure requires students to read aloud as many words as possible from a list of non-words (e.g., knop, plod). The Sight Word Efficiency measure assesses students' accuracy and fluency in reading aloud phonetically regular and irregular words. Results indicated that NWF was significantly correlated with the Phonetic Decoding Efficiency ($r = .75$) and Sight Word Efficiency ($r = .68$) subtests of the TOWRE for a sample of first-grade students.

Contrary to the many studies that have supported the utility of pseudoword reading measures, there has also been some criticism that pseudoword reading measures are unrelated to comprehension and inappropriate as indicators of reading achievement for ELLs (Goodman, 2006). This argument has been based on the notion that students with limited English proficiency may be able to identify letter sounds and fluently read pseudowords, but may not have the necessary vocabulary skills required for adequate comprehension. In particular, it has been suggested that students whose native language is based on an alphabetic writing system that is similar to English (e.g., Spanish) may be able to identify letter sounds without necessarily comprehending text. Several studies, however, have found that pseudoword reading is also significantly related to measures of reading fluency and comprehension for ELLs (Fien et al., 2008; Lesaux & Siegel, 2003; Vanderwood et al., 2008).

Vanderwood et al. (2008) examined the relationship between first-grade NWF and third-grade reading outcome with a sample of ELLs primarily composed of Latinos (90%) and Asians (7%). Third-grade reading outcome measures included oral reading fluency (ORF), Maze, and the California Achievement Test Sixth Edition (CAT-6), a state mandated reading assessment. With ORF, students are asked to read a grade-level passage of connected text aloud for 1 min, during which the examiner counts the number of errors and the number of words read correctly. Maze is a multiple-choice cloze task that students complete while reading silently. The first sentence of each passage is left intact. Thereafter, every n^{th} word is deleted and replaced with 3 multiple-choice alternatives inside parentheses. Significant correlations were found for first-grade NWF scores and third-grade ORF ($r = .65$), Maze ($r = .54$), and the California Achievement Test reading composite ($r = .39$). In addition, when the variance associated with English language proficiency level was removed through hierarchical linear regression, NWF continued to account for a significant amount of variance in all three outcome measures.

Fien et al. (2008) evaluated the concurrent and predictive validity of NWF as an index of beginning reading proficiency for ELLs and native English speakers in kindergarten through second-grade. A total of five cohorts from Oregon participated, with each cohort representing approximately 2,400 students. Criterion measures of reading achievement included ORF and the Stanford Achievement Test (SAT-10), a group administered, norm-referenced test of overall reading proficiency. ORF was administered to students in the spring of first- and second-grade and the SAT-10 was administered in the spring of kindergarten, first- and second-grade.

When examining the results of the entire sample (ELL and non-ELL students), Fien et al. (2008) found concurrent correlations between NWF and the criterion measures to remain relatively stable across the different data collection periods. For example, the concurrent correlation between NWF and SAT-10 in the spring of kindergarten and the spring first-grade were .73 and .65, respectively. The concurrent correlation between NWF and ORF in the spring of first-grade was .76. Predictive validity correlations for NWF and the outcome measures were also found to be significant when examining data from the entire sample (Fien et al., 2008). The correlation for NWF in the winter of kindergarten with ORF and SAT-10 in the spring of first- grade were .65 and .63, respectively. Moderate correlations were also found for NWF in the winter of kindergarten with ORF and SAT-10 in the spring of second- grade ($r = .51$ and $.56$, respectively). In addition, the initial 1 min. administration of NWF collected in the winter of kindergarten accounted for 31% of the variance in SAT-10 scores collected in the spring of second-grade. This pattern of correlations is consistent with other past studies that have found NWF to be significantly associated with future reading outcomes (e.g., Good et al., 2001; Vanderwood et al., 2008).

Separate analyses of ELLs and native English speakers indicated that the magnitude of the relationship between NWF and criterion reading measures were typically as strong for ELLs as native English speakers (Fien et al., 2008). Of the 24 correlations between NWF and the outcome measures (ORF and SAT-10) collected at different time points between kindergarten and second-grade, only 7 of the correlations were found to be significantly different for ELLs and native English speakers. In other

words, for 79% of the comparisons, the correlations were statistically equivalent for ELLs and native English speakers, suggesting that NWF appears to function fairly similarly for both groups. Interestingly, of the correlations that were found to be significantly different, 5 of these correlations involved the initial administration of NWF, which occurred during the winter of kindergarten. When a statistically significant difference was found between NWF and outcome measures for ELLs and native-English speakers, correlations were higher for native-English speakers. These results suggest that assessments of NWF in kindergarten may be less predictive of future reading outcomes for ELLs than for native English speakers.

One major limitation with the Fien et al. (2008) study is that the impact of varying levels of language proficiency was not taken into account in their analyses. If trends in Oregon are similar to those found in California, the significantly different correlations between ELLs and native English speakers in kindergarten may be attributed to high numbers of students in kindergarten who are in the beginning stages of second language acquisition. Based on English language proficiency data from California, approximately 30% of kindergarten English learners were performing at the Beginning, or lowest level of language proficiency in 2010-2011, while only 9% of students were performing at the Beginning level in first-grade, as assessed by the California English Language Development Test (CELDT; California Department of Education, 2011b). Assessment results from 2002 to 2006 also reflect larger percentages of students at the Beginning level of language proficiency in kindergarten when compared to first-grade (California Department of Education). It is possible that the predictive validity of literacy

measures, such as NWF, may differ for students with limited English proficiency skills. While an assessment of NWF provides useful information regarding students' acquisition of the alphabetic principle, further examination is required in order to determine whether it is predictive of future reading performance for students with varying levels of English proficiency.

Word Identification. Word identification fluency (WIF) assesses automatic word recognition skills by having students read isolated words from a high-frequency word list for a specified amount of time, typically 1 min (Fuchs, Fuchs, & Compton, 2004). In an early study by Deno, Mirkin, and Chiang (1982), WIF was found to be significantly correlated with the reading comprehension subtest of the Peabody Individual Achievement Test ($r = .76$) with a sample of third-grade students. A more recent study by Fuchs et al. examined whether NWF or WIF was a better predictor of reading outcomes for a sample of at-risk first-grade native English speakers and ELLs. Results indicated that fall measures of WIF and assessments of WIF slope across time were superior to NWF (fall assessment and slope) in predicting end-of-the-year reading outcomes, as measured by the Woodcock Reading Mastery Tests (WRMT) Word Identification and Word Attack subtests, and the Comprehensive Reading Assessment Battery (CRAB). Although it found that WIF was superior to NWF in predicting end-of-the-year reading outcomes, one major limitation of this study was the restricted sample of students. The children who were included in this study were the lowest performing students in their class and at-risk for poor reading outcomes, limiting the generalizability of these findings. Another significant limitation of this study was that separate analyses

were not conducted for native English speakers and ELLs. Therefore, it cannot be determined whether WIF functions similarly for both groups of students.

While several studies have found measures of pseudoword reading to be predictive of reading skills for ELLs (e.g., Fien et al., 2008; Lesaux & Siegel, 2003; Vanderwood et al., 2008), there have been fewer studies that have examined the predictive validity of WIF for this population of students. With NWF, students must apply their knowledge of letter-sound correspondences in order to decode the pseudoword. Consequently, measures of pseudoword reading can be used to obtain a measure of a student's decoding skill without any confounding of sight word familiarity or vocabulary knowledge. In contrast, WIF uses high-frequency words, which may be quickly recognized and understood by native English-speakers, but may not be as familiar to students who are learning English. It is also possible that limited exposure to English words and English vocabulary may impact students' fluent recognition of these words.

Oral Reading Fluency. The ultimate goal of reading instruction is to help children acquire the skills necessary to derive meaning from text. Comprehending text requires more than a good vocabulary. It involves accurate and fluent decoding skills as well as the ability to use syntax to anticipate words in a sentence, monitor context, and make inferences on the basis of background knowledge (McGuinness, 2004). In Gough's "simple view of reading", reading is characterized as the product of two separate, but interdependent component skills, comprehension and decoding ($R = C \times D$; Hoover & Gough, 1990). In first-grade, decoding ability has been found to account for about 80%

of the variance in reading comprehension (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998). According to Perfetti's Verbal Efficiency Theory (1985), fluent readers are able to read text quickly and accurately, which allows them to expend less cognitive resources on decoding and have a higher capacity for comprehension. In contrast, non-fluent readers experience tremendous difficulty with word recognition, which demands excessive cognitive resources and negatively impacts comprehension.

Typically, an assessment of ORF requires students to read a passage aloud, during which the examiner counts the number of errors and the number of words read correctly. Although the use of ORF in schools has become widespread in recent years, a concern that some teachers have expressed about this measure is that it is solely a measure of speed reading and not an index of students' comprehension of the text (Fuchs, Fuchs, & Maxwell, 1988). Both Goodman (2006) and Samuels (2007) have strongly criticized the use of ORF, suggesting that it emphasizes speed rather than comprehension, and that it may actually penalize students who are carefully searching for meaning within the text.

There has also been some debate among reading researchers regarding the existence and characteristics of word callers. Word callers are individuals who appear to efficiently decode text, but do not comprehend (Stanovich, 1986). Consequently, using ORF with these students would greatly overestimate their reading abilities. Hamilton and Shinn (2003) explored this topic by examining whether oral reading measures would in fact overestimate the ability of students who were identified by their teachers as word callers. The results of this study did not support the existence of word callers with a

population of native English speakers. Students who were identified as word callers by their teachers did not fit the expected profile- high decoding and low comprehension skills. Instead, students who were identified as word callers exhibited lower comprehension and lower decoding skills relative to their peers. In fact, average performance of the word callers on the Passage Comprehension Test of the Woodcock Reading Mastery Test (WRMT) indicated performance below the 30th percentile compared to the normative sample. Consistent with their low comprehension scores, the word callers also had significantly lower oral reading fluency scores than comparison students who were perceived by their teachers to have similar fluency skills.

Despite the poor face validity of ORF among teachers (Fuchs et al., 1988; Good & Kaminski, 2002), a number of past studies have found moderate to strong correlations between ORF and measures of reading comprehension for both native English speakers and ELLs (e.g., Baker & Good, 1995; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Good & Kaminski, 2002; Shinn, Good, Knutson, Tilly, & Collins, 1992). In addition, past studies have demonstrated that ORF is superior to other, more “direct” assessments of comprehension, including oral and written retell of stories, cloze passages, Maze, and question answering (Ardoin et al., 2004; Fuchs & Fuchs, 1992; Fuchs et al., 1988; Wiley & Deno, 2005)

Baker and Good (1995) evaluated the relationship between ORF and reading comprehension with a sample of Spanish-English bilingual students and native English speakers. ORF was administered to students 2 times per week, for a total of 10 weeks. Results indicated that ORF was sensitive to growth and significantly correlated with the

Stanford Diagnostic Reading Comprehension subtest for both native English speakers and ELLs, with correlations ranging from .51 to .82. Correlations between ORF and the Stanford Diagnostic Reading Comprehension subtest were not found to be significantly different for native English speakers and ELLs, indicating that the magnitude of the relationship between ORF and the reading outcome measures was comparable for both groups of students.

Fuchs et al. (1988) compared the relationship between the Reading Comprehension subtest of the SAT-10 and four indices of reading comprehension, including oral reading fluency, oral recall, question and answer, and cloze. The reading fluency measure required students to read a 400-word traditional folktale aloud for 5 min, while the examiner scored omissions, repetitions, substitutions, and mispronunciations as errors. Student performance was reported in terms of average numbers of words read correctly and incorrectly per min and the average numbers of omissions, substitutions, repetitions, and hesitations per minute. For the retell procedure, students were required to retell the folktale that they had just read without referring back to the text. The question and answering assessment required students to respond to 10 short-answer comprehension questions regarding critical ideas from the folktale they had read. Finally, a random cloze procedure was created, involving the 400-word traditional folktales, with every 7th word deleted from the passage and replaced with a blank. Subjects were required to read the passage and supply the missing words within 10 min. Although oral and written recall measures, question and answer assessments, and cloze have been regarded by teachers to have higher face validity than fluency measures (Fuchs & Fuchs,

1992), results of the correlational analyses indicated that students' oral reading rate score was most strongly correlated with the Stanford Reading Comprehension subtest ($r = .91$). ORF was found to be a robust measure for assessing overall reading proficiency, including comprehension with this population of students. However, since this study examined the validity of these measures with a fairly restricted sample (mildly and moderately handicapped middle and junior high school boys), these findings may not be generalizable to other groups of students.

There have been several studies that have examined the relationship between Maze and criterion measures of reading comprehension. The popularity of this particular measure may be due to the ease of administration and scoring, high face validity, and relatively strong correlations with other measures of reading comprehension (Fuchs & Fuchs, 1992). A study conducted by Ardoin et al. (2004) examined the predictive validity of ORF and Maze with a sample of third-grade students. Criterion measures of comprehension included the Woodcock Johnson III Achievement Test Broad Reading Scale (WJ-III BR) and the Passage Comprehension subtest (WJ-III PC). The Broad Reading Scale of the WJ-III is comprised of three subtests: Letter-Word Identification, Reading Fluency, and Passage Comprehension. Although both ORF and Maze were significantly correlated with WJ-III BR ($r = .73$ and $.50$, respectively) and PC scores ($r = .41$ and $.31$, respectively), Maze did not explain significant variance in both WJ-III BR and PC scores after the predictive value of reading fluency scores were taken into account.

Wiley and Deno (2005) evaluated the association between ORF, Maze, and the Minnesota Comprehensive Assessment (MCA), a statewide reading assessment, with a sample of ELLs. Participants included students whose home languages were Hmong, Spanish, and Somali. As expected, ORF was significantly correlated with the MCA for both third- ($r = .61$) and fifth-grade ELs ($r = .69$). Although Maze was also significantly correlated with the MCA for both third- ($r = .52$) and fifth-grade students ($r = .57$), consistent with findings by Ardoin and colleagues (2004) with native English speakers, Maze did not account for significant variance in MCA scores after ORF scores were taken into account.

While the developers of DIBELS have suggested that ORF is “one of the best measures of reading competence, including comprehension, for children in first through third grades (Good & Kaminski, 2002) they also suggest using an additional measure of Retell Fluency (RTF) immediately following ORF for two purposes: 1) first, to identify children whose comprehension is not consistent with their fluency; and 2) second, to increase the face validity of ORF. However, one major concern that has been raised about the RTF measure is whether it can be reliably scored (Goodman, 2006). Scoring procedures for fluency measures (oral and retell) can vary dramatically (e.g., total words retold, total number of content words, total idea units) and these methods are often be very unreliable and time consuming (Fuchs & Fuchs, 1992; Fuchs et al., 1988).

Riedel (2007) evaluated the predictive validity of ORF and RTF with a total of 1,518 first- and second-grade students comprised primarily of native English speakers. Reading outcomes were assessed using the Group Reading Assessment and Diagnostic

Evaluation (GRA+DE), a standardized, group administered test of overall reading ability. An interesting finding was that the relationship between ORF and GRA+DE was stronger for ELLs ($r = .80$) than non-ELLs ($r = .67$) at the end of first-grade. Although the correlations between RTF and GRA+DE were lower than the correlations between ORF and GRA+DE, the relationship between RTF and GRA+DE was again stronger for ELLs ($r = .69$) than non-ELLs ($r = .51$). Another important finding was that end of first-grade, ORF was able to predict first- and second-grade reading comprehension status (satisfactory vs. poor) with 80% and 71% accuracy, respectively. Although RTF is designed to be a measure of comprehension, it was found to be weaker than ORF in predicting comprehension. Moreover, the combination of RFT and ORF was not found to improve the predictive power beyond that provided by ORF alone. While alternative measures of reading comprehension such as Maze or RTF may have greater face validity, results of past studies with both native English speakers and ELLs indicate that these measures do not provide a significant amount of information regarding students' reading abilities beyond that explained by ORF.

Diagnostic Accuracy

The purpose of benchmark goals is to provide educators with standards for determining the progress of all students (University of Oregon Center on Teaching and Learning, 2008). Benchmarks represent minimum levels of performance that students must reach in order to be considered on track for achieving early literacy goals (Good, Simmons, Kame'enui, Kaminski, & Wallin, 2002). In addition to benchmarks, DIBELS also provides cutoff scores for three levels of risk status- at risk, some risk, and low risk.

Students with scores below these cutoffs are extremely unlikely to meet subsequent literacy goals unless additional support is provided. While demonstrating a strong relationship between screening tools and reading outcome measures is important, established cutoff scores for these measures, such as those developed by DIBELS, must also be evaluated in order to determine whether they accurately identify students who are at-risk and not at-risk for poor reading outcomes.

An examination of sensitivity, specificity, positive predictive power, negative predictive power, and hit rate is essential when evaluating the predictive validity of screening measures (Rathvon, 2004). The sensitivity index refers to the percentage of children predicted to be poor readers who actually become poor readers (true positives). In contrast, the specificity index refers to the percentage of children predicted to be good readers who actually become good readers (true negatives). According to Rathvon, sensitivity and specificity indices should be at least 75-80% in order for the measure to conform to reliability standards for screening measures. The positive predictive value reflects the proportion of valid positives compared with the total number identified as at risk, whereas the negative predictive value refers to the proportion of valid negatives compared with the total students identified as not at risk. Finally, the hit rate, or the overall effectiveness of the screening measure, refers to the percentage of correctly identified students (valid positives + valid negatives / total number of students screened (see Table 1).

