USING STATIC AND DYNAMIC MEASURES TO ESTIMATE READING DIFFICULTY FOR HISPANIC CHILDREN

by

Douglas B. Petersen

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Disability Disciplines

Approved:

[Signatures]

Dr. Ronald B. Gillam
Major Professor

Dr. Timothy A. Slocum
Committee Member

Dr. Thomas Bohman
Committee Member

Dr. Sandra L. Gillam
Committee Member

Dr. Julie Wolter
Committee Member

Dr. Byron Bunnham
Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2009
ABSTRACT

Using Static and Dynamic Measures to Estimate Reading Difficulty for Hispanic Children

by

Douglas B. Petersen, Doctor of Philosophy
Utah State University, 2009

Major Professor: Dr. Ronald B. Gillam
Departments: Special Education and Rehabilitation and Communicative Disorders and Deaf Education

This study investigated the validity of measures that were hypothesized to account for significant variance in English reading ability. During kindergarten, 63 bilingual Hispanic children completed letter identification, English and Spanish phonological awareness, rapid automatized naming, and sentence repetition static assessment tasks. They also completed a dynamic assessment nonsense-word decoding task that yielded pretest to posttest gain score, response to decoding strategy, and temporally related working memory information. One week prior to kindergarten, information was gathered regarding socioeconomic status, preschool attendance, English and Spanish language dominance, and language ability. At the end of first grade, the same children completed word identification, decoding, and reading fluency tasks designed to represent the narrow view of reading. Reliability, content relevancy, construct validity, and predictive evidence of validity were examined.
The letter identification task, the English-only and Spanish-only tasks, and a composite of the participants' best English and Spanish scores accounted for significant variance in first-grade word-level reading. However, the Spanish and BLS static measures did not account for significant, unique variance over and above English-only static measures, and the English-only static measures did not account for significant, unique variance over and above the letter identification static measure. The dynamic assessment measure pertaining to the response to reading strategy instruction accounted for equivalent variance in first-grade word-level reading when compared to a combination of letter identification and BLS static measures. The dynamic assessment measure yielded the highest classification accuracy, with sensitivity and specificity at or above 80% for all three formative criterion reading measures, including 100% sensitivity for two out of the three first-grade measures.

The dynamic assessment of reading strategy surfaced as a parsimonious, valid means of predicting first-grade word-level reading ability for Hispanic, bilingual children. When compared to multiple English, Spanish, and BLS static measures, the dynamic measure accounted for equivalent variance in the majority of first-grade reading measures and had superior classification accuracy.

(171 pages)
ACKNOWLEDGMENTS

I would like to thank my committee members, Drs. Thomas Bohman, Sandi L. Gillam, Timothy A. Slocum, and Julie Wolter, for their extraordinary feedback and support. Their areas of expertise were the perfect complement to the design and content of this dissertation. Their respective knowledge, perspective, and experience—apportioned through published works and personal consultation—were of immeasurable value. I would also like to thank Dean Carol Strong for planting the seeds of research in my heart and for being willing to serve on my committee.

I will be forever grateful to my major professor, Dr. Ronald B. Gillam, for his willingness to take me under his wing and share his exceptional knowledge of research. He has spent countless hours helping me channel my passion into something tangible and refined. Dr. Gillam is the best mentor anyone could ever hope to have. He is dedicated, knowledgeable, personable, and truly concerned for the best interest of his students. Dr. Gillam has been an example of excellence and accomplishment. His influence on my life, both personally and professionally, has been profound. I will never be able to thank him enough for his willingness to guide and support my education.

I want to express my gratitude to my family and colleagues. I have leaned on them so many times for support and encouragement, and they are always there for me. My four wonderful children have been understanding of my limited time, and continually remind me why I have the research focus that I do—they, just as all children, are precious and deserve the most out of life. My parents have always believed in me, and their love and encouragement have carried me through challenging times. My colleagues have been examples to me of efficiency, precision, and hard work. Trina Spencer has been a dear
friend throughout, and our discussions, however intense, have always brought us closer
to a better understanding of language and literacy and its wonderful intricacies. My
newest colleagues at the University of Wyoming have rendered support and friendship,
and have played a role in the completion of this dissertation. I love working with them,
and look forward to many years of collaboration.