Some measures with high hit rates may accurately predict children who *do not* become poor readers but poorly predict those who *do* develop reading problems. For

example, Felton (1992) found that kindergarten measures of PA, phonological memory, and rapid naming had an overall hit rate of 80%, as determined by students' performance on reading measures in third-grade. While the valid negative rate was 97%, the valid positive rate was only 31%, and the false positive rate was 69%. These percentages indicate that only a small percentage of poor readers were correctly identified (valid positive) and many students who eventually became good readers were misidentified (false positive). Thus, the overall hit rate of 80% made the screening measures appear more useful than they actually were.

There is a reciprocal relationship between sensitivity and specificity. The sensitivity of a test may be increased, but only at the expense of sensitivity and vice versa. For example, using kindergarten "at-risk" DIBELS cutoff scores for NWF, Nelson (2008) found that assessments of NWF in January had a sensitivity index of .67 and a specificity index of .73 when the Woodcock Johnson Letter-Basic Reading Skills Cluster was used as the criterion measure in May of kindergarten. Using "at-risk" criteria for PSF, both sensitivity and specificity were found to be .67. Using the "some-risk" criteria yielded higher sensitivity indices for both NWF (.94) and PSF (.94), but significantly lower specificity indices (.53 and .38, respectively).

While benchmarks and cutoff scores have been established for many early literacy measures, there have been very few studies that have examined whether these scores are valid for different subgroup populations ELLs. Vanderwood et al., (2008) evaluated the predictive validity of DIBELS NWF with an ELL population. Using decision rules established by DIBELS (Good, Wallin, Simmons, Kame'enui, & Kaminski,

2002), students who achieved scores below 50 sounds/min on NWF in the spring of first-grade were classified as “at-risk”, and students with scores above 50 sounds/min were classified as “not at-risk.” Cut-scores for third-grade outcome measures (Maze, ORF, and CAT-6) were also clearly determined, such that scores below the 25th percentile were considered “below expectations”, and scores at or above the 25th percentile were “at or above expectations”. Results of the diagnostic accuracy analysis indicated that NWF in first-grade demonstrated adequate specificity indices (.79-.82) and hit rates (.65-.75), but significantly lower sensitivity indices (.43-.55). Using a 1-min assessment in first-grade, NWF was able to correctly identify approximately 80% of the students who scored above 25th percentile on all three outcome measures in third-grade. However, NWF was not as accurate at predicting who would perform below the 25th percentile on third-grade outcome measures (sensitivity).

Interestingly, results from Vanderwood et al. (2008) also indicated that the majority of the students (80%) who were classified as “not at risk” with NWF, yet were below expectations in third-grade (false negatives), were classified as having the lowest levels of English language proficiency in first-grade, as assessed by the California English Language Development Test (CELDT). These findings suggest that NWF may function differently for ELLs with lower levels of English proficiency. Although past studies have found strong correlations between NWF and criterion measures of reading proficiency, very few studies have examined whether this measure is predictive of future reading performance for students with varying levels of English proficiency. For students who are in the very early stages of English acquisition, these results suggest that different

or additional screening measures may be necessary in order to accurately determine whether these students are at-risk and in need of early literacy intervention.

In order for tests to have high rates of accuracy in differentiating students who are at-risk from students who are not at-risk, appropriate cutoffs must be established such that sensitivity and specificity are maximized. Although measures of PA and pseudoword reading have been found to be significantly correlated with later reading outcomes, it is critical to examine whether established cutoffs, such as those recommended by DIBELS, are also valid for different subgroup populations of ELLs with varying levels of English proficiency.

Language Proficiency Assessments

In accordance with Title III of NCLB, all states are required to establish English proficiency standards and assess ELLs with a statewide English language proficiency (ELP) assessment that reflects these standards. In the state of California, all students whose primary language is not English, as indicated in their Home Language Survey, must be tested with the CELDT annually. There are 3 primary aims of the CELDT: 1) to identify students as ELLs in grades K-12; 2) to determine English language proficiency levels; and 3) to monitor students' annual progress in acquiring the skills of listening, reading, speaking, and writing in English (California Department of Education, 2011a). The initial development of the CELDT was based on the ELP standards drafted by a panel of practitioners and experts in 1999-2000. During its early development in 2000, the CELDT consisted primarily of items from the Language Assessment Scales. However, more recent versions of the CELDT have replaced these

items with questions that are more closely aligned with the California ELD standards (Linguanti & George, 2007). Based on the technical report, the CELDT has also been found to have strong psychometric properties, with test-retest reliability coefficients between 0.86 to 0.90 across all grades and subtests (CTB/McGraw-Hill, 2004).

Students' performance in the areas of reading, listening, speaking, and writing are used to classify them to one of five English proficiency levels: Beginning, Early Intermediate, Intermediate, Early Advanced, and Advanced. Students are re-assessed with the CELDT every year until they are reclassified as Fluent English Proficient (RFEP; California Department of Education, 2011a). Although students who are acquiring English as a second language are classified as ELLs, these students vary greatly in their development of English proficiency. CELDT assessment results from 2010-2011 indicated that 9% of first-grade ELLs performed at the Beginning proficiency of level, 19% at Early Intermediate, 39% at Intermediate, 26% at Early Advanced, and 6% at Advanced (California Department of Education, 2011b).

ELLs represent a diverse group of students with varying levels of English proficiency levels. Consequently, it cannot be assumed that assessments that are predictive of future reading outcomes for ELLs with advanced levels of English proficiency will also be predictive for students who are still in the early stages of English language development. One goal of this study was to examine whether students' level of English language proficiency, as assessed by the CELDT, impacts the relationship between early literacy measures and future reading performance.

Research Questions

The primary purpose of this study was to examine the predictive validity of early literacy measures with a sample of first-grade Korean ELLs representing varying levels of English proficiency. The goal of this study was to answer the following research questions:

1. What is the concurrent relationship between winter measures of PSF, NWF, and WIF?
2. What is the concurrent relationship between spring measures of PSF, NWF, WIF, and ORF?
3. What is the concurrent relationship between students' fluency and comprehension in the spring, as assessed by ORF and the WRMT, respectively?
4. How much variance in spring outcomes (ORF and WRMT) is explained by winter predictor variables (PSF, NWF, and WIF)?
5. After controlling for student's performance on winter early literacy measures, does CELDT level explain additional variance in spring outcomes?
6. What is the predictive accuracy (sensitivity, specificity, positive predictive power, negative predictive power, and hit rate) of winter PSF, NWF, and WIF with respect to spring outcomes?

Chapter 3: Methods

Participants and Setting

Approval from the Institutional Review Board was obtained prior to conducting this study. This study took place at three public elementary schools in Southern California. All three schools operate on traditional-year school schedules and are comparable in their school and student profiles. School A serves approximately 510 students in grades K-5. Seventy-six percent of the students participate in free or reduced-price lunch programs and 61% of the students are classified as ELLs. Among the ELLs at this school, approximately 59% speak Spanish, 35% of the students speak Korean, and the remaining 6% of the ELLs speak other languages including Bengali, Tagalog, and Arabic. The ethnic distribution of school A is largely Hispanic (54%), Asian (38%), African American (1.5%), and White (1.3%). A total of 47 students from school A participated in this study. School B serves approximately 957 students in grades K-6. Sixty-five percent of the students participate in free or reduced-price lunch programs and 62% of the students are classified as ELLs. The home languages spoken by these students are primarily Spanish (52%), Korean (46%), and other languages (2%). Hispanic (51%), Asian (41%), African American (4%), and other ethnicities (4%) make up school B. A total of 25 students from school B participated in this study. Finally, School C serves 980 students in grades K-6. Eighty-five percent of the students participate in free or reduced-price lunch programs and 62% of the students are classified as ELLs. Similar to Schools A and B, the primary language spoken by these students include Spanish (52%), Korean (46%), and other languages (2%). The ethnic distribution of school C is largely Hispanic

(58%), Asian (32%), Filipino (5%), and African American (2.7%). Thirty students from school C participated in this study.

Participants were composed of first-grade Korean ELLs ($N = 102$). During the first year of data collection, a total of 26, 1st grade students participated in the initial screening. An additional 76 students were screened during the second year of data collection. Students' ELL status and home language were confirmed with official school files. These students were also assessed by the school district with the CELDT. A total of 2 children moved during the study, and their scores were excluded from analyses. Of the final sample of 102 students, 55 (54%) of the participants were females and 47 (46%) were males. Four of the participants received special education services under the disability categories of a specific learning disability ($N = 1$) and autism ($N = 3$). All of these students were mainstreamed in the general education classroom and received pull-out services for part of the day. These students were also included in the final sample. Students received Open Court reading instruction, a district- and state-adopted reading-language arts curriculum. In addition, students participated in the Structured English Immersion (SEI) program, which consists of daily instruction in English language development (listening, speaking, reading, and writing) and thirty to forty-five minute blocks of daily ELD instruction.

Of the final sample, a total of 3 students were at the Beginning CELDT level, 24 were at the Early Intermediate level, 42 were at the Intermediate level, and 33 were at the Early Advanced or Advanced level. Prior to data collection, it was determined that the sample would be separated into four groups (Beginning, Early Intermediate,

Intermediate, and Early Advanced/Advanced) based on CELDT classifications. However, due to the limited number of students at the Beginning level of English proficiency, the students in this group were combined with the Early Intermediate group. The decision to combine the Early Advanced and Advanced proficiency students was based on state guidelines for reclassification. Reclassification is the process by which students, who have been identified as English learners, are reclassified as fluent English proficient (R-FEP). California state guidelines require students' CELDT scores to be at the Advanced or Early Advanced level in order to be considered for reclassification (California Department of Education, 2011a). Following data collection, students were grouped into three English language proficiency levels: (1) the Beginning/Early Intermediate (B/EI) group ($N = 27$), the Intermediate group ($N = 42$), and the Early Advanced/Advanced (EA/A) group ($N = 33$).

Procedures

Students were tested in the winter (January/February) and spring (June) of first-grade during both year 1 and year 2. The following measures were administered in the winter: PSF, NWF, and WIF. Spring assessments included PSF, NWF, WIF, ORF and the WRMT. All of the assessments were administered by the principal investigator and school psychology graduate students trained in assessment methodology and the administration of the literacy measures. Standardized administration procedures, as indicated in the DIBELS and WRMT administration manual, were followed by the examiners. All of the measures were administered in English and testing took place in a separate room to minimize distractions. Testing took approximately 5 to 10 min per

student in the winter and 25 to 30 min per student in the spring. Examiners were provided with a detailed training on administration and scoring. Moreover, examiners had an opportunity to observe appropriate test administration and were provided with written instructions on test administration and scoring. Interobserver agreement data was calculated on 20% of the probes administered in the winter and 20% of the probes administered in the spring during the first year only. Interobserver agreement was calculated for each probe by dividing the number of agreements per probe by the number of disagreements plus disagreements, and multiplying by 100.

Pearson correlation coefficients were calculated to examine the concurrent relationships between measures, as well as to determine the predictive validity of winter screening measures with respect to end of the year reading outcomes. A series of multiple regression analyses were also conducted to determine the variance in spring outcomes explained by each predictor variable (PSF, NWF, and WIF). All results were analyzed at the $p < .05$ significance level. Finally, receiver operating characteristic (ROC) curves were generated for screening measures and area under the curve (AUC) values evaluated to examine the diagnostic accuracy of DIBELS cutoffs with respect to each of the spring reading outcomes.

Hierarchical Regression Analyses. While standard multiple regression is used to examine the relationship between a set of independent variables (IV) and a dependent variable (DV), with hierarchical regression, the proportion of variance accounted by all of the IVs (R^2) is partitioned incrementally (Pedhazur, 1997). The predictor variables are entered into the regression analysis in a pre-determined order, which allows the

researcher to examine the increase in the proportion of variance accounted for by each IV (ΔR^2). Thus, hierarchical regression is used to evaluate the relationship between a set of IVs and the DV after having controlled for the impact of other variables.

In order to examine the proportion of variance explained by winter screening measures, as well as language proficiency level to spring reading performance, a series of hierarchical regression models were tested for each of the reading outcomes (ORF and WRMT Total Reading- Short Scale score). The purpose of these analyses was to examine the amount of variance associated with each predictor variable (PSF, NWF, and WIF) and to determine whether language proficiency level explains additional variance in spring outcomes after the variance associated with early literacy measures is removed from the model. All regression coefficients (β and B), the amount of variance associated with the addition of each predictor variable (R^2 and ΔR^2), and F-values are reported.

Violations of assumptions including, linearity, independence, and homoscedasticity were examined through visual analysis of residual plots, bivariate scatterplots, and descriptive statistics for the independent variables. Although predictor variables (winter PSF and NWF) were slightly skewed, analysis of residuals for outcome variables indicated that regression assumptions were met. Multicollinearity diagnostics were examined in two different ways. First, the following series of regressions analyses were conducted for the three predictors: (a) PSF was regressed on NWF and WIF, (b) NWF was regressed on WIF and PSF, and finally (c) WIF was regressed on PSF and NWF. Resulting R^2 values ranged from .17 to .39. Second, Variance Inflation Factor values were examined for each of the predictor variables using Wetherill's (1986)

recommendation ($VIF < 10$). The VIF is $1/\text{Tolerance}$ and measures the impact of multicollinearity among the variables in a regression model. Since the highest VIF value for the independent variables was significantly lower than this cutoff point, multicollinearity in the explanatory variables was not of concern with this data set.

A four-step hierarchical regression model was tested for each outcome variables. The purpose of these analyses was to examine the amount of unique variance associated with each predictor variable based on an a priori developed sequence. The sequence in which the variables were entered into the regression model was based upon the order in which these skills are typically developed for beginning readers. Variables that were entered at an earlier stage of the analyses were assumed to affect the variables that followed. According to Chall (1983), children typically progress through several stages before becoming fluent readers. During the earliest stages of literacy development, students begin to demonstrate an understanding that spoken words are made up of sounds, which eventually leads to the ability to identify the smallest units of sounds within words, or phonemic awareness. Following the development of phonemic awareness, children learn to associate symbols with sounds and also begin to understand that sounds, when blended together in particular sequences, make up different words. During the next stages, children continue to further develop their fluency and decoding abilities, and eventually, as decoding skills become automatic, they are able to focus their cognitive resources on comprehension. Using this model of reading development, PSF was entered into the first step, followed by NWF, and WIF. Language proficiency level was added to the regression model in the final step to determine whether CELDT level would explain

additional variance in spring outcomes after controlling for student's performance on winter early literacy measures. A dummy variable for language proficiency status was created and the EA/A language proficiency group served as the reference group for each of the models.

ROC Analysis. ROC analysis is a statistical method for exploring the diagnostic accuracy of a test by providing the ratio of true positive/false positive and true negative/false negative decisions. The ROC curve is a graph of sensitivity (y-axis) vs. 1 – specificity (x-axis). ROC curves were generated and area under the curve (AUC) values evaluated to examine the diagnostic accuracy of DIBELS 6th Edition Revised cutoffs for PSF and NWF with respect to spring reading outcomes. The range of cutoffs for PSF, NWF, and WIF were also evaluated to determine statistically optimal scores (optimal ratio of true positive and false positive decisions) with respect to spring outcomes.

The AUC represents the diagnostic accuracy of the instrument, such that values closer to 1 (the ROC curve will near the upper left corner) indicate that the screening measure perfectly distinguished between at-risk and not-at-risk students, while a value of .50 (the ROC curve coincides with the diagonal) suggests that the measure is no better than chance at making correct predictions of at-risk status. An AUC of 0.87 means that a randomly selected student from the ORF at-risk group has a PSF score that is lower than that for a randomly chosen individual from the not-at-risk group 87% of the time (Zweig & Campbell, 1993). AUC values were interpreted using the following scale proposed by Swets (1988): $AUC > .90$ (good), $0.90 > AUC > 0.70$ (useful), and $AUC < 0.70$ (poor). Decision matrices were also created for PSF, NWF, and WIF by each outcome variable

(ORF and WRMT Total Reading-Short Scale). Sensitivity, specificity, positive predictive values, negative predictive values, and hit rates were evaluated.

Materials

Phoneme Segmentation Fluency (PSF). PSF is an individually administered test of PA. DIBELS 6th Edition Revised PSF was administered to all students in the winter and spring of first-grade. As indicated in the DIBELS administration manual, the examiner presented three- and four-phoneme words and asked the student to produce the individual phonemes for each word. Each correctly spoken phoneme was awarded 1 point. For example, if the examiner said the word “mop” and the student responded with, “/m/ /o/ /p/”, the student received 3 points for the word. After the student’s response, the examiner immediately presented the next word. Final PSF scores were determined by counting the number of correctly produced phonemes in 1 min (Good & Kaminski, 2002).

According to the DIBELS administration manual, the alternate-form reliability of PSF is .88 for 2 weeks and .79 for 1 month (Kaminski & Good, 1996). Concurrent criterion validity of PSF with the Readiness Clusters score of the Woodcock-Johnson Psycho-Educational Battery is reported to be .54 in the spring of kindergarten (Good, & Kaminski, 2002). The predictive validity of PSF in the spring of kindergarten with the Total Reading Cluster of the Woodcock-Johnson Psycho-Educational Battery and ORF in the spring of first-grade is reported to be .68. and .62, respectively (Good et al., 2001). Lower concurrent and predictive validity correlations have been found when PSF has been administered in first-grade. The concurrent validity of DIBELS PSF with ORF and

NWF in the middle of first grade has been reported to be .30 and .40, respectively (Burke & Hagan-Burke, 2007). The correlation between PSF administered in the beginning, middle, and end of first-grade with 2nd grade Terra Nova reading assessment scores has been reported to be .26, .18, and .23, respectively (Riedel, 2007).

Nonsense Word Fluency (NWF). NWF assesses children's ability to identify letter-sound correspondences and blend letters into pseudowords. DIBELS 6th Edition Revised NWF was administered to all students in the winter and spring of first-grade. Students were presented with a page of 50 vowel-consonant (e.g., *ov*) and consonant-vowel-consonant pseudowords (e.g., *sig*), and asked to either produce the individual letter sounds constituting the pseudowords or to read the whole pseudowords. For example, for the word "sim" students were allowed to either sound out "/s/ /i/ /m/" or read the whole word, "sim". The final score was determined by counting the number of correctly produced letter sounds in 1 min.

Word Identification Fluency (WIF). WIF was administered to all students in the winter and spring of first-grade. The WIF measure created by Fuchs et al. (2004) was used in this study. This first-grade word list includes 5 columns of 20 isolated words, for a total of 100 high-frequency words. Students were asked to read the words as quickly and carefully as they can. If the student hesitated in reading a word for 2 sec, the student was prompted to try the next word. If the student attempted to sound out a word, the examiner waited 5 sec before prompting the student to try the next word. The final score was determined by counting the number of words read correctly in 1 min.

The 2-week alternate-test form reliability coefficient is .92 and the concurrent validity with the WRMT-R Word Identification Subtest is .77 in the fall of first-grade (Compton, Fuchs, Fuchs, & Bryant, 2006). Predictive validity with the WRMT-R Word Identification Subtest and the Comprehensive Reading Assessment Battery (CRAB) are .63 and .80, respectively in fall of first-grade (Fuchs et al., 2004).

Oral Reading Fluency (ORF). DIBELS 6th Edition Revised ORF was administered to all students in the spring of first-grade. Students were asked to read a grade-level passage of connected text aloud for 1 min during which the examiner recorded the number of errors and the number of words read correctly. If the child delayed longer than 3 sec on a word, the examiner recorded the word as an error and supplied the child with the word. In order to control for the varying degrees of difficulty and content areas of the reading passages, three 1 min passages were administered to the student and the median score was used in the analyses (Good & Kaminski, 2002). Test-retest reliabilities for elementary students range from .92 to .97 and alternate-form reliabilities range from .89 to .94 (Good & Kaminski).

Woodcock Reading Mastery Tests- Revised (WRMT-R). The Woodcock Reading Mastery Tests- Revised- Normative Update Form G (WRMT-R/NU; Woodcock, 1998) is widely used in schools and in reading research to assess students' reading readiness, basic reading, and reading comprehension skills (Salvia & Ysseldyke, 1998; Woodcock). The Word Identification (WI) and Passage Comprehension (PC) subtests of the WRMT-R/NU were administered to all students in the spring of first-grade. All basal and ceiling rules, as listed in the technical manual, were followed during test administration.

The WI test requires students to read as many words as possible from a list consisting of 100 words ordered by difficulty. Students earned 1 point for each correctly pronounced word. The raw score was determined by counting the number of words read correctly. The PC test measures the student's ability to comprehend a short reading passage and identify a key word omitted from the passage. For each blank, the student was asked to supply a word that would be appropriate in the context of that passage. Testing for all students was discontinued after six consecutive errors. The combination of the WI and PC test scores were used to calculate the Total Reading- Short Scale score. The Total Reading-Short Scale score was used as an overall measure of reading ability. The split-half, internal consistency reliability coefficient for the Total Reading-Short Scale is reported to be .98 in first-grade. Concurrent validity of the Total Reading- Short Scale score with respect to the Woodcock Johnson (WJ) Psycho-Educational Battery at first-grade is .61 with WJ Letter-Word Identification, .48 with WJ Word Attack, .78 with WJ Passage Comprehension, and .79 with WJ Total Reading.

California English Language Development Test (CELDT). The CELDT is annually administered to all ELLs by the district. According to the technical report from 2005, test-retest reliability coefficients for the 2004-2005 version of the CELDT were reported to be between 0.86 to 0.90 across all grades and subtests (CTB/McGraw-Hill, 2004). Specifically for grades K-1, test-retest reliability coefficients were .86 for listening/speaking, .88 for reading, and .89 for writing (CTB/McGraw-Hill).

DIBELS Benchmarks. DIBELS 6th Edition Revised emerging/some-risk cutoffs for the winter of first-grade were used to determine "at-risk" status for PSF (< 35) and

NWF (< 50). Students who scored above these cutoffs were classified as “not-at-risk” for the measure (see Table 2). For spring criterion measures, all students who scored below the 25th percentile on the WRMT Total Reading-Short Scale were classified as “below expectations”. For spring ORF, all students who scored below the some-risk criteria for ORF (< 40) were classified as “below expectations”. Students who scored above the 25th percentile on the Total Reading-Short Scale or above the some-risk criteria for ORF (> 40) were classified as “at or above expectations”.

Chapter 4: Results

Descriptive statistics, including the range of scores, means, and standard deviations were examined for all variables included in the analyses. The distribution of scores, including skewness, kurtosis, and outliers, was also evaluated. A series of correlational and multiple regression analyses were conducted to examine the following: (a) the concurrent relationship between winter screening assessments (PSF, NWF, and WIF); (b) the relationship between winter screening measures and spring reading outcomes; and (c) the concurrent relationship between spring reading outcomes, specifically, the relationship between students' performance on fluency and comprehension measures. In addition, the diagnostic accuracy of DIBELS 6th Edition Revised cutoffs for PSF and NWF was evaluated for each spring outcome variable (ORF and WRMT Total Reading-Short Scale). Results are reported for the entire sample, as well as by language proficiency group.

Descriptive Statistics for the Entire Sample

Descriptive statistics for the entire sample are presented in Table 3. In the winter of first-grade, students provided an average of 37.82 phonemes ($SD = 16.26$) on PSF, 68.38 sounds ($SD = 38.32$) on NWF, and 57.07 sight words ($SD = 26.65$) on WIF. By the spring, students provided an average of 44.43 phonemes ($SD = 14.72$), 105.45 sounds ($SD = 51.98$), and 69.68 sight words ($SD = 21.97$). Reading fluency scores ranged from 16 to 171 words per minute, with an average of 90.28 words ($SD = 37.65$), as measured by ORF. The average raw score on the WI subtest of the WRMT was 49.37 words correct ($SD = 14.83$) and the average raw score on the PC subtest of the WRMT was 20.04 ($SD =$

7.67) correct answers. Total Reading Short-Scale scores on the WRMT ranged from standard scores of 70 to 126, with an average of 108.40, which falls within the average range of standard scores.

Of the 102 students who were tested during the winter screening, 7 students scored in the deficit range on PSF ($PSF < 10$), 27 students scored in the emerging range ($10 \leq PSF < 35$), and 68 students scored in the established range ($PSF \geq 35$). On NWF, 16 students scored in the at-risk range ($NWF < 30$) and 20 students scored in the some-risk range ($30 \leq NWF < 49$). A total of 65 students, representing approximately 60% of the students, scored in the low-risk range on NWF ($NWF \geq 50$; see Table 4) during the winter screening.

Spring screening results indicated the following: On spring ORF, 2 students scored in the at-risk range ($ORF < 20$) and 9 students scored in the some-risk range ($20 \leq ORF < 40$; see Table 4). Although 24 students scored in the deficit or emerging range on PSF and 12 students scored in the deficit or emerging range on NWF during the spring screening, approximately 90% of the students (91 students) who were tested scored in the low-risk range on ORF. Among the 24 students who scored in the deficit or emerging range on PSF, 10 students were from the B/EI group, 10 students were from the Intermediate group, and 4 of the students were from the EA/A group. Among the 12 students who scored in the deficit or emerging range on NWF, 8 of these students were from the B/EI language proficiency group and 4 students were from the Intermediate group. No students from the EA/A group scored in the deficit or emerging range on NWF during the spring. Finally, of the 11 students who scored in the at-risk or some-risk

range on spring ORF (ORF < 40), 8 students were from the B/EI English language proficiency group and 3 students were from the Intermediate group. No students from the EA/A group scored in the at-risk or some-risk range on spring ORF. Interestingly, of these 11 students who scored in the at-risk or some risk range on spring ORF, 6 students scored in the established category for PSF (PSF \geq 35) and 4 students scored in the established category for (NWF \geq 50). In other words, these results indicate that the students who scored lowest on ORF during the spring screening did not necessarily score within in the at-risk or some-risk categories on the other spring reading indicators.

Correlational Analyses for the Entire Sample

In order to examine the strength of the relationship between winter and spring measures, Pearson correlation coefficients were calculated for all screening and outcome measures. Of the fluency measures administered in the winter, WIF was most strongly correlated with spring reading performance, as measured by ORF and the WRMT (see Table 5). Winter WIF scores were significantly correlated with performance on spring ORF ($r = .87; p < .01$), WI subtest ($r = .79; p < .01$), PC subtest ($r = .72; p < .01$), and the WRMT Total Reading- Short Scale ($r = .78; p < .01$). Moderate to high correlations were also found between winter NWF scores and performance on spring ORF ($r = .56; p < .01$), WI subtest ($r = .54; p < .01$), PC subtest ($r = .50; p < .01$), and the WRMT Total Reading- Short Scale scores ($r = .53; p < .01$). Smaller correlations were found between winter PSF scores and spring outcome measures. Students' performance on winter PSF was significantly correlated to their performance on spring ORF ($r = .35; p < .01$), the WI

subtest ($r = .36; p < .01$), PC subtest ($r = .28; p < .01$), and the Total Reading- Short Scale score ($r = .34; p < .01$).

An examination of the concurrent relationship between winter screening measures and spring measures indicated the following: Winter PSF scores were significantly correlated with winter NWF and WIF ($r = .37$ and $.36$, respectively; $ps < .01$). Winter NWF and WIF scores were also significantly correlated ($r = .60; p < .01$). Among the spring measures, a small to moderate correlation was found between spring PSF and NWF ($r = .21; p < .05$), as well as between spring PSF and ORF ($r = .21; p < .05$), but a significant correlation was not found between spring PSF and WIF. Spring NWF was significantly correlated with spring WIF ($r = .69; p < .01$) and ORF ($r = .75; p < .01$). Among the spring fluency measures, the largest correlation was found between spring WIF and spring ORF ($r = .86; p < .01$).

On the WRMT, students' performance on the WI subtest was significantly correlated to their comprehension skills, as measured by the PC subtest ($r = .87; p < .01$; see Table 5). In addition, students' spring ORF scores were significantly correlated with their performance on the WI and PC subtests ($r = .87$ and $.82$, respectively; $ps < .01$), as well as their Total Reading- Short Scale Scores ($r = .87; p < .01$). Similarly, students' spring performance on WIF was significantly correlated with their scores on the WI and PC subtests ($r = .84$ and $.78$, respectively; $ps < .01$), as well as their overall reading performance, as measured by the Total Reading- Short Scale Scores ($r = .85; p < .01$).

Descriptive Statistics Aggregated by English Proficiency Level

Descriptive statistics aggregated by English proficiency level are presented in Table 6. Average spring reading fluency scores for students in the B/EI English group ranged from 16 to 139 words per minute, with an average of 65.85 words ($SD = 36.70$), as measured by ORF. WRMT Total Reading Short-Scale scores ranged from standard scores of 70 to 118, with an average of 98.26 ($SD = 14.26$), which falls within the average range of standard scores. For students in the Intermediate group, average spring reading fluency scores ranged from 26 to 171 words per minute, with an average of 92.31 words ($SD = 36.17$). Total Reading Short-Scale scores ranged from standard scores of 89 to 125, with an average of 110.24 ($SD = 8.42$), which falls within the average to high average range of scores. Students who were in the EA/A English proficiency group read between 53 to 170 words per minute on ORF, with an average score of 107.70 words ($SD = 29.71$). Total Reading Short-Scale scores ranged from standard scores of 96 to 126, with an average of 114.36 ($SD = 7.26$), which falls within the average to high average range of scores. These results indicate that students in the EA/A group, on average, scored higher than the B/EI group on spring reading fluency and WRMT Total Reading Short-Scale scores.

Correlational Analyses Aggregated by English Proficiency Level

Of the winter screening measures that were administered, winter WIF was most strongly correlated with spring reading performance for all English language proficiency groups. This is consistent with what was found with the larger sample of students. Correlations between winter WIF scores and spring ORF were $r = .91$, $.84$, and $.77$ (ps

<.01), for students in the B/EI, Intermediate, and EA/A language proficiency groups, respectively. Moderate to large correlations were also found between winter WIF and spring WRMT Total Reading- Short Scale, WI subtest, and PC subtest scores, for all language proficiency groups. Correlations between winter WIF and spring WI ranged from $r = .59$ to $.84$, with the strongest relationship found for students in the B/EI group and the smallest relationship found for students in the EA/A group. Winter WIF and spring PC correlations ranged from $r = .39$ to $r = .80$, again, with the strongest relationship found for students in the B/EI group and the smallest found for students in the EA/A group. Finally, moderate to large correlations were found between winter WIF and spring WRMT Total Reading scores. Correlations ranged from $r = .55$ to $.84$, with the largest correlation found for students in the B/EI group and a moderate correlation found for students in the EA/A group.

Winter NWF was found to be correlated with spring reading performance, however, correlations were smaller in magnitude when compared to winter WIF and only significant for students in the B/EI and Intermediate groups. Correlations between winter NWF and spring outcomes for the B/EI group were as follows: spring ORF ($r = .71, p < .01$), WI subtest ($r = .63, p < .01$), PC subtest ($r = .60, p < .01$), and WRMT Total Reading scores ($r = .62, p < .01$). Among students in the Intermediate group, winter NWF was correlated with spring ORF, WI and PC scores, as well as Total Reading scores with correlations ranging between $r = .54$ to $.59$. Surprisingly, winter NWF was not found to be significantly correlated with any of the spring reading outcomes for students in the EA/A language proficiency group.

Although winter PSF was significantly correlated to spring ORF, WI scores, and WRMT Total Reading- Short Scale Scores for students in the B/EI and Intermediate groups, correlations between winter PSF and spring outcomes measures were smaller in magnitude when compared to winter WIF and NWF, ranging from $r = .37$ to $.51$. In addition, winter PSF was significantly correlated to spring PC scores ($r = .41$, $p < .05$) for students at the B/EI group, but not for the other language proficiency groups. Finally, for students at the EA/A language proficiency group, winter PSF was not correlated with any of the spring outcome measures. In fact, the relationship between winter PSF and spring reading outcomes was negative for students in the EA/A language group

An examination of the concurrent relationship between all winter and spring measures aggregated by English proficiency group indicate the following: Among students in the B/EI and Intermediate English proficiency groups, winter PSF scores were significantly correlated with winter NWF and WIF, with correlations ranging between $r = .44$ to $.62$ ($ps < .01$; see Tables 7 to 9). Winter NWF and WIF scores were also significantly correlated for students in the B/EI and Intermediate English proficiency groups ($r = .72$ and $= .55$, respectively; $p < .01$). Spring PSF was not found to be significantly correlated with spring ORF and spring WIF for the B/EI and Intermediate English proficiency groups. In contrast, significant and large correlations were found between spring NWF, WIF, and ORF, with correlations ranging between $r = .66$ and $.95$ ($p < .01$). Among the spring fluency measures, the largest correlation was found between spring WIF and spring ORF for both the B/EI and Intermediate English proficiency group

($r = .95$ and $.84$, respectively; $p < .01$), which is consistent with what was found with the larger sample.

Among students in the B/EI and Intermediate groups, significant correlations were found between spring fluency measures and spring WRMT subtests. Largest correlations were found between students' spring performance on WIF and WRMT Total Reading Short-Scale scores for both the B/EI ($r = .92$; $p < .01$) and Intermediate groups ($r = .82$; $p < .01$). A significant correlation was also found between spring NWF and WRMT Total Reading Short-Scale scores for both the B/EI ($r = .71$; $p < .01$) and Intermediate groups ($r = .62$; $p < .01$). While significant correlations were not found between students' spring PSF scores and WRMT Total Reading Short-Scale scores for the B/EI group, moderate correlations were found for the Intermediate group ($r = .32$; $p < .05$). Finally, spring ORF scores were significantly correlated with spring Total Reading-Short Scale scores, for both the B/EI ($r = .91$; $p < .01$) and Intermediate group ($r = .83$; $p < .01$).