My deepest gratitude is reserved for my wife, Olivia, who is the most precious
blessing to me. She is the facilitator of my dreams. She has somehow held our family
together while working full-time, managing a house full of busy children, and making up
for the loss of a husband who has been gone 18 hours or more most days over several
years. Her belief in me, and her understanding of my passion, has made all the difference.

Douglas B. Petersen
**CONTENTS**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Reading Difficulty for Hispanic Children</td>
<td>1</td>
</tr>
<tr>
<td>Valid Early Reading Assessment</td>
<td>2</td>
</tr>
<tr>
<td>Reading Defined</td>
<td>6</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>10</td>
</tr>
<tr>
<td>General Population Research</td>
<td>11</td>
</tr>
<tr>
<td>Language Minority Research</td>
<td>18</td>
</tr>
<tr>
<td>Research on Hispanic Children</td>
<td>22</td>
</tr>
<tr>
<td>Dynamic Assessment</td>
<td>33</td>
</tr>
<tr>
<td>Conclusions and Purpose</td>
<td>41</td>
</tr>
<tr>
<td>III. METHODS</td>
<td>44</td>
</tr>
<tr>
<td>Participants</td>
<td>44</td>
</tr>
<tr>
<td>Procedures</td>
<td>45</td>
</tr>
<tr>
<td>IV. RESULTS</td>
<td>67</td>
</tr>
<tr>
<td>Descriptive Statistics and Correlations</td>
<td>67</td>
</tr>
<tr>
<td>Reliability</td>
<td>70</td>
</tr>
<tr>
<td>Validity</td>
<td>73</td>
</tr>
<tr>
<td>Assumptions of Linear Multiple Regression</td>
<td>77</td>
</tr>
<tr>
<td>Multiple Regression Analysis</td>
<td>87</td>
</tr>
<tr>
<td>Classification Analysis</td>
<td>99</td>
</tr>
<tr>
<td>Results Summary</td>
<td>103</td>
</tr>
<tr>
<td>V. DISCUSSION</td>
<td>105</td>
</tr>
<tr>
<td>Study Overview</td>
<td>105</td>
</tr>
<tr>
<td>Correlations</td>
<td>107</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Average Correlation Coefficients ($r$) Between Pre-literacy Measures and Reading Ability as Reported by Hammill (2004)</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>Participant Characteristics</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>Percentage of English Scores and Spanish Scores That Comprised the Best Language Score</td>
<td>58</td>
</tr>
<tr>
<td>4</td>
<td>Descriptive Statistics of Kindergarten Predictor Measures and First Grade Criterion Measures</td>
<td>68</td>
</tr>
<tr>
<td>5</td>
<td>Correlations and Intercorrelations Among Kindergarten Measures and First Grade Criterion Measures</td>
<td>71</td>
</tr>
<tr>
<td>6</td>
<td>Multiple Regression Analysis Models</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>Classification Analysis</td>
<td>103</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linear regression correlation coefficients ($R$) and variance ($R^2$) estimates derived from a review of the predictive reading research that included Hispanic participants.</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>Histograms representing the distribution of nonsense word fluency, oral reading fluency, word identification, and narrow view of reading composite raw scores.</td>
<td>85</td>
</tr>
<tr>
<td>3</td>
<td>Q-Q plots representing the distribution of the residuals of nonsense word fluency, oral reading fluency, word identification, and narrow view of reading composite.</td>
<td>86</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Reading Difficulty for Hispanic Children

Hispanic children comprise a large, disproportionate percentage of the millions of school-age students in the U.S. who are classified as having reading difficulty. According to the National Assessment of Educational Progress (NAEP, 2007), Hispanic children account for approximately one third of all fourth-grade students reading below grade level. This overrepresentation of Hispanic children among struggling readers is greatly concerning. Reading is a fundamental skill required for academic achievement, and is essential for success in today's society (Snow, Burns, & Griffin, 1998).

Reading difficulty, once manifest, is pervasive and difficult to remediate. Research has indicated that nearly 75% of first grade students who are at-risk for reading problems will continue to have difficulty reading into adulthood (Lyon, 2004; Scarborough, 1998). The early identification and subsequent prevention of reading problems is the most effective method to reduce the prevalence of reading difficulty (Durlak, 1997; Gersten & Dimino, 2006; Walker, 1996), and researchers have demonstrated that reading problems can be corrected at an early age when evidence-based practices are applied (Brown & Felton, 1990; Elbro & Scarborough, 2004; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Justice & Pullen, 2003; Morris, Tyner, & Perney, 2000; National Reading Panel [NRP], 2000; Torgesen et al., 2001; Vaughn et al., 2006; Vellutino, Scanlon, & Tanzman, 1998). The early, accurate identification of children at risk for reading difficulty is one of the first steps in preventing reading problems from manifesting.
Valid Early Reading Assessment

In order to most effectively reduce the prevalence of reading problems, children at risk for reading difficulty must first be identified at an early age. Early reading assessments that require limited time to administer and that are sensitive to a diverse student population need to be implemented in U.S. schools. As with all testing instruments, an early reading assessment must demonstrate suitable psychometric properties, the sum of which offer evidence of validity.