An examination of the concurrent relationship between winter and spring assessments for the EA/A group was inconsistent with what was found with the B/EI and Intermediate groups, specifically with regards to winter and spring PSF. Winter PSF was not found to be significantly correlated with winter NWF and WIF for students in the EA/A group. In fact, the relationship between winter PSF and the other winter screening measures was found to be negative ($r = -.08$ and $-.24$, for NWF and WIF, respectively) for this group of students. In addition, PSF scores from the spring were not found to be significantly correlated with any of the spring outcome measures (ORF and WRMT). In

contrast, winter NWF and winter WIF scores were significantly and positively correlated ($r = .44; p < .01$) for students in the EA/A group, which is consistent with what was found for the B/EI and Intermediate groups. Spring NWF was also found to be significantly correlated with spring WIF ($r = .70; p < .01$) and ORF ($r = .71; p < .01$). Large correlations were also found between spring WIF and ORF ($r = .72; p < .01$). Consistent with what was found with the other English language proficiency groups, both spring ORF and WIF were also significantly correlated with Total Reading- Short Scale Scores ($r = .84$ and $.58$, respectively; $ps < .01$) for students in the EA/A group.

Hierarchical Linear Regression Analyses

A series of hierarchical regression analyses were conducted to answer the following research question: After controlling for student's performance on winter early literacy measures, does CELDT level explain additional variance in spring outcomes, as assessed by ORF and the WRMT? Results of the hierarchical analysis examining the relationship between winter screening measures and spring ORF are presented in Table 10. PSF was first entered into the model and accounted for a significant proportion of the variance in ORF (12.3%); however, PSF was non-significant after NWF was entered in Step 2. The addition of NWF in step 2 resulted in a large and significant change in R^2 [$F_{(2, 99)} = 25.70, p < .001, \Delta R^2 = .22$] and accounted for approximately 34.2% of the variance in ORF. With the addition of WIF, however, NWF became non-significant. The addition of WIF resulted in a large and significant change in R^2 [$F_{(3, 98)} = 101.20, p < .001, \Delta R^2 = .41$] and accounted for a total of 75.6% of the variance in ORF. After controlling for the contribution of winter screening measures, the addition of language proficiency level in

step 4, with the EA/A group serving as the reference group, did not account for additional variance in ORF.

The final model was statistically significant from zero and accounted for approximately 76.6% of the variance in spring ORF, $F(5, 96) = 62.97, p < .001$. The regression model, $ORF = B_0 + B_1 PSF + B_2 NWF + B_3 WIF + B_4 B/EI + B_5 Intermediate$, was used to yield the following regression equation at step 4, $Y' = 26.98 + .05X_{1i} + .04X_{2i} + 1.12X_{3i} - 10.68X_{4i} - 6.52X_{5i}$. It is important to note, however, that from the winter screening measures, only WIF was significant at the 0.05 level ($t = 12.24, p < .001$) in the final model. Standardized beta values (β) for PSF, NWF, and WIF were .02, .05, and .79, respectively. The dummy coded variable for B/EI language proficiency was also found to be significant at the .05 level ($t = -2.01, p < .05$), indicating that there is a statistically significant difference in spring ORF performance between students in the B/EI group and EA/A group. On average, students from the B/EI group are expected to score 10.68 points lower on spring ORF than students from the EA/A group. The dummy coded variable for the Intermediate group was not statistically significant. This indicates that on average, students from the Intermediate group did not score significantly lower or higher than students from the EA/A group on spring ORF.

The results of the hierarchical analysis examining the relationship between winter screening measures and WRMT-Total Reading Scores are also presented in Table 10. PSF was entered into the model in step 1 and accounted for 11.5% of unique variance in overall WRMT scores, $F(1, 100) = 13.04, p < .001$. The addition of NWF in the second step accounted for approximately 30.5% of the variance in WRMT scores and a significant

change in R^2 [$F(2, 99) = 21.71, p < .001, \Delta R^2 = .19$]. Similar to what was found with ORF as the outcome measure, the addition of WIF resulted in a significant and large change in R^2 [$F(3, 98) = 52.45, p < .001, \Delta R^2 = .31$] and accounted for a total of 61.6% of the variance in ORF. In the final step, the addition of language proficiency scores also resulted in a small, yet significant change in R^2 ($F(5, 96) = 40.67, p < .001, \Delta R^2 = .06$). The final model was statistically significant from zero and accounted for approximately 67.9% of the total variance in WRMT-Total Reading scores.

The regression model, $WRMT\ Total\ Reading\ Score = B_0 + B_1\ PSF + B_2\ NWF + B_3\ WIF + B_4\ B/EI + B_5\ Intermediate$, was used to yield the following prediction equation at step 4, $Y' = 93.299 + .02X_{1i} + .02 X_{2i} + .28 X_{3i} - 8.04 X_{4i} - 1.71 X_{5i}$. Similar to what was found with ORF, of the winter screening measures, only WIF was significant at the 0.05 level ($t = 8.24, p < .001$) in the final model. Standardized beta values (β) for PSF, NWF, and WIF were .03, .06, and .63, respectively. The dummy coded variable for B/EI language proficiency was also found to be significant ($t = -4.14, p < .001$), suggesting that there is a statistically significant difference in spring WRMT-Total Reading Scale performance between the B/EI group and EA/A group. On average, students from the B/EI group are expected to score 8.05 points lower on the WRMT-Total Reading Scale than students from the EA/A group. The dummy coded variable for the Intermediate group was not statistically significant, indicating that on average, students from the Intermediate group did not score significantly different than students from the EA/A group on the WRMT.

Diagnostic Accuracy of DIBELS Cutoffs for Entire Sample

DIBELS emerging/some-risk cutoffs for the winter of first-grade were used for determining at-risk status for PSF (< 35) and NWF (< 50). Students who scored above these cutoffs were classified as “not-at-risk” for the measure. On spring measures, DIBELS emerging/some-risk cutoffs for the spring of first-grade were used for determining at-risk status for PSF (< 35) and NWF (< 50). For spring criterion measures, students who scored below the 25th percentile on the WRMT Total Reading-Short Scale were classified as “below expectations”. On ORF, all students who scored below the some-risk criteria for ORF (< 40) were “below expectations”. Students who scored above the 25th percentile on the Total Reading-Short Scale or above the some-risk criteria for ORF (> 40) were classified as “at or above expectations”.

Using DIBELS cutoffs, decision matrices were also created for PSF and NWF with respect to each outcome variable (ORF and WRMT Total Reading-Short Scale). When examining the entire sample of students ($N = 102$), 2 students scored in the at-risk range ($ORF < 20$) and 9 students scored in the some-risk range ($20 \leq ORF < 40$) on spring ORF, for a total of 11 students in the “below expectations” category. For spring WRMT, a total of 8 students fell in the “below expectations” category and 94 students scored “at or above expectations”.

Sensitivity values for winter PSF and NWF with respect to both ORF and WRMT scores were above 73% and specificity values were between 68% - 71% (see Figures 3 to 6). Hit rates were between 70- 72%, however, positive predictive values, were much lower than expected, ranging between 17% - 26%. In other words, many students who

scored within the “at-risk” range for PSF and NWF in the winter scored “at or above expectations” on ORF in the spring. During the winter screening, a total of 34 students scored within the at-risk range for PSF. However, of this group of students, only 8 scored “below expectations” and 26 of these students scored “at or above expectations” on spring ORF, which resulted in a positive predictive value of 24%. Similarly, of the 39 students who scored within the at-risk range for winter NWF, only 10 of these students scored “below expectations” and 29 scored “at or above expectations”, resulting in a positive predictive value of 26%. Positive predictive values for PSF and NWF with respect to spring WRMT were 17% and 19%, respectively.

Negative predictive values were above 95% for all tests (see Figures 3 to 6). Among the 68 students who scored within the low-risk range for winter PSF (> 35), 65 of these students scored “at or above expectations” on spring ORF, and 3 students scored below expectations on ORF, which resulted in a 95% negative predictive value. Similarly, among the 63 students who scored within the low-risk range for winter NWF (> 50), 62 of these students scored “at or above expectations” on spring ORF and 1 student scored below expectations on ORF, which resulted in a negative predictive value of 98%. Negative predictive values for PSF and NWF with respect to spring WRMT were 97% and 98%, respectively.

ROC curves were generated and AUC values were evaluated to examine the accuracy of PSF and NWF in predicting each of the spring reading outcomes. Using Swet’s (1988) criteria, AUC values for winter PSF with respect to ORF (AUC = .82) and WRMT (AUC = .80) fell within the “useful” range. AUC values for winter NWF with

respect to ORF (AUC = .93) and WRMT (AUC = .92) fell within the “good” range. The range of scores for PSF and NWF were also evaluated to determine cutoffs (optimal ratio of true positive and false positive decisions) with respect to spring outcomes.

Specifically, optimal cutoffs for PSF and NWF were determined to be the point at which there was the smallest difference between sensitivity and specificity. Results indicated that lowering the cutoffs for both measures to 28 for PSF and 37 for NWF, with respect to both outcomes, resulted in more accurate predictions with respect to both ORF and WRMT (see Table 15). According to Rathvon (2004), sensitivity and specificity indices should be at least 75-80% in order for the measure to conform to reliability standards for screening measures. This standard was met for NWF, as the lower cutoff resulted in sensitivity and specificity values greater than .82 with respect to both spring outcomes. However, for PSF, using a cutoff of 28 to predict ORF scores resulted in a sensitivity index of .73 and a specificity index of .85. Sensitivity and specificity indices for PSF in predicting spring WRMT scores were .75 and .83, respectively.

Diagnostic Accuracy of WIF

AUC values for winter WIF with respect to spring ORF (AUC = .97) and spring WRMT (AUC = .98) were in the “good” range. Currently, the DIBELS screening battery does not include a word identification fluency measure. As a result, the range of scores for WIF was evaluated to determine optimal cutoffs (optimal ratio of true positive and false positive decisions) with respect to spring outcomes. Based on an analysis of the coordinates of the ROC curve, optimal WIF cutoffs were determined with respect to each of the spring outcomes. A cutoff of 26 correct words on winter WIF to predict spring

ORF, and a cutoff of 21 correct words to predict spring WRMT resulted in hit rates of .93 and .94, respectively. Sensitivity and specificity rates were greater than .92 and negative predictive values were 1.0 with respect to both outcome measures. Positive predictive values for WIF were .61 and .57 for ORF and WRMT, respectively.

Diagnostic Accuracy of DIBELS Cutoffs for B/EI Group

Decision matrices for students in the B/EI group with respect to each outcome variable (ORF and WRMT Total Reading-Short Scale) were also examined (see Tables 11 to 14). Decision matrices were not examined for the Intermediate and EA/A groups, as only a total of 3 students from the Intermediate group and no students from the EA/A performed below expectations on spring ORF and WRMT. When examining the B/EI group, 8 students scored “below expectations” on ORF and 19 students scored “at or above expectations”. With respect to WRMT scores, 7 B/EI students scored “below expectations” and 20 students scored “at or above expectations”. Sensitivity values for winter PSF and NWF for the B/EI group with respect to spring outcomes were above 71%. Specificity values were between 55% - 58%. Hit rates ranged between 59% - 67%.

Positive predictive values were low for the B/EI group, ranging between 34% - 47%, but were higher than the values found for the larger sample. During the winter screening, a total of 14 students scored within the at-risk range for PSF. From this group of students, 6 students scored “below expectations” on spring ORF and 8 students scored “at or above expectations”, resulting in a positive predictive value of 43% for winter PSF with respect to spring ORF. The positive predictive value for winter NWF with respect to

spring ORF was 47%. Positive predictive values for PSF and NWF with respect to spring WRMT were 34% and 40%, respectively.

Negative predictive values were at or above 85% for all tests with the B/EI group. Among the 13 students who scored within the low-risk range for winter PSF (> 35), 11 of these students scored “at or above expectations” on spring ORF and 2 students scored below expectations on ORF, resulting in a negative predictive value of 85%. Similarly, among the 12 students who scored within the low-risk range for winter NWF (> 50), 11 of these students scored “at or above expectations” on spring ORF and 1 student scored below expectations on ORF, resulting in a negative predictive value of 92%. Negative predictive values for PSF and NWF with respect to spring WRMT were 85% and 92%, respectively.

Chapter 5: Discussion

While a growing body of research has demonstrated that many of the same reading performance indicators that are used with native English speakers can improve outcomes for ELLs, current demographics indicate that ELLs represent a diverse group with students with varying home languages and degrees of English proficiency (California Department of Education, 2011b; Kindler, 2002). As a result, students' language background must be carefully considered when selecting, administering, and interpreting test performance. In addition, as recommended by the *Standards*, it is important to examine the validity of assessment tools and test norms with students from different linguistic subgroups (AERA, APA, & NCME, 1999). A recent review of the current literature indicated that there were no published studies that had examined the use of early literacy assessments with Korean ELLs. Consequently, the primary aim of this study was to examine the predictive validity of widely used early literacy screening measures with 1st grade Korean ELLs.

This discussion will address the results of the research questions and summarize the key findings from this study. First, students' progress between the winter and spring screenings will be reviewed. This will be followed by a summary regarding the concurrent and predictive validity of DIBELS screening measures, including a review of which measures accounted for robust variance in spring reading outcomes, followed by a discussion regarding the relationship between fluency and comprehension measures, and ending with an evaluation of the results of the diagnostic accuracy analyses. This section

will conclude by presenting some of the major limitations of this study, as well as potential applications to educational practice.

Concurrent and Predictive Validity of Screening Tools

Examination of PSF, NWF, and WIF scores during the winter and spring screenings indicated that the students, as a group, had demonstrated growth in phonemic awareness, alphabetic knowledge, and word identification skills, respectively, between the screening periods. Average ORF scores in the spring also showed that many of the students had achieved high levels of fluency by the end of first-grade. Specifically, spring reading fluency scores ranged from 16 to 171 words per minute, with an average of 90.28 words. Of the 102 students who were tested in the spring, only 11 students scored in the at-risk or some-risk range on ORF based on DIBELS cutoffs.

Consistent with prior studies, (e.g., Compton, Fuchs, Fuchs, Bouton, et al., 2010; Deno et al., 1982; Fien et al., 2008; Lesaux & Siegel, 2003; Vanderwood et al., 2008) moderate to high correlations were found for winter NWF and WIF scores with respect to spring outcomes for the larger sample, as well as for the B/EI and Intermediate groups. Although winter PSF was also correlated to spring outcomes for these groups, correlations were smaller in magnitude when compared to winter NWF and WIF. Interestingly, very different results were found for the EA/A group. While moderate correlations were found between winter WIF and spring reading measures for the EA/A group, which is consistent with what was found with the other groups, a significant correlation between winter NWF and spring outcomes was not found. In addition, a

negative relationship was found between winter PSF and spring outcomes for the EA/A group.

When examining the concurrent relationship between measures for the larger sample, as well as the B/EI and Intermediate groups, moderate correlations were found between NWF and WIF during both the winter and spring screenings. Moderate to large correlations were also found for spring WIF and NWF with respect to spring ORF. Although winter PSF was correlated to winter NWF and WIF, spring PSF was not found to be related to spring reading fluency and comprehension scores for these groups. Once again, differences were found when examining the concurrent validity of screening measures with the EA/A language proficiency group. While significant correlations were found between NWF, WIF, and ORF during both the winter and spring screenings for the EA/A group, a negative correlation was found for winter PSF with respect to both winter NWF and WIF, as well as spring PSF with respect to spring NWF and WIF.

Interestingly, during the winter screening, 34 students out of the total who were screened scored in the deficit or emerging range on PSF, and 36 students scored in the at-risk or some-risk range on NWF. Furthermore, during the spring screening, 24 students continued to score in the deficit or emerging range on PSF, and 12 students scored in the deficit or emerging range on NWF. Although many students scored below the cutoffs for winter, as well as spring PSF and NWF, 90% of the students attained adequate levels of reading fluency on spring ORF as evidenced by their low-risk scores on this measure. Of the 11 students who scored lowest on spring ORF, 6 of these students scored within the established category for spring PSF, and 4 students scored within the established

category for spring NWF. Furthermore, of the 8 students who scored “below expectations” on the WRMT, 4 of these students scored within the established category for PSF, and 2 students scored within the established category for spring NWF. In other words, students who scored lowest on ORF and WRMT during the spring screening did not necessarily score within the at-risk range on the other spring reading indicators.

These results clearly indicate that while phonemic awareness and alphabetic knowledge are prerequisite early literacy skills that should lead to fluent reading, these skills may not be sufficient for students to fluently read and comprehend text. Another factor that plays an important role in reading development and should be carefully examined is vocabulary knowledge. An examination of students’ vocabulary development is especially important when examining the reading progress of English language learners. While native English speakers have already learned 5,000-7,000 words before they begin formal reading instruction (Biemiller & Slonin, 2001), limited English vocabulary may pose another significant challenge for ELLs who are learning how to read. The results of this study indicate that while phonemic awareness and alphabetic knowledge are critical skills that are taught to beginning readers, the acquisition of these skills may not necessarily lead students to become good readers. Other important factors including vocabulary knowledge should also be examined with ELLs, as these skills may also play a significant role in students’ reading development.

While it may be difficult to generalize the results of this study to groups outside of this study, as the overall number of participants in this study and the number of students within each language proficiency group was small, these results provide some

preliminary data regarding the predictive and concurrent validity of literacy screening measures with students representing varying language proficiency levels. It appears that language proficiency may play a significant role when examining the validity of these measures with ELLs. In particular, among students who were at the higher levels of English proficiency, PA appeared to be unrelated to students' performance on pseudoword reading, word identification, fluency, and reading comprehension tasks. While there are numerous studies that have found phonological awareness to be strongly related to students' reading acquisition (e.g., Cunningham & Stanovich, 1997; Stanovich, 1986; Wagner & Torgesen, 1987), the results of the present study do not support these past findings. In fact, during the winter and spring of first-grade, PA skills did not appear to be a prerequisite to reading, as many of the students who scored within the at-risk and emerging range on this measure were able to read fluently and accurately, as measured by ORF.

These results are consistent with the prior findings by Riedel (2007) who examined the predictive validity of PSF with first-grade, predominantly native-English speakers. In this study, middle and end-of-the year PSF scores were found to predict first- and second-grade comprehension at a rate that was only slightly better than chance. By the winter of first-grade, it appeared that many students in Riedel's study, including students who were behind their peers in reading development, had mastered the lower level skills required by PSF, making the measure less likely to distinguish between good and poor readers. This appears to be the case in the present study, as many students who had mastered the skills required by PSF, but nevertheless continued to perform in the at-

risk range on ORF and WRMT. Several interesting observations were also made during both the winter and spring screening with regards to the administration of PSF with high performing readers. First, although sample items were reviewed with all students, several of the students who were already known to be fluent readers immediately attempted to spell out the words. In fact, some of these students appeared to have more difficulty with segmenting words into phonemes than with spelling the words. Other students were observed repeating words or providing rhyming words rather than segmenting the words.

It is important to note that with DIBELS measures, emphasis is placed on whether a student is above or below a cutoff at a specified time rather than his/her absolute performance on the measure. For example, when determining whether a student is at-risk or not-at-risk, a score of 40 is no different from a score 60 on PSF, as both scores are within outside the at-risk range for the winter of first-grade. It can also be assumed that both students have established PA skills and it is also quite possible that each of these students may be approaching a growth asymptote. As a result, these students may show little or no growth in PSF scores during the winter and spring screenings. In contrast, when PSF is used as a quantitative predictor of later reading skills and its relationship to ORF is examined, the difference between a score of 40 and 60 may greatly influence the magnitude of results. Although the results of this study provide valuable data regarding the predictive and concurrent validity of early literacy measures, these results should be interpreted with caution as PSF was examined as a quantitative predictor for parts of this study. It also appears that increased emphasis should be placed on examining the accuracy of DIBELS cutoffs in predicting future reading performance, as these criteria

are currently being used across schools throughout the country to identify students who need intervention services (University of Oregon Center on Teaching and Learning, 2009).

According to Paris (2005), it is also important to carefully consider differences in the developmental trajectories of reading skills, including skill onset, duration of acquisition, and asymptotic levels of performance when examining the predictive and concurrent validity of measures. For example, letter-names and letter-sound correspondences are constrained skills because the number of elements to be mastered is relatively small and finite. As a result, the duration of acquisition is brief and the trajectory of mastery is steep. Skills may also have transitory importance to future reading performance. In other words, the predictive validity of measures may be short-lived. For example, the magnitude of the correlation between letter knowledge and later reading proficiency, while important among younger children, may not be as correlated to reading performance during a period when many students have already mastered this skill. Furthermore, while some reading skills may not be mastered perfectly, they may approach a growth asymptote as acquisition slows or a ceiling is attained. For example, ceiling effects have often been observed with letter naming tasks when administered to students at the end of kindergarten and first-grade (Speece, Mills, Ritchey, & Hillman, 2003; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). All of these factors must be taken into account when examining relationships between early literacy measures, as they may greatly influence the significance of the data, the magnitude of the correlations, as well as the interpretation of the results.

In the present study, it appears that the predictive validity of PSF may have been short-lived and that many students, including both poor readers and good readers, may have approached a growth asymptote on this measure by the winter of first-grade. This appears to have also influenced the magnitude of the correlations between PSF and spring reading performance measure. Interestingly, following the proposal of this project and the first year of data collection, an updated version of DIBELS, known as DIBELS Next, was released with revised screening tools and updated benchmark goals. According to the University of Oregon Center on Teaching and Learning (2011), several revisions were made including the removal of PSF from the 1st grade winter and spring screening battery.

Summary of Regression Analyses

In order to examine the contribution of winter measures of PA, alphabetic principles, and word identification, as well as the language proficiency level to spring reading performance, a series of hierarchical regression models were tested for each of the reading outcomes (ORF and WRMT Total Reading- Short Scale score). Results of the regression models further highlighted the importance of WIF in predicting reading outcomes. When examining the entire sample, each of the winter screening variables, when entered independently, were found to account for a significant and unique proportion of the variance in spring ORF and WRMT scores. However, results of hierarchical models with PSF, NWF, and WIF entered respectively, indicated that the addition of WIF resulted in a large and significant change in R^2 with respect to both spring ORF and WRMT scores. Final models accounted for a total of 75.6% of the

variance in spring ORF and 60.4% of the variance in WRMT-Total Reading scores. However, in the final model, of the winter screening measures, only winter WIF was found to be significant at the 0.05 level with respect to both spring ORF and WRMT scores.

After controlling for the contribution of winter screening measures, the addition of language proficiency level in the final step did not account for additional variance in ORF scores, but was found to add a small, yet significant change in R^2 with WRMT scores. Based on the current sample, on average, students from the B/EI group were expected to score lower on the WRMT than students from the EA/A. Differences in performance between the Intermediate and EA/A groups were not found to be statistically significant. These results are not surprising given that WRMT scores are also not only a reflection of the child's decoding skills, but also his/her ability to comprehend text, which may be impacted by limited English vocabulary and proficiency.

It was not surprising to find that winter WIF, a measure assessing a student's ability to quickly read isolated words from a list of high-frequency words, was not only highly correlated with spring outcomes, but also accounted for a significant and unique proportion of the variance in spring ORF and WRMT scores. This is consistent with earlier findings (e.g., Compton et al., 2010; Deno et al., 1982; Fuchs et al., 2004) regarding the utility of WIF as a screening tool. Specifically, when compared to winter PSF and NWF, this 1 min screening measure was found to be most correlated with end-of-the year reading fluency and comprehension for the larger sample, as well all language proficiency groups. Concurrent validity correlations between spring WIF with respect to

spring ORF and WRMT scores were also statistically significant for the larger sample, as well as all language proficiency groups.

Relationship Between Fluency and Comprehension

While there has been some criticism that ORF is solely a measure of speed reading and that it may even penalize students who search for meaning within text (Goodman, 2006; Samuels, 2007), there have been numerous studies with both native English speakers and with ELLs that have found moderate to strong correlations between ORF and measures of reading comprehension (e.g., Baker & Good, 1995; Fuchs et al., 2001; Good & Kaminski, 2002; Shinn et al., 1992). Consistent with these past findings, results of the present study also found a strong relationship between students' fluency and comprehension. When examining the performance of the larger sample, students' spring ORF scores were significantly correlated with their performance on the WRMT PC subtest, as well as their Total Reading- Short Scale Scores. Furthermore, when examining this relationship by language proficiency group, ORF scores were significantly correlated with PC subtest scores for the B/EI, Intermediate, as well as the EA/A groups. Significant correlations were also found between ORF scores and Total Reading Scores for all language proficiency groups, with the strongest correlation found for the B/EI group. In the present study, ORF was found to be a robust measure for assessing overall reading proficiency and was significantly related to students' comprehension.

Examination of Diagnostic Accuracy

In order for tests to have high rates of accuracy in differentiating students who are at-risk from students who are not at-risk, appropriate cutoffs must be established such

that sensitivity and specificity are maximized. While there have been many studies that have found DIBELS to be related to later reading outcomes (Good & Kaminski, R, 2002), the results of the present study also highlight the importance of examining the appropriateness of DIBELS cutoffs with ELLs, as these are the criteria that are currently used by schools to identify at-risk students.

While AUC values for PSF and NWF fell within the “useful” and “good” range, using Swet’s (1988) criteria, an examination of specificity, sensitivity, and hit rates provided additional information regarding the appropriateness of DIBELS cutoffs for 1st grade Korean ELLs. While the overall effectiveness of the screening assessments and DIBELS cutoffs, as measured by hit rates, was between 70- 72%, and negative predictive values were above 95% for all tests, positive predictive values were much lower than expected, ranging between 17% - 26%. In other words, many students who scored within the “at-risk” range for PSF and NWF in the winter scored “at or above expectations” on ORF and the WRMT in the spring.

Based on an examination of ROC coordinates, lowering DIBELS cutoffs for PSF and NWF to 28 correct phonemes and 37 correct letter sounds, respectively, were found to result in more accurate predictions with respect to both ORF and WRMT. Using these cutoffs, sensitivity and specificity values for both NWF and PSF were above 75% with respect to spring WRMT scores. Sensitivity and specificity values were also above 75% for NWF with respect to spring ORF scores, however using the lower PSF cutoff to predict spring ORF yielded a sensitivity index of .73 and a specificity index of .85. In other words, PSF failed in its attempts to accurately predict spring ORF performance over

a range of cutoffs. While PSF was able to adequately predict true negatives, the percentage of students predicted to be poor readers who actually become poor readers (true positives) was lower than expected. In fact, using a range of cutoffs, PSF was found to over-identify students who were likely to demonstrate difficulties in reading fluency. The results of the present study are consistent with an earlier study by Hintze, Ryan, and Stoner (2003) that examined the diagnostic accuracy of PSF in predicting performance on the Comprehensive Test of Phonological Processing (CTOPP). While results showed moderate to strong correlations between the DIBELS PSF and CTOPP scores, using a range of cutoffs for PSF (10 to 34 phonemes) also resulted in the over-identification of students who were likely to demonstrate phonological awareness problems that were not corroborated by true CTOPP scores.

When comparing the diagnostic accuracy of winter PSF, NWF, and WIF in predicting spring outcomes, the results of the current study found highest sensitivity, specificity, and hit rates for WIF. Among the winter screening measures, highest AUC values were also found for winter WIF with respect to spring ORF and spring WRMT. Based on an analysis of the range of scores for WIF, optimal cutoffs were determined with respect to spring outcomes. Sensitivity and specificity rates using these cutoffs were greater than .92 and negative predictive values were 1.0 with respect to both outcomes. However, similar to what was found with PSF and NWF, positive predictive values for WIF were .61 and .57 for ORF and WRMT, respectively. In other words, if a student did not meet the winter WIF cutoff, there was a 61% chance that the student would truly be at-risk on ORF and a 39% chance that the student would be above-expectations on ORF.

With respect to the WRMT, if the student did not meet the winter cutoff for WIF, there was a 57% chance that the student would truly be at-risk and a 43% chance that the student would be above-expectations. However, if a student met the winter WIF cutoff, there was a 100% chance that this student was going to be above expectations on ORF and WRMT.

Limitations

By far, the greatest challenge of this study was gaining access to schools and students. This resulted in a significantly smaller sample size than what was anticipated during the planning phases of this study. While a primary aim of this study was to examine the predictive validity of early literacy measures with Korean ELLs representing varying levels of English proficiency, due to the limited number of participants gathered, language proficiency groups were small in number and unequal in size. It is also important to note that the decision to combine the Early Advanced and Advanced English proficiency students was based on state guidelines for reclassification, as the state recommendation is that CELDT scores be at either the Early Advanced or Advanced level for students to be considered for reclassification. In contrast, the decision to combine the Beginning and Early Intermediate language proficiency groups was solely based on the limited number of students at the Beginning level of English proficiency.

A second limitation of this study is that the initial screening, which was projected to take place in the fall (September/October), was pushed back to the winter following difficulties with gaining access to schools. Additional components of this study were adjusted due to difficulties with gaining access to schools, as well as limited personnel.

For example, interobserver agreement data was calculated on 20% of the probes administered in the winter and 20% of the probes administered in the spring during the first year of data collection only. Interobserver agreement data was not gathered during the second year of data collection.

Another limitation of this study is that a second English proficiency measure was not able to be administered to be compared with CELDT language proficiency classifications. While psychometric properties for the CELDT have been reported to be adequate, a significant limitation of English language proficiency (ELP) assessments is that there is currently no consensus among researchers regarding the nature of language proficiency and how to best measure it. The nature of language proficiency has been understood by some researchers as consisting of 64 separate language components, whereas others have suggested that one global factor accounts for the majority of the variance in language proficiency test scores (Cummins, 1984).

One final area that was unable to be addressed by this study was the relationship between students' levels of Korean literacy and English reading outcomes. Despite the considerable differences in orthography and phonological structures, studies have found evidence of cross-linguistic transfer of PA with Korean-English bilingual kindergarten students (Kim, 2008). Although specific data regarding students' Korean literacy skills was not able to be collected, informal interviews with several of the students indicated that many of them were in the process of also learning how to read in their home language either through a Korean-language school or with their parents. In fact, according to a study by Zhou and Kim (2007), approximately 500 Korean-language

schools were registered in the Korean School Association of America (KSAA) by the end of the 1980s, and by 2005, the number of registered Korean-language schools in the Los Angeles area alone stood at 254. In their study, they also found that over half of the Korean youth interviewed attended a Korean-language school for some time during their primary-school years. This is an important factor to consider because while phoneme-level awareness has been found to be predictive of reading development for English speakers, due to the high salience of syllables in the Korean script, both syllable- and phoneme-level awareness have been found to predict Korean reading acquisition (Cho and McBride-Chang, 2005). In the present study, students' exposure to reading instruction in two different languages may have played an important role in how students learned to read in English.

Potential Applications to Educational Practice

In order to accurately identify students who are at-risk, it is critical to not only identify valid and reliable screening measures, but also to establish cutoffs, specifically for ELLs, that result in true positives approaching 100% while minimizing the number of false negatives. The consequences of failing to identify at-risk children far outweigh the consequences of over-identification, such that if these children are not identified, they will fail to receive additional support. While a high rate of false positives may not be as great of a concern, it is nevertheless important to find tools that also minimize the number of false positives so that resources are not allocated to children who are not truly at-risk for poor reading outcomes.

The results of the current study clearly indicate that relying on DIBELS cutoffs with Korean ELLs may inaccurately identify these students. In fact, schools with large numbers of ELLs that are currently utilizing DIBELS cutoffs to determine at-risk status should carefully consider examining the appropriateness of these cutoffs with their ELLs. The findings from the present study are also supported by the results of an earlier, large-scale study by Johnson, Jenkins, Petschcer, and Catts (2009). This study examined the diagnostic accuracy of beginning-of-the-year DIBELS measures in predicting end-of-the-year reading performance for 12,055 first-grade students comprised of both native English speakers and ELLs. When comparing native English speakers and ELLs, results of this study indicated that while the strongest predictors for the full sample were also the best predictors for ELLs, cutoffs that produced at least 90% sensitivity rates differed for ELLs and native-English speakers. While an ORF cutoff of 21 resulted in 90% sensitivity and 62% specificity for native English speakers, a lower ORF cutoff of 16 was needed to yield the minimum 90% sensitivity, but resulted in a specificity index of only 45%, indicating a greater rate of false positives. These results have very important implications, such that if cutoffs that are used with native English speakers are used with ELLs, ELLs will have a greater chance of being misidentified. Again, these results emphasize the importance of examining students' language background when examining the validity of test norms and interpreting test performance.

While there have been several studies have found measures of pseudoword reading to be predictive of reading skills for ELLs (e.g., Fien et al., 2008; Lesaux & Siegel, 2003; Vanderwood et al., 2008), there have been far fewer studies that have

examined the predictive validity of WIF as an early literacy screening tool. Furthermore, while earlier versions of DIBELS, as well as the most recent update, DIBELS Next, include NWF as a screening measure for students in kindergarten through second-grade, a WIF measure is currently not part of the DIBELS screening battery. Although the results of this study are difficult to generalize to students outside of this sample, the present data clearly indicate that WIF administered in the winter of first-grade is strongly predictive of students' end-of-the-year reading fluency and comprehension. In order to maximize the diagnostic accuracy of screening measures, schools should examine the utility of WIF in conjunction with other DIBELS measures with a larger sample of first-grade ELLs. The results of the present study also suggest that the incorporation of additional key pieces of data, including students' English language proficiency status, as well as their vocabulary knowledge (Riedel, 2007) may improve screening accuracy.

Instead of relying solely on assessment scores to identify at-risk status, additional information may also need to be considered to make more accurate decisions. According to Berliner (2004), while screening and progress monitoring tools provide objective assessments of students' reading progress, teachers' judgments may also provide additional data towards evaluating students' academic performance. Teachers' perceptions influence daily instructional decisions, such as student groupings, curricula, instructional strategies, and educational placement. Recent studies by Begeny, Krouse, Groce, and Mann (2011) and Martin and Shapiro (2011) have found moderate relationships between teachers' judgments of students' reading performance and students' actual reading abilities. Interestingly, both studies also found that some teachers

also made inaccurate judgments, at times overestimating students' reading abilities. In fact, Begeny et al. found that teachers were better at judging the reading performance of high-performing readers than low- or average-performing readers. In contrast, Martin and Shapiro (2011) found that teachers made more accurate judgments for lower-achieving students than typical-achieving students. Based on these findings, it appears that while teachers' perceptions alone may overestimate or underestimate students' actual performance, the use of this information in conjunction with data from screening measures may provide educators with a more accurate model for differentiating at-risk and not-at-risk students.