The measures of an early reading assessment, and the results derived thereof, need to have acceptable reliability and validity. Reliability is a prerequisite of validity (Guilford, 1956; Nunnally, 1978). A measure is considered reliable when it yields consistent scores over a variety of conditions (Gall, Gall, & Borg, 2007). Reliability can be measured through self-consistent analyses or through retesting procedures. Nunnally (1967) suggested that relatively low reliability coefficients (.5) are acceptable in early stages of research, but that once measures are used to make important decisions regarding student placement and treatment, reliability should be very high. Although not designed to conclusively classify students in one category or another (e.g., good and poor readers), an early reading assessment should have as little measurement error as possible.

Validity, according to the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, and National council on Measurement in Education, 1999) is the “degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests.” (p. 9). As pointed out by Gall et al. (2007), and Messick (1995), it is the interpretation of test...
scores that hold validity. That is, the results of an early reading assessment need to be interpretable, meaningful, and actions based on test results need to be appropriate (Gall et al., 2007; Messick, 1989, 1995).

In the case of an early reading assessment designed to classify children at-risk or not at-risk for future reading difficulty, the predictor-criterion relationship must be strong, and the construct upon which the predictor and criterion measures are founded must be clearly defined and related. The intended interpretation and application of the results of an early reading assessment must be specified, and the lucidity of that specification is dependent on the careful consideration of the construct that the test is proposed to measure. Once the construct is carefully defined, the content of the predictor measures and the content of the criterion measures can be carefully specified to yield scores that reflect that construct.

Test validity can be measured by accumulating evidence to support the proposed application of test results. When examining the validity of a predictive assessment, the content relevancy of the predictor measures and the content relevancy of the criterion measures administered at a later time represent evidential, interrelated facets of the unitary concept of validity (Pedhazur & Schmelkin, 1991). The examination of test content can include theoretical and empirical analyses along with more surface-level analyses such as logical and expert judgment. Correlational analyses and multiple regression analysis can be useful in extracting influential predictor variables from a large group of variables accumulated through theory or empirical evidence. Each of these analyses should focus on the extent that the content of a test reflects the construct of interest. The standards for education and psychological testing (American Psychological
Association, 1985) stated that "...there is often no sharp distinction between test content and test construct" (p. 11). Test content that is appropriately selected engenders test scores that accurately represent a construct, leading to valid inference.

Evidence of validity can also be derived by examining the relationship between the results of the predictor measure and the results of other measures designed to assess the same construct. This examination is designed to demonstrate the extent that an instrument matches comparable, previously validated instruments in rank ordering individual performances (Bachman, 1990; McNamara, 2000). This convergent-correlational evidence of validity is highly dependent on the existence of valid, external tests.

_Predictive evidence_ and _classification results_ are additional means of assessing validity that are particularly relevant to an early reading assessment. Predictive evidence pertains to the extent that a predictor measure can accurately reflect scores on a criterion measure. Classification results indicate the extent that a predictor measure (or combination of predictor measures) can accurately classify individuals based on performance on a criterion measure. Any assessment that is designed to classify children into differing groups must have adequate sensitivity and specificity. In relation to the identification of reading difficulty, sensitivity refers to the ability of a single measure or group of measures to correctly identify children at-risk for reading difficulty (true positives). Specificity refers to the correct identification of children who are not at-risk for reading difficulty (true negatives). Previous studies related to the prediction of reading ability have revealed that a balance between good sensitivity and good specificity is difficult to obtain (Scarborough, 1998). Many studies involving young children have
over-predicted risk status, reporting false positives (over-identification) ranging from 47% to 67% (O’Connor & Jenkins, 1999; Uhry, 1993). Of potentially even greater concern, rates of false negatives (under identification) have been high, ranging from 21% to 69% (Coleman & Dover, 1993; Mantzicopoulos & Morrison, 1994; Torgesen, Burgess, Wagner, & Rashotte, 1996).

The construct that the results of an assessment are designed to represent is foundational to test validity. Therefore, it is of considerable importance to carefully specify the construct that is under examination when designing an assessment (Hager & Slocum, 2008). Construct validity refers to the extent that an assessment measures the theoretical construct it is intended to measure (Anastasi, 1976). Messick (1993) suggested that the two major threats to construct validity concern the underrepresentation of a construct and construct irrelevant variance. Construct underrepresentation is evident when an assessment does not account for all essential aspects of a construct. Construct irrelevant variance is evident when variables consistently account for (undesirable) variance in assessment results. When discussing the analysis of construct validity, Pedhazur and Schmelkin (1991) suggested that “probably the most important, certainly the first, aspect of logical analysis is to scrutinize the definition of the construct” (p. 60). A construct can have different meanings to different people, and those constructs that are poorly defined render greater ambiguity, both explicit and implicit. This ambiguity leads to imprecise measurement, resulting in poor assessment validity. When developing an instrument designed to predict a specific construct such as reading, it is important that the construct be carefully defined.
Reading Defined

The Simple View of Reading

Theory and empirical evidence can help clarify the definition of a construct, and in order for a construct to be scientifically meaningful, it must be part of an implicit or explicit theoretical framework that consists of empirically-derived formative indicators (Pedhazur & Schmelkin, 1991). Reading, broadly defined, is the result of various formative sub-skills, which lead to the development of more intermediary formative component skills. The intermediary component skills found to be causally related to reading through empirical research include word identification, word decoding, reading fluency, vocabulary, and comprehension (National Reading Panel, 2000; Puranik, Petscher, Al Otaiba, Catts, & Lonigan, 2008). These component skills are often classified into two broad groups: word recognition and comprehension. This classification is referred to as the simple view of reading (Gough & Tunmer, 1986). Generally speaking, word identification, word decoding, and reading fluency pertain to word recognition, while vocabulary and reading comprehension relate more to comprehension. Word identification and decoding are formative sub-skills necessary for accurate, fluent word recognition, and while reading fluency clearly has a facilitative role in the comprehension of text (Perfetti, 1985; Slocum, Street, & Gilberts, 1995; Stanovich, 1990), we consider it to be a means to convey a message, similar to sign language or speech, which must still be understood using skills appurtenant to a comprehension construct. The simple view of reading combines these two broad groups of skills, word recognition and comprehension, and generalizes them under one recrementitious construct.
It can be difficult to interpret the results of a measure that reflect a construct that is too broadly defined. Theory and empirical evidence must align to form a coherent, parsimonious definition of a construct (Kaplan, 1964). Many educational high stakes assessments categorize children into two distinct groups, those that are reading at or above grade level, and those that are reading below grade level. This aggregated reading classification, based on the simple view of reading, can lead to uninterpretable test scores, resulting in poor test validity. Labeling a child as having reading difficulty when neither word recognition nor reading comprehension are expressly implicated renders an imprecise diagnosis which leads to non-specific action for intervention. Thus, an assessment that defines reading globally through the simple view of reading can have poor construct validity.

The NAEP assessment renders an excellent example of the limited validity of an assessment that is conceptually based on the simple view of reading. The NAEP assessment exclusively uses measures of reading comprehension to classify reading ability. Whether the nearly 50% of Hispanic children reading below grade level according to the NAEP assessment (2007) performed poorly because of deficit word recognition skills, language comprehension, or both is not known. However, Nakamoto, Lindsey, and Manis (2007) examined word decoding and reading comprehension measures from first through sixth grade for a large sample of Spanish-speaking English language learners and found that it was the children’s comprehension, not word recognition skills that began to fall behind the normative sample starting in the third grade. The findings of this study suggest that the Hispanic children scoring poorly on the NAEP assessment did so likely because of language comprehension difficulty, and not
word recognition problems - yet it would be impossible to confirm this from the results of the NAEP assessment. The 2007 NAEP assessment lacked construct validity, which yielded ambiguous results. Again, this lack of construct validity was the product of the simple view of reading.