Another possible strategy for improving the diagnostic accuracy of screening models is to utilize a multi-gated screening approach. Compton, Fuchs, Fuchs, Bouton, et al. (2010) examined a two-step screening approach to identify at-risk status with a group of first-grade students. The aim of this study was to identify measures that when added to a first-grade screening model would lower the rate of false positives and increase screening efficiency. In the first stage, a standardized high frequency word list was presented for 1 min during which a child was asked to read as many words as he/she could. In the second stage, children who failed the initial screening were assessed with additional measures to discriminate true positives from false positives. Results indicated that the multi-step screening approach significantly decreased the number of false positives.

Finally, the high rate of false positives in this study also supports the importance of progress monitoring within an RtI model. With the current sample, several of the

students were found to be at-risk for PSF and NWF during the winter and spring screenings. Despite their at-risk status on these measures, many of these students were found to score within the low-risk range on spring ORF. Within an RtI model, students who are found to be at-risk during universal screening would be provided with more intensive support. During the intervention, these students would be regularly monitored to determine whether they are benefiting from the intervention. The incorporation of progress monitoring would also allow teachers to use multiple data points to make adjustments to the intervention, as well as to identify those students from the group who may have been misidentified.

As a result of the growing interest and implementation of RtI models, DIBELS is currently being used by schools throughout the country as a universal screening tool. According to the University of Oregon Center on Teaching and Learning (2009), more than 15,000 schools have adopted DIBELS as of 2009. Although past studies have reported that many of the same reading indicators and assessments tools that have been used to screen native English speakers can be used with ELLs (e.g., Baker & Good, 1995; Haagar & Windmueller, 2001), the results of this present study indicate that there is clearly a need for the continued development of appropriate assessment tools and decision rules that are sensitive to the unique cultural and linguistic backgrounds of ELLs.

References

- Abedi, J. (2006). Psychometric issues in the ELL assessment and special education eligibility. *Teachers College Record, 108*, 2282-2303.
- AERA, APA, & NCME. (1999). Testing individuals of diverse linguistic backgrounds. In AERA/APA/NCME (Ed.), *Standards for educational and psychological testing* (pp. 91-97). Washington, D.C.: American Educational Research Association.
- Aidinis, A., & Nunes, T. (2001). The role of different levels of phonological awareness in the development of reading and spelling in Greek. *Reading and Writing: An Interdisciplinary Journal, 14*, 145-177.
- Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L., et al. (2004). Examining the incremental benefits of administering a maze and three vs. one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review, 33*, 218 – 233.
- Artiles, A. J., Rueda, R., Salazar, J., & Higarreda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children, 71*, 283-300.
- August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth*. Retrieved from http://www.cal.org/projects/archive/nlpreports/executive_summary.pdf
- Baker, S. K., & Good, R. H. (1995). Curriculum-based measurement of English reading with bilingual Hispanic students: A validation study with second-grade students. *School Psychology Review, 24*, 561-578.
- Barnes, J. S., & Bennett, C. E. (2002). *The Asian population: 2000. Census 2002 Brief*. Retrieved from Census Bureau website: <http://www.census.gov/prod/2002pubs/c2kbr01-16.pdf>
- Begeny, J.C., Krouse, H.E., Groce, M.K., & Mann, C.M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review, 40*, 23-38.
- Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology & Society, 24*, 200-212.

- Blachman, B. A. (1994). What we have learned from longitudinal studies of phonological processing and reading, and some unanswered questions: A response to Torgesen, Wagner, and Rashotte. *Journal of Learning Disabilities, 27*, 287-291.
- Burke, M. D. & Hagan-Burke, S. (2007). Concurrent criterion-related validity of early literacy indicators for middle of first grade. *Assessment for Effective Intervention, 32*, 66-77.
- Butler, F. A., & Stevens, R. (2001). Standardized assessment of the content knowledge of English language learners K-12: Current trends and old dilemmas. *Language Testing, 18*, 409-427.
- California Department of Education, Educational Demographics Office (2011). *Languages of English learner students. Los Angeles Unified School District 2009-2010*. Retrieved from http://www.eddata.k12.ca.us/App_Resx/EdDataClassic/fsTwoPanel.aspx?#!bottom =/_layouts/EdDataClassic/profile.asp?Tab=1&level=06&reportnumber=16&county=19&district=64733#englishlearners
- California Department of Education (2011a). *California English Language Development Test- CalEd Facts*. Retrieved from California Department of Education website: <http://www.cde.ca.gov/ta/tg/el/cefceldt.asp>
- California Department of Education (2011b). *CELDT Number and Percent of Students at Each Overall Performance Level*. Retrieved from California Department of Education website: [http://dq.cde.ca.gov/dataquest/CELDT/results.aspx?year=2010 2011&level= state&assessment=2&subgroup=1&entity=](http://dq.cde.ca.gov/dataquest/CELDT/results.aspx?year=2010%202011&level=state&assessment=2&subgroup=1&entity=)
- Chafouleas, S. M., Lewandowski, L. J., Smith, C. R., & Blachman, B. A. (1997). Phonological awareness skills in children: Examining performance across tasks and ages. *Journal of Psychoeducational Assessment, 15*, 334-347.
- Chall, J. S. (1983). *Stages of reading development*. New York, New York: McGraw-Hill.
- Cho, J. R., & McBride-Chang, C. (2005). Correlates of Korean Hangeul acquisition among kindergartners and second graders. *Scientific Studies of Reading, 9*, 3-16.
- Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K. Barquero, L. A.,...Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102*, 327-340.
- Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*, 394-409.

- CTB/McGraw-Hill. (2005). *California English Language Development Test 2004-2005 Form D Technical Report*. Monterey, CA: Author.
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49, 222–251.
- Cummins, J. C. (1984). *Bilingual and special education: Issues in assessment and pedagogy*. Austin, TX: Pro-Ed.
- Daly, E. J., Chafouleas, S. M., & Skinner, C. H. (2005). *Interventions for reading problems: Designing and evaluating effective strategies*. New York, NY: The Guilford.
- D'Angiulli, A., Siegel, L. S., & Serra, E. (2001) The development of reading in Italian and English in bilingual children. *Applied Psycholinguistics*, 22, 479-507.
- Davidson, F. (1994). Norms appropriate of achievement tests: Spanish-speaking children and English children's norms. *Language Testing*, 11, 83-95.
- Deno, S. L., Mirkin, R. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children*, 49, 36-45.
- Durgunoglu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology*, 85, 453-465.
- Durgunoglu, A. Y. & Öney, B. (1999). A cross-linguistic comparison of phonological awareness and word recognition. *Reading and Writing: An Interdisciplinary Journal*, 11, 281-299.
- Dynamic Measurement Group, (2011). *DIBELS Next Technical Manual*. Retrieved from DIBELS Next website: <https://dibels.org/next/download.php?agree=2>
- Fien, H., Baker, S. K., Smolkowski, K., Smith, J. L., Kame'enui E. J., & Beck, C. T. (2008). Using nonsense word fluency to predict reading proficiency in kindergarten through second grade for English learners and native English speakers. *School Psychology Review*, 37, 391-408.
- Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology*, 90, 37-55.
- Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review*, 21, 45-58.

- Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*, 7–21.
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239-256.
- Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*, 20-28.
- Genesee, F., E. Geva, C. Dressler & M. L. Kamil (2006). Synthesis: Cross-linguistic relationships. In D. August & T. Shanahan (eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-minority Children and Youth* (pp. 153-183). New York: Routledge.
- Gersten, R., Baker, S. K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades: A practice guide*. Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Geva, E., Yaghoub-Zadeh, Z., & Schuster, B. (2000). Understanding individual differences in word recognition skills of ESL children. *Annals of Dyslexia, 50*, 123-154.
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Education Achievement.
- Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, L., et al. (2004). *Adequacy and Decision Making Utility of DIBELS* (Technical Report No. 7). Eugene, OR: University of Oregon.
- Good, R. H., Simmons, D.C., & Kame'enui, E.J. (2001). The importance and decision making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*, 259-290.
- Good, R. H., Simmons, D.C, Kame'enui, E. J, Kaminski, R. A., & Wallin, J. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade* (Technical Report No. 11). Eugene, OR: University of Oregon.

- Good, R. H., Simmons, D. C., & Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *Educational and Child Psychology, 15*, 56-70.
- Goodman, K. S. (2006). A critical review of DIBELS. In K. S. Goodman (Ed.), *The truth about DIBELS: What it is, what it does* (pp. 1-39). Portsmouth, NH: Heinemann.
- Goswami, U. (2000). Phonological Representations, Reading Development and Dyslexia: Towards a Cross-Linguistic Theoretical Framework. *Dyslexia, 6*, 133-151.
- Goswami, U., & Mead, F. (1992). Onset and rime awareness and analogies in reading. *Reading Research Quarterly, 27*, 153-162.
- Goswami, U., Ziegler, J. & Richardson, U. (2005). The effects of spelling consistency on phonological awareness: A comparison of English and German. *Journal of Experimental Child Psychology, 92*, 345-365.
- Green, S.B. (1991) How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26*, 499-510.
- Haager, D. S., & Windmueller, M. P. (2001). Early reading intervention for English language learners at-risk for learning disabilities: Student and teacher outcomes in an urban school. *Learning Disabilities Quarterly, 24*, 235-250.
- Hamilton, C., & Shinn, M. R. (2003). Characteristics of word callers: An investigation into the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review, 32*, 228 – 240.
- Harn, B. A., Stoolmiller, M, & Chard, D. J. (2008). Measuring the dimensions of alphabetic principle on the reading development of first graders: The role of automaticity and unitization, *Journal of Learning Disabilities, 41*, 143-157.
- Healy, K., Vanderwood, M., & Edelston, D. (2005). Early literary interventions for English language learners: support for an RTI model. *The California School Psychologist, 10*, 55-63.
- Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent Validity and Diagnostic Accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review, 32*, 541-556.

- Høien, T., Lundberg, I., Stanovich, K.E. & Bjaalid, I. K. (1995). Components of phonological awareness, *Reading and Writing: An Interdisciplinary Journal* 7: 171–188.
- Holmes, D., & Duron, S., (2000). *LEP Students and High-Stakes Assessment*. Retrieved from [The National Clearinghouse for Bilingual Education \(NCBE\) website: http://www.ncela.gwu.edu/files/rcd/BE021069/LEP_Students_and_High_Stakes.pdf](http://www.ncela.gwu.edu/files/rcd/BE021069/LEP_Students_and_High_Stakes.pdf)
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127-160.
- Hopstock, P. J., & Stephenson, T. G. (2003). *Special Topic Report #1: Native languages of LEP students*. Retrieved from National Clearinghouse for English Language Acquisition website: http://www.ncela.gwu.edu/files/rcd/BE021196/native_languages1.pdf
- Johnson, E. S., Jenkins, J. R., Petscher, Y., Catts, H. W. (2009). How Can We Improve the Accuracy of Screening Instruments? *Learning Disabilities Research & Practice*, 24, 174-185.
- Juel, C. (1988). Learning to read and write: a longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*.
- Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, 25, 215-227.
- Kang, H., & Simpson, G. B. (1996). Development of semantic and phonological priming in a shallow orthography. *Developmental Psychology*, 32, 860–866.
- Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–85). Amsterdam: Elsevier.
- Kim, Y. S. (2007). Phonological awareness and literacy skills in Korean: An examination of the unique role of body-coda units. *Applied Psycholinguistics*, 28, 67–93.
- Kim, Y. S. (2008). Crosslinguistic influence on phonological awareness for Korean-English bilingual children. *Reading and Writing*.

- Kindler, A. L. (2002). *Survey of the states' limited English proficient students and available educational programs and services: 2000-2001 summary report*. Retrieved from National Clearinghouse for English Language Acquisition website: http://www.ncele.gwu.edu/files/rcd/BE021853/Survey_of_the_States.pdf
- Klingner, J. K., Artiles, A. J., & Barletta, L. M. (2006). English language learners who struggle with reading: Language acquisition or LD? *Journal of Learning Disabilities, 39*, 108-128.
- Klingner, J. K., & Edwards, P. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly, 41*, 108-117.
- Knofczynski, G., & Mundfrom, D. (2008). Sample sizes when using multiple linear regression for prediction. *Educational and Psychological Measurement, 68*, 431-442.
- Korean American Coalition-Census Information Center. (2003). *Korean population in the United States, 2000*. Retrieved from California State University, Center for Korean American and Korean Studies website: http://www.calstatela.edu/centers/ckaks/census/7203_tables.pdf
- Langdon, H. W. (1989). Language disorder or difference? Assessing the language skills of Hispanic students. *Exceptional Children, 56*, 160-167.
- Lesaux, N. K. (2006). Building consensus: Future directions for research on English language learners at risk for learning difficulties. *Teachers College Record, 108*, 2406 – 2438.
- Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Findings from a 5-year longitudinal study. *Journal of Educational Psychology, 99*, 821-834.
- Lesaux, N. K. & Siegel, L. S. (2003). The Development of Reading in Children Who Speak English as a Second Language. *Developmental Psychology, 39*, 1005-1019.
- Linquanti, R., & George, C. (2007). Establishing and utilizing an NCLB Title III accountability system: California's approach and findings to date. In J. Abedi (Ed.), *English language proficiency assessment in the Nation: Current status and future practice* (pp. 105-120). Davis, CA: The Regents of the University of California.
- Martin, S. D. & Shapiro, E. S. (2011). Examining the accuracy of teachers' judgments of DIBELS performance. *Psychology in the Schools, 48*, 343–356.

- McGuinness, D. (2004). *Early Reading Instruction: What Science Really Tells Us about How to Teach Reading*. Cambridge, MA: MIT Press.
- McLaughlin, B. (1992). *Myths and misconceptions about second language learning: What every teacher needs to unlearn*. Washington, DC: National Center for Research on Cultural Diversity and Second Language Learning.
- Migration Policy Institute. (2010). *Top languages spoken by English language learners nationally and by state*. Retrieved from Migration Policy Institute website: http://www.migrationinformation.org/ellinfo/FactSheet_ELL3.pdf
- National Center for Education Statistics. (2009). *Reading 2009. National Assessment of Educational Progress at grades 4 and 8*. Retrieved from National Assessment for Educational Progress website: http://nationsreportcard.gov/reading_2009/nat_g4.asp
- National Clearinghouse for English Language Acquisition. (2011). *The growing numbers of English learner students 1998/99-2008/09*. Retrieved from National Clearinghouse for English Language Acquisition website: http://www.ncela.gwu.edu/files/uploads/9/growingLEP_0809.pdf
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Retrieved from National Reading Panel website: <http://www.nationalreadingpanel.org/Publications/summary.htm>
- National Research Council. (2002). *Minority students in special and gifted education. Committee on Minority Representation in Special Education*. Retrieved from http://www.nap.edu/catalog.php?record_id=10128
- Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, 40, 184-202.
- Pedhazur, E. P. (1997). *Multiple regression in behavioral research: Explanation and prediction*. Fort Worth, TX: Harcourt Brace.
- Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.
- Perfetti, C. A., & Dunlap, S. (2008). Learning to read: General principles and writing system variations. In K. Koda & A. Zehler (Eds.). *Learning to read across languages* (pp. 13-38). Mahwah, NJ: Erlbaum.
- Perfetti, C. A., & Liu, Y. (2005). Orthography to phonology and meaning: Comparisons across and within writing systems. *Reading and Writing*, 18, 193-210.

- Quiroga, T., Lemos-Britton, Z., Mostafapour, E., Abbott, R. D., & Berninger, V. W. (2002). Phonological awareness & beginning reading in Spanish-speaking ESL first graders: Research into practice. *Journal of School Psychology, 40*, 85-111.
- Rathvon, N. (2004). *Early Reading Assessment: A Practitioner's Handbook*. New York: Guilford.
- Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first grade students. *Reading Research Quarterly, 42*, 546-567.
- Salvia, J., & Ysseldyke, J.E. (1998). *Assessment (7th ed.)*. Boston: Houghton Mifflin Company.
- Samuels, S. J. (2007). Commentary: The DIBELS tests: Is speed of barking at print what we mean by reading fluency? *Reading Research Quarterly, 42*, 563-566.
- Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96*, 265-282.
- Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*, 459-479.
- Snow, C. E., Burns, S. M., & Griffin, P. (Eds.). (1998). *Preventing Reading Difficulties in Young Children*. Washington, D.C.: National Academy Press.
- Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *Journal of Special Education, 36*, 223-233.
- Stahl, S. A., & Murray, B. A. (1994). Defining PA and its relationship to early reading. *Journal of Educational Psychology, 86*, 221-234.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.
- Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38*, 175-190.