The Narrow View of Reading

The simple view of reading conflates word recognition and comprehension under one general construct. Reading comprehension, however, can be considered a contextual modification of language comprehension, which pertains to its own theoretical and empirical definition (cf. Zwaan & Radvansky, 1998). Consequently it is more appropriate and more parsimonious to refer to reading comprehension as a product of language comprehension (not as a product of the context in which language comprehension occurs). With comprehension appropriately denoted, the construct of reading can reflect a more concise set of skills—those that fall under word recognition. The resultant view of reading has been referred to as the narrow view of reading (Kamhi, 2007, 2009). The narrow view of reading focuses exclusively on word-level reading (word identification, word decoding, and reading fluency), and implies that a reading impairment is synonymous with difficulty in accurate and fluent word recognition.

The narrow view of reading does not diminish the importance of comprehension (Catts, 2009). On the contrary, reading comprehension becomes dissociated from the peripheral (e.g., word recognition) and is defined according its appurtenant construct: language comprehension. With reading comprehension appropriately conceptualized, an increase in the validity of the results of reading comprehension assessment may ensue.
Equivalently, by redefining the construct of reading to reflect a more narrow domain, a more accurate description of reading impairment or deficit emerges, leading to potentially increased validity of the interpretation of a reading assessment.

The narrow view of reading was used as the conceptual framework for the development and inclusion of the measures used in the current study. The predictor measures administered at kindergarten were designed to yield accurate predictive information about word recognition ability. We hypothesized that the narrow view of reading would focus the content of the test on a single, carefully defined construct. A comprehensive review of the extant literature was conducted to identify measures previously recognized as predictor variables of reading ability.

Measures that were expected to offer information about comprehension were also included in this study for two purposes (a) these measures were previously used in studies that defined reading according to the simple view of reading, making it difficult, if not impossible to determine whether such measures were predictive of word recognition, comprehension, or both constructs; and (b) by including predictive measures appertaining to comprehension, we would obtain information useful for the future analysis of the participants’ language comprehension, measured in the context of reading at a later time.
CHAPTER II

LITERATURE REVIEW

The review of literature is organized so that general population-inclusive meta-analyses, research syntheses, and large-scale longitudinal studies are considered first, followed by a review of population-exclusive research involving language minority individuals. Finally, we conducted a comprehensive review of the literature on Hispanic participants.

Several syntheses of the research and recent large-scale studies related to reading correlates have been published. These comprehensive reviews, meta-analyses, and longitudinal studies do not necessarily implicate any one particular population. However, they are of considerable interest—potentially lending valuable insight into what might be applicable to Hispanic children. Additionally, because Hispanic children are often English language learners, population-exclusive research that has concentrated on language-minority individuals may be relevant. Studies that either included a large percentage of Hispanic children or that provided sufficient information for the disaggregation of data for that specific population are likely most applicable.

The information from the following reviews, and the data obtained from individual studies were analyzed through a focused lens. The information was examined with the purpose of identifying those methods that would be appropriate for screening purposes, predictive of future problems with reading in English, applicable to kindergarten-age children, and when possible, applicable to Hispanic children. The
dependent or criterion measures of each study were classified into two distinct constructs whenever possible: reading, as defined by the narrow view, and comprehension.

**General Population Research**

**Meta-Analysis of General Population Research**

Hammill (2004) synthesized the results of three different meta-analyses on the correlations between a variety of measures and reading outcomes. Hammill's meta-analysis focused on syntheses produced by Hammill and McNutt (1981), Scarborough, (1998), and Swanson, Trinin, Necoechea, and Hammill (2003). Hammill and McNutt (1981) included both concurrent correlational studies and longitudinal studies in their meta-analysis. Scarborough (1998) only included studies that were longitudinal over the course of at least 1 year and that comprised a large number of participants. Swanson et al. (2003) only evaluated concurrent correlational studies. The three meta-analyses included publications between 1950 and 2002 and evaluated a total of 450 studies with nearly 11,000 different correlation coefficients. Hammill used superordinate ability clusters to organize the multiple independent variables used in the hundreds of studies, and calculated correlation coefficients (r) to determine the magnitude of predictive power for each ability cluster in relation to reading (defined through the simple view). Hammill used Hopkins' (2002) suggestions to describe the degree of relationship between the ability clusters and reading ability. Coefficients between .00 and .10 were considered very small, between .10 and .30 were considered small, between .30 and .50 were considered moderate, between .50 and .70 were considered large, between .70 and .90
were considered very large and coefficients between .90 and 1.0 were considered
nearly perfect. Other researchers have used this classification approach for reading (e.g.,
Caffrey, Fuchs, & Fuchs, 2008), however, it is not without controversy. Correlations
equal to or greater than .40 have been considered by some to be large (Cohen, 1977,
1988; Lipsey & Wilson, 2000). The results of Hammill’s meta-analysis are summarized
in Table 1.