- Swanson, L. H., Trainin, G., Necochea, D. M., & Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research, 73*, 407–440.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240*, 1285–1293.
- Taylor, I. (1980). The Korean writing system: An alphabet? A syllabary? A logography? In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language 2* (pp. 67–82). New York: Plenum Press.
- Taylor, I., & Taylor, M. M. (1995). *Writing and Literacy in Chinese, Korean, and Japanese*. Philadelphia: John Benjamins.
- University of Oregon Center on Teaching and Learning. (2008). *General Information about DIBELS*. Retrieved from DIBELS website: <https://dibels.uoregon.edu/docs/dibelsinfo.pdf>
- University of Oregon Center on Teaching and Learning. (2009). *DIBELS data system brochure*. Retrieved from DIBELS website: https://dibels.uoregon.edu/resources/data_system/brochure/dds_brochure_color_low.pdf
- University of Oregon Center on Teaching and Learning. (2011). *DIBELS Next vs. 6th Edition for DDS*. Retrieved from DIBELS website: https://dibels.uoregon.edu/docs/DIBELS_6thvsNextfortheDDS.pdf
- U.S. Department of Education & National Institute of Child Health and Human Development. (2003). *National symposium on learning disabilities in English language learners: Symposium summary*. Retrieved from U.S. Department of Education website: <http://www2.ed.gov/about/offices/list/osers/products/ld-ell/ld-ell.pdf>
- Vanderwood, M. L., Linklater, D., & Healy, K. (2008). Predictive accuracy of nonsense word fluency for English language learners. *School Psychology Review, 37*, 5-17.
- Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., et al. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal, 107*, 153-180.
- Vellutino, F. R., & Scanlon, D.M. (1987). Phonological coding, phonological Awareness and reading ability: Evidence from a longitudinal and experimental Study. *Merrill-Palmer Quarterly, 33*, 321-363.

- Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1998). The case for early intervention in diagnosing specific reading disability. *Journal of School Psychology, 36*, 367-397.
- Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). The development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85*, 1-20.
- Wang, M., Park, Y., & Lee, K. (2006). Korean-English biliteracy acquisition: Cross-language phonological and orthographic transfer. *Journal of Educational Psychology, 98*, 148-158.
- Wetherill, G.B. (1986). *Regression analysis with applications*. Chapman and Hall, London.
- Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26*, 207-14.
- Wimmer, H., Landerl, K., Linortner, R. & Hummer, P. (1991). The relationship of phonemic awareness to reading acquisition: More consequence than precondition but still important. *Cognition, 40*, 219-249.
- Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests—Revised Normative Update: Examiner's Manual*. Circle Pines, MN: American Guidance Service.
- Yi, K. O. (1998). The internal structure of the syllable in Korean: Onset–rime or body-coda? *Korean Journal of Experimental and Cognitive Psychology, 10*, 67–83.
- Zehler, A., Fleischman, H., Hopstock, P., Stephenson, T., Pendzick, M., & Sapru, S. (2003). *Policy report: Summary of findings related to LEP and SPED-LEP students*. Retrieved from http://onlineresources.wnylc.net/pb/orcdocs/LARC_Resources/LEPTopics /ED/DescriptiveStudyofServicestoLEPStudentsandLEPStudentswithDisabilities.pdf
- Zhou, M. & Kim, S. (2007, May). After-school institutions in Chinese and Korean immigrant communities: A model for others? *Migration Information Source*. Retrieved from <http://www.migrationinformation.org>
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3-29.

Table 1. Predictive Accuracy Table

Spring Criterion Measure (ORF, WRMT)	Predictor Variable (PSF, NWF)		Indices of Predictive Accuracy
	At-Risk	Not At-Risk	
Below Expectations	True Positive (TP)	False Negative (FN)	Sensitivity $TP / (TP + FN)$
At or Above Expectations	False Positive (FP)	True Negative (TN)	Specificity $TN / (TN + FP)$
Total	Positive Predictive Value $TP / (TP + FP)$	Negative Predictive Value $TN / (TN + FN)$	Hit Rate $(TP + TN) / (TP + FN + TN + FP)$

Note. NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency; WRMT = Woodcock Reading Mastery Tests; TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative.

Table 2. Winter and Spring DIBELS 6th Edition Revised Benchmarks for First-Grade

Winter & Spring PSF	Deficit	0-9
	Emerging	10-34
	Established	35 and above
Winter & Spring NWF	At-Risk	0-29
	Some Risk	30-49
	Low Risk	50 and above
Spring ORF	At-Risk	0-19
	Some Risk	20-39
	Low Risk	40 and above

Note. PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency.

Table 3. Descriptive Statistics of Measures (N = 102)

Measure	Range	<i>M</i>	<i>SD</i>
Winter			
PSF1	0- 76	37.82	16.26
NWF1	4- 204	68.38	38.32
WIF1	2- 106	57.07	26.65
Spring			
PSF2	0- 82	44.03	14.72
NWF2	15- 264	105.45	51.98
WIF2	8- 113	69.68	21.97
Criterion Measures			
ORF2	16- 171	90.28	37.65
WI	5- 72	49.37	14.83
PC	2- 32	20.04	7.67
TR	70- 126	108.40	11.76

Note. PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; WIF= Word Identification Fluency; ORF= Oral Reading Fluency; WI= Word Identification Subtest; PC = Passage Comprehension Subtest; TR = Total Reading Score.

Table 4. Results of Winter and Spring Screening (N = 102)

Winter Screening Results

	Deficit (< 10)	Emerging ($10 \leq \text{PSF} < 35$)	Established (≥ 35)
PSF	$N = 7$	$N = 27$	$N = 68$
	At-risk (< 30)	Some-Risk ($30 \leq \text{NWF} < 50$)	Low-Risk (≥ 50)
NWF	$N = 16$	$N = 21$	$N = 65$

Spring Screening Results

	Deficit (< 10)	Emerging ($10 \leq \text{PSF} < 35$)	Established (≥ 35)
PSF	$N = 1$	$N = 23$	$N = 78$
	Deficit (< 30)	Emerging ($30 \leq \text{NWF} < 50$)	Established (≥ 50)
NWF	$N = 5$	$N = 7$	$N = 90$
	At-risk (< 20)	Some-Risk ($20 \leq \text{ORF} < 40$)	Low-Risk (≥ 40)
ORF	$N = 2$	$N = 9$	$N = 91$

Note. PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency.

Table 5. Intercorrelations of Measures for Entire Sample (N = 102)

Measure	PSF1	NWF1	WIF1	PSF2	NWF2	WIF2	ORF2	WI	PC	TR
1. PSF1	—	.37**	.36**	.59**	.22*	.34**	.35**	.36**	.28**	.34**
2. NWF1	.37**	—	.60**	.19	.43**	.53**	.57**	.54**	.50**	.53**
3. WIF1	.36**	.60**	—	.19	.69**	.84**	.87**	.79**	.72**	.78**
4. PSF2	.59**	.19	.19	—	.21*	.17	.21*	.24*	.21*	.23*
5. NWF2	.22*	.43**	.69**	.21*	—	.69**	.75**	.65**	.60**	.65**
6. WIF2	.34**	.53**	.84**	.17	.69**	—	.86**	.84**	.78**	.85**
7. ORF2	.35**	.57**	.87**	.21*	.75**	.86**	—	.87**	.82**	.87**
8. WI	.36**	.54**	.79**	.24*	.65**	.84**	.87**	—	.87**	.99**
9. PC	.28**	.50**	.72**	.21*	.60**	.78**	.82**	.87**	—	.93**
10. TR	.34**	.53**	.78**	.23*	.65**	.85**	.87**	.99**	.93**	—

Note. ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; WIF = Word Identification Fluency; WI = Word Identification Subtest; PC = Passage Comprehension Subtest; TR = Total Reading Score.

* $p < .05$ ** $p < .01$

Table 6. Descriptive Statistics of Measures for B/EI (N = 27), Intermediate (N = 42), and EA/A Groups (N = 33)

Measure	Range			<i>M</i>			<i>SD</i>		
	B/EI	Int	EA/A	B/EI	Int	EA/A	B/EI	Int	EA/A
Winter									
PSF1	0- 76	0-57	0-70	33.04	36.21	43.82	19.98	13.84	14.33
NWF1	4- 143	10-124	23-204	51.37	66.67	84.46	35.44	31.51	42.75
WIF1	2- 106	7-104	29-102	40.77	59.91	66.80	29.01	26.13	18.73
Spring									
PSF2	0- 82	10-73	12-63	43.37	42.91	47.24	18.99	13.69	11.77
NWF2	28-173	15-264	53-221	82.46	107.33	121.88	40.82	56.00	49.25
WIF2	8- 113	30-100	55-103	56.70	72.05	77.28	28.10	19.35	13.92
Criterion Measures									
ORF2	16- 139	26-171	53-170	65.85	92.31	107.70	36.70	36.17	29.71
WI	5-61	20-72	35-72	37.07	51.62	56.58	18.15	10.59	9.75
PC	2-27	7-32	9-32	13.89	20.88	24.00	7.88	6.67	5.41
TR	70- 118	89-125	92-126	98.26	110.24	114.36	14.26	8.42	7.26

Note. PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; WIF= Word Identification Fluency; ORF= Oral Reading Fluency; WI= Word Identification Subtest; PC = Passage Comprehension Subtest; TR = Total Reading Score.

Table 7. Intercorrelations of Measures for B/EI Group (N = 27)

Measure	PSF1	NWF1	WIF1	PSF2	NWF2	WIF2	ORF2	WI	PC	TR
1. PSF1	—	.62**	.49**	.52**	.36	.43*	.51**	.44*	.41*	.44*
2. NWF1	.62**	—	.72**	.09	.47*	.66**	.71*	.63**	.60**	.62**
3. WIF1	.49**	.72**	—	.22	.71**	.93**	.91**	.84**	.80**	.84**
4. PSF2	.52**	.09	.22	—	.22	.20	.22	.26	.34	.28
5. NWF2	.36	.47*	.71**	.22	—	.75**	.72**	.73**	.64**	.71**
6. WIF2	.43**	.66**	.93**	.20	.75**	—	.95**	.92**	.83**	.92**
7. ORF2	.51*	.71**	.91**	.22	.72**	.95**	—	.92*	.87**	.91**
8. WI	.44*	.63**	.84**	.26	.73**	.92	.92**	—	.88**	.99**
9. PC	.41*	.60**	.80**	.35	.63**	.83**	.87**	.88**	—	.93**
10. TR	.44*	.63**	.84**	.28	.71**	.92**	.91**	.99**	.93**	—

Note. ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; WIF = Word Identification Fluency; WI = Word Identification Subtest; PC = Passage Comprehension Subtest; TR = Total Reading Score.
 * $p < .05$ ** $p < .01$

Table 8. Intercorrelations of Measures for Intermediate Group (N = 42)

Measure	PSF1	NWF1	WIF1	PSF2	NWF2	WIF2	ORF2	WI	PC	TR
1. PSF1	—	.46**	.44**	.58**	.27	.39*	.40**	.42**	.25	.37*
2. NWF1	.46**	—	.55**	.37*	.42**	.44**	.54**	.59**	.41**	.54**
3. WIF1	.44**	.55**	—	.27	.71**	.77**	.84**	.74**	.67**	.74**
4. PSF2	.58**	.37*	.27	—	.34*	.23	.25	.40**	.18	.32*
5. NWF2	.27	.42**	.71***	.34*	—	.66**	.75**	.62**	.59**	.65**
6. WIF2	.39*	.44**	.77**	.23	.66**	—	.84**	.79**	.78**	.82**
7. ORF2	.40**	.54**	.84**	.25	.75**	.84**	—	.82**	.75*	.83*
8. WI	.43**	.59**	.74**	.40**	.62**	.79**	.82**	—	.82**	.97**
9. PC	.25	.41**	.67**	.18	.59**	.78**	.75**	.82**	—	.92**
10. TR	.37*	.54**	.74**	.32*	.65**	.82**	.83**	.97**	.92**	—

Note. ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; WIF = Word Identification Fluency; WI = Word Identification Subtest; PC = Passage Comprehension Subtest; TR = Total Reading Score.
 * $p < .05$ ** $p < .01$

Table 9. Intercorrelations of Measures for EA/A Group (N = 33)

Measure	PSF1	NWF1	WIF1	PSF2	NWF2	WIF2	ORF2	WI	PC	TR
1. PSF1	—	-.08	-.24	.71**	-.17	-.21	-.20	-.21	-.27	-.23
2. NWF1	-.08	—	.44**	.03	.28	.38*	.31	.23	.30	.25
3. WIF1	-.24	.44*	—	-.10	.56**	.66**	.77**	.59**	.39*	.55**
4. PSF2	.71	.03	-.10	—	-.05	-.11	.03	-.04	-.02	-.05
5. NWF2	-.17	.28	.56**	-.05	—	.70**	.71**	.63**	.45**	.62**
6. WIF2	-.21	.38*	.66**	-.11	.70**	—	.72**	.58**	.48**	.58**
7. ORF2	-.20	.31	.77**	.03	.71**	.72**	—	.86**	.69**	.84**
8. WI	-.21	.23	.59**	-.04	.63**	.58**	.86**	—	.79**	.98**
9. PC	-.27	.30	.39*	-.02	.45**	.48**	.69**	.79**	—	.89**
10. TR	-.23	.25	.55**	-.05	.62**	.58**	.84**	.98**	.89**	—

Note. ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; WIF = Word Identification Fluency; WI = Word Identification Subtest; PC = Passage Comprehension Subtest; TR = Total Reading Score.
 * $p < .05$. ** $p < .01$

Table 10. Hierarchical Multiple Regression Analyses Predicting Spring ORF and WRMT Performance From Winter Early Literacy Screening Measures (N = 102)

		ORF				
		R^2	ΔR^2	Final β	Final B	F
Step 1:	PSF	.12**	.12**	.02	.05	14.01**
Step 2:	NWF	.34**	.22**	.05	.04	25.70**
Step 3:	WIF	.76**	.41**	.79**	1.12**	101.20**
Step 4:	B/EI			-.13*	-10.68*	
	Int			-.09	-6.52	
	Lang Prof	.77	.01			62.97**
		WRMT- Total Reading				
		R^2	ΔR^2	Final β	Final B	F
Step 1:	PSF	.11**	.12**	.03	.02	13.04**
Step 2:	NWF	.31**	.19**	.06	.02	21.71**
Step 3:	WIF	.62**	.31**	.63**	.28**	52.45**
Step 4:	B/EI			-.30**	-8.05**	
	Int			-.07	-1.71	
	Lang Prof	.68**	.06**			40.67**

Note. PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; WIF= Word Identification Fluency; ORF= Oral Reading Fluency; WRMT = Total Reading Score ; B/EI = Beginning/Early Intermediate Dummy Variable; Int = Intermediate Dummy Variable

* $p = .05$ ** $p < .001$

Table 11. Predictive Accuracy of PSF With Respect to Spring ORF for B/EI Students (N= 27)

Spring Criterion Measure (ORF)	Predictor Variable (PSF)		Indices of Predictive Accuracy
	At-Risk PSF < 35	Not At-Risk PSF ≥ 35	
Below Expectations N = 8	TP N = 6	FN N = 2	Sensitivity .75
At or Above Expectations N = 19	FP N = 8	TN N = 11	Specificity .58
Total	PPV .43	NPV .85	Hit Rate .67

Note. PPV = Positive Predictive Value; NPV = Negative Predictive Value; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency; WRMT = Woodcock Reading Mastery Tests.

Table 12. Predictive Accuracy of NWF With Respect to Spring ORF for B/EI Students (N= 27)

Spring Criterion Measure (ORF)	Predictor Variable (NWF)		Indices of Predictive Accuracy
	At-Risk NWF < 50	Not At-Risk NWF ≥ 50	
Below Expectations N = 8	True Positive N = 7	False Negative N = 1	Sensitivity .88
At or Above Expectations N = 19	False Positive N = 8	True Negative N = 11	Specificity .56
Total	PPV .47	NPV .92	Hit Rate .67

Note. PPV = Positive Predictive Value; NPV = Negative Predictive Value; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency; WRMT = Woodcock Reading Mastery Tests.

Table 13. Predictive Accuracy of PSF With Respect to Spring WRMT for B/EI Students (N= 27)

Spring Criterion Measure (WRMT < 25%)	Predictor Variable (PSF)		Indices of Predictive Accuracy
	At-Risk PSF < 35	Not At-Risk PSF ≥ 35	
Below Expectations N = 7	True Positive N = 5	False Negative N = 2	Sensitivity .71
At or Above Expectations N = 20	False Positive N = 9	True Negative N = 11	Specificity .55
Total	PPV .34	NPV .85	Hit Rate .59

Note. PPV = Positive Predictive Value; NPV = Negative Predictive Value; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency; WRMT = Woodcock Reading Mastery Tests.

Table 14. Predictive Accuracy of NWF With Respect to Spring WRMT for B/EI Students (N= 27)

Spring Criterion Measure (WRMT < 25%)	Predictor Variable (NWF)		Indices of Predictive Accuracy
	At-Risk NWF < 50	Not At-Risk NWF ≥ 50	
Below Expectations N = 7	True Positive N = 6	False Negative N = 1	Sensitivity .86
At or Above Expectations N = 20	False Positive N = 9	True Negative N = 11	Specificity .55
Total	PPV .40	NPV .92	Hit Rate .63

Note. PPV = Positive Predictive Value; NPV = Negative Predictive Value; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency; WRMT = Woodcock Reading Mastery Tests.

Table 15. Optimal PSF and NWF DIBELS cutoffs to Predict "Below Expectations" Performance on Spring ORF and WRMT

PSF	Outcome	Sens	Spec	PPV	NPV	HR
DIBELS Cutoff- 35	ORF	.73	.71	.24	.95	.72
Optimal Cutoff- 28	ORF	.73	.85	.36	.96	.83
DIBELS Cutoff- 35	WRMT	.75	.70	.17	.97	.71
Optimal Cutoff- 28	WRMT	.75	.83	.27	.98	.83
NWF	Outcome	Sens	Spec	PPV	NPV	HR
DIBELS Cutoff- 50	ORF	.91	.68	.26	.98	.71
Optimal Cutoff- 37	ORF	.82	.88	.45	.98	.85
DIBELS Cutoff- 50	WRMT	.88	.68	.19	.98	.70
Optimal Cutoff- 37	WRMT	.87	.84	.32	.98	.84

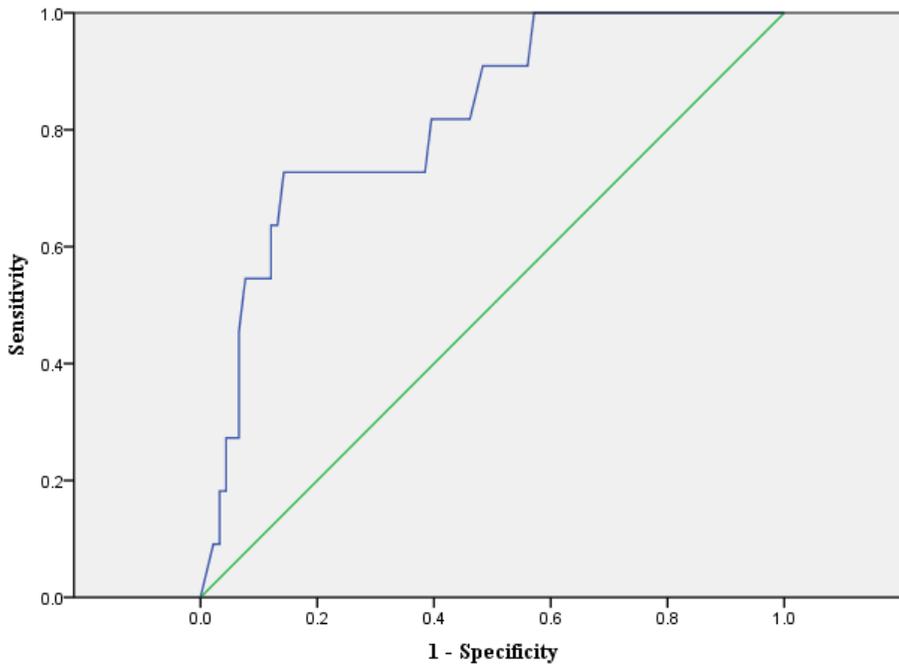
Figure 1. Examples of Korean words with their English pronunciation and translation. Symbols are arranged from left-to-right and top-to-bottom to form square blocks that correspond to single syllables.

English word	horse
Hangul Word	말
Hangul Syllables	말
Hangul Phonemes	ㅁ ㅏ ㄹ
Pronunciation	m a l
English word	market
Hangul Word	시장
Hangul Syllables	시 장
Hangul Phonemes	ㅅ ㅣ ㅈ ㅏ ㅇ
Pronunciation	s i j a ng
English word	lunch box
Hangul Word	도시락
Hangul Syllables	도 시 락
Hangul Phonemes	ㄷ ㅕ ㅅ ㅣ ㄹ ㅏ ㄱ
Pronunciation	d o s i r a g

Figure 2. The 14 consonants and 10 vowels in Hangul.

Consonants	ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
	/g/	/n/	/d/	/l,r/	/m/	/b/	/s/	/ŋ/	/j/	/ch/	/k/	/t/	/p/	/h/
Vowels	ㅏ	ㅑ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ				
	/a/	/ya/	/ǒ/	/yǒ/	/o/	/yo/	/u/	/yu/	/ü/	/i/				

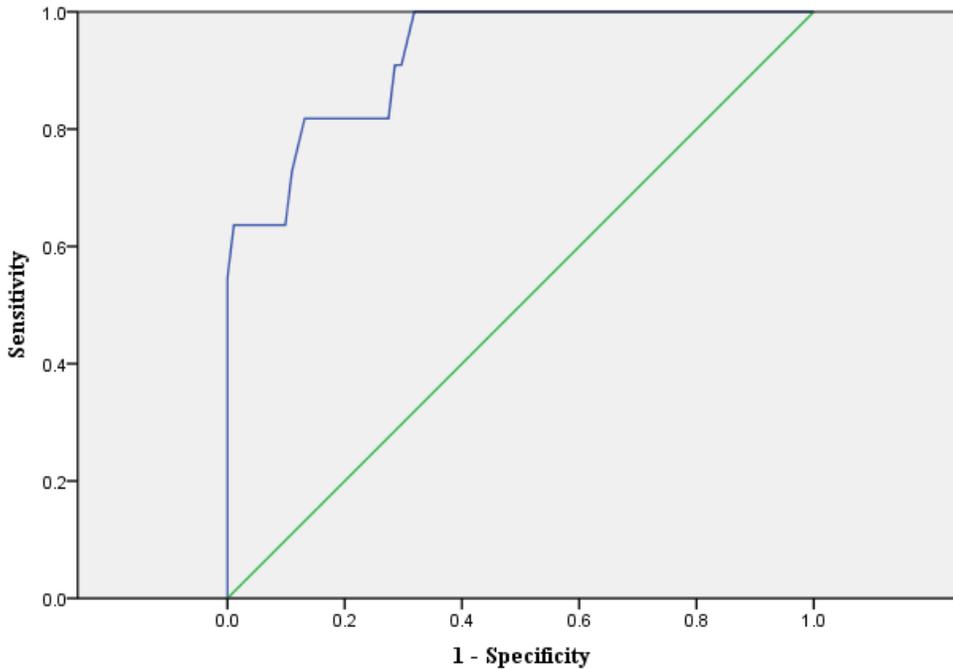
Figure 3. ROC Curve: Winter PSF DIBELS Benchmarks to Predict "Below Expectations" Performance on Spring ORF (N = 102)



Spring Measure (ORF < 40)	At-Risk (PSF < 35)	Not At-Risk (PSF ≥ 35)	Indices of Predictive Accuracy
Below Expectations N = 11	TP N = 8	FN N = 3	Sensitivity .73
At or Above Expectations N = 91	FP N = 26	TN N = 65	Specificity .71
AUC = .82	PPV .24	NPV .95	Hit Rate .72

Note. AUC = Area Under the Curve; PPV = Positive Predictive Value; NPV = Negative Predictive Value; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency.

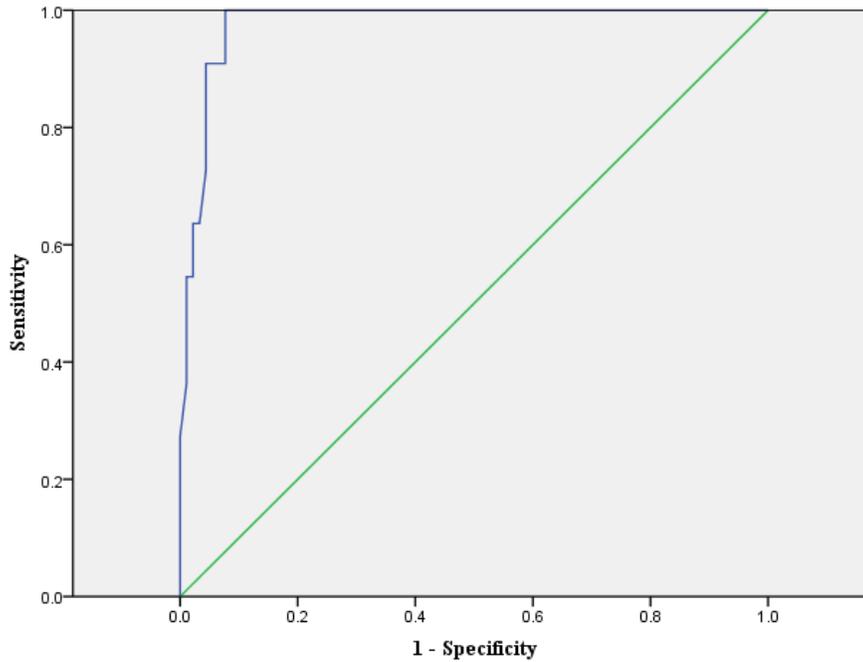
Figure 4. ROC Curve: Winter NWF DIBELS Benchmarks to Predict "Below Expectations" Performance on Spring ORF (N = 102)



Spring Measure (ORF < 40)	At-Risk (NWF < 50)	Not At-Risk (NWF ≥ 50)	Indices of Predictive Accuracy
Below Expectations N = 11	True Positive N = 10	False Negative N = 1	Sensitivity .91
At or Above Expectations N = 91	False Positive N = 29	True Negative N = 64	Specificity .68
AUC = .93	PPV .26	NPV .98	Hit Rate .71

Note. AUC = Area Under the Curve; PPV = Positive Predictive Value; NPV = Negative Predictive Value; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency.

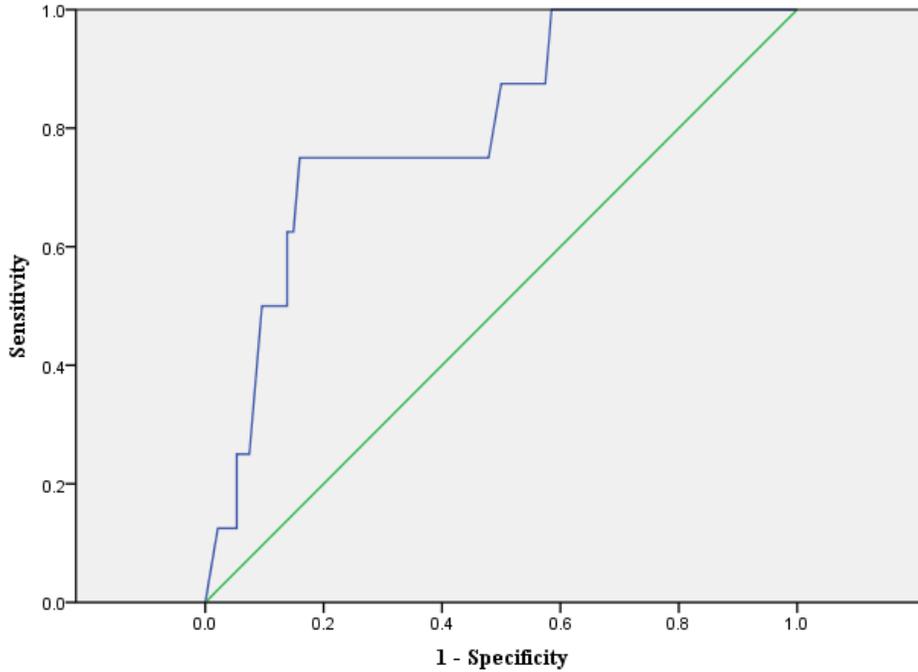
Figure 5. ROC Curve: Winter WIF Cutoffs to Predict "Below Expectations" Performance on Spring ORF (N = 102)



Spring Measure (ORF < 40)	At-Risk (WIF < 26)	Not At-Risk (WIF ≥ 26)	Indices of Predictive Accuracy
Below Expectations N = 11	True Positive N = 11	False Negative N = 0	Sensitivity 1.0
At or Above Expectations N = 91	False Positive N = 7	True Negative N = 84	Specificity .92
AUC = .97	PPV .61	NPV 1.0	Hit Rate .93

Note. AUC = Area Under the Curve; PPV = Positive Predictive Value; NPV = Negative Predictive Value; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; PSF = Phoneme Segmentation Fluency.

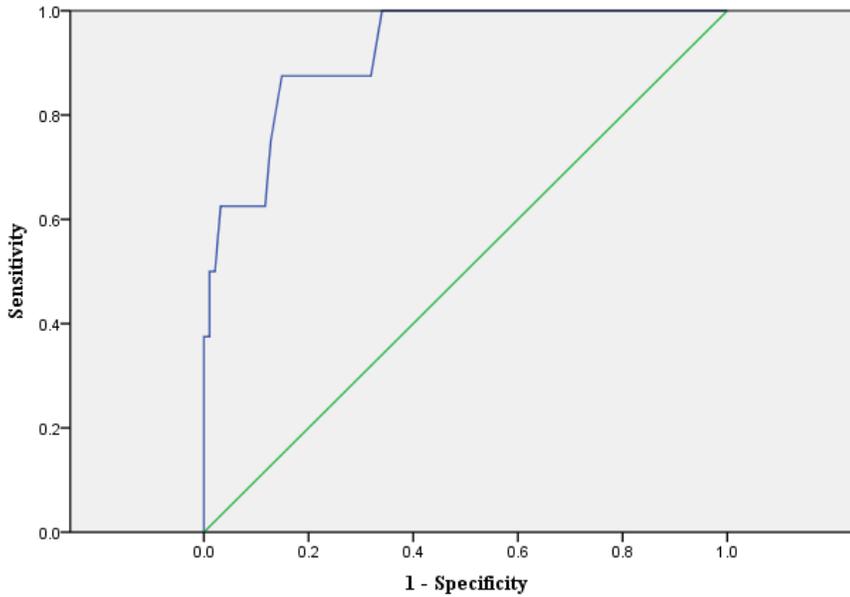
Figure 6. ROC Curve: Winter PSF DIBELS Benchmarks to Predict "Below Expectations" Performance on Spring WRMT (N = 102)



Spring Measure (WRMT < 25%)	At-Risk (PSF < 35)	Not At-Risk (PSF ≥ 35)	Indices of Predictive Accuracy
Below Expectations N = 8	True Positive N = 6	False Negative N = 2	Sensitivity .75
At or Above Expectations N = 94	False Positive N = 28	True Negative N = 66	Specificity .70
AUC = .80	PPV .17	NPV .97	Hit Rate .71

Note. AUC = Area Under the Curve; PPV = Positive Predictive Value; NPV = Negative Predictive Value; NWF = Nonsense Word Fluency; PSF = Phoneme Segmentation Fluency; WRMT = Woodcock Reading Mastery Tests.

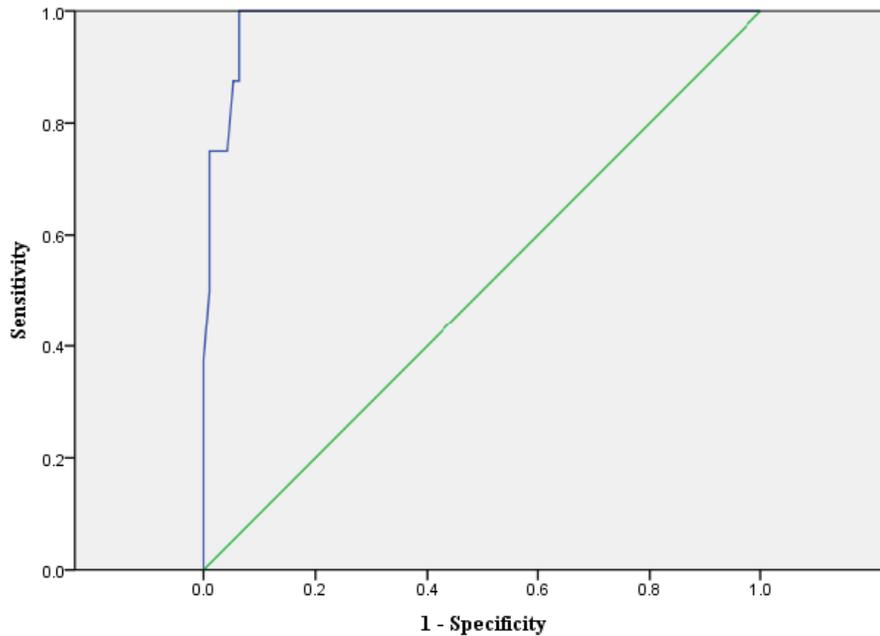
Figure 7. ROC Curve: Winter NWF DIBELS Benchmarks to Predict "Below Expectations" Performance on Spring WRMT (N = 102)



Spring Measure (WRMT < 25%)	At-Risk (NWF < 50)	Not At-Risk (NWF ≥ 50)	Indices of Predictive Accuracy
Below Expectations N = 8	True Positive N = 7	False Negative N = 1	Sensitivity .88
At or Above Expectations N = 94	False Positive N = 30	True Negative N = 64	Specificity .68
AUC = .92	PPV .19	NPV .98	Hit Rate .70

Note. AUC = Area Under the Curve; PPV = Positive Predictive Value; NPV = Negative Predictive Value; NWF = Nonsense Word Fluency; PSF = Phoneme Segmentation Fluency; WRMT = Woodcock Reading Mastery Tests.

Figure 8. ROC Curve: Winter WIF Cutoffs to Predict "Below Expectations" Performance on Spring WRMT (N = 102)



Spring Measure (WRMT < 25%)	At-Risk (WIF < 21)	Not At-Risk (WIF ≥ 21)	Indices of Predictive Accuracy
Below Expectations <i>N</i> = 8	True Positive <i>N</i> = 8	False Negative <i>N</i> = 0	Sensitivity 1.0
At or Above Expectations <i>N</i> = 94	False Positive <i>N</i> = 6	True Negative <i>N</i> = 88	Specificity .92
AUC .98	PPV .57	NPV 1.0	Hit Rate .94

Note. AUC = Area Under the Curve; PPV = Positive Predictive Value; NPV = Negative Predictive Value; NWF = Nonsense Word Fluency; PSF = Phoneme Segmentation Fluency; WRMT = Woodcock Reading Mastery Tests.