Hammill (2004) found that measures most reflective of reading (e.g., the
recognition and comprehension of real and pseudo words, \( r = .71 \)), and writing
conventions (\( r = .62 \)) were the two best predictors of reading ability. The finding that
reading real or nonsense words is most predictive of reading ability is not particularly
surprising, and may have limited utility when considered in the context of predictive
assessment and prevention. That is, if the object is to predict reading difficulty before a
child has learned to read, using actual reading measures would, on the surface, appear to
be an unlikely option.

The second most predictive cluster, writing conventions, was defined as measures
related to spelling, punctuation, capitalization, book-handling skills or basic print
concepts. This heterogeneous group of measures have at least one thing in common
according to Hammill, “...they are all arbitrary conventions used in writing” (2004, p.
456). While this may be true, grouping each of the above measures together under one
superordinate ability cluster draws attention to the extremely large unit of analysis his
meta-analysis presents. For example, book-handling skills would certainly not be a
sensitive predictor of reading ability for older students, and may only be appropriate for
preschool-age children. Conversely, assessing punctuation or capitalization would likely
not be appropriate for younger children. The results of Hammill’s review suggest that measures related to writing conventions are highly predictive of reading for the general population, and measures appropriate for kindergarten-age children should be explored.

The third highly predictive ability cluster was associated with letter concepts \((r = .52)\). Letter concepts included measures associated with naming letters, distinguishing letters from non-letters, distinguishing 1 letter from another, recognizing upper- and lower-case letters, and associating speech sounds with particular letters and letter patterns.

Written language \((r = .47)\), rapid naming \((r = .44)\), phonological awareness \((r = .40)\), intelligence \((r = .35)\) and memory \((r = .30)\) were moderately correlated to reading ability. According to Hammill’s (2004) classifications, written language included measures that were linked to linguistic aspects of writing, in particular those pertaining to vocabulary composition and morphosyntax. Rapid naming measures were defined as measures that assessed the ability to quickly name familiar symbols, pictures or other stimuli such as letters, colors, numbers, or simple objects (Wolf, Bowers, & Biddle, 2000). Phonological awareness measures were those measures that assessed the ability to focus on and manipulate the phonological and phonemic features of spoken words, including tests that involved comparing and contrasting speech sounds or segmenting and blending phonemes, identifying phonemes within spoken words, or combining phonemes to make spoken words. Intelligence was defined as "...the ability to adjust or adapt to the environment, the ability to learn, or the ability to perform abstract thinking (e.g., to use symbols and concepts)" (Sattler, 2001, p. 135). Memory was defined as the ability to recall auditory or visual material after a brief lapse of time. Measures of memory
included the repetition of a series of sounds, phonemes, or spoken digits or the drawing of geometric figures from memory. Two composite abilities were reported in Hammill’s synthesis to only have a weak correlation with reading outcomes: spoken language and perceptual/motor skills.

Hammill’s (2004) conclusions pertaining to the development of preschool and kindergarten reading screening measures were that such screening assessments should include a majority of items that are related to print. A fairly comprehensive list of measures associated with print were offered.

Table 1

*Average Correlation Coefficients (r) Between Pre-literacy Measures and Reading Ability as Reported by Hammill (2004)*

<table>
<thead>
<tr>
<th>Pre-literacy Measures</th>
<th>Reading Ability (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>0.71</td>
</tr>
<tr>
<td>Writing conventions</td>
<td>0.62</td>
</tr>
<tr>
<td>Letter concepts</td>
<td>0.52</td>
</tr>
<tr>
<td>Written language</td>
<td>0.47</td>
</tr>
<tr>
<td>Rapid naming</td>
<td>0.44</td>
</tr>
<tr>
<td>Phonological awareness</td>
<td>0.40</td>
</tr>
<tr>
<td>Intelligence</td>
<td>0.35</td>
</tr>
<tr>
<td>Memory</td>
<td>0.30</td>
</tr>
<tr>
<td>Oral language</td>
<td>&lt;0.30</td>
</tr>
<tr>
<td>Perceptual/motor</td>
<td>&lt;0.30</td>
</tr>
</tbody>
</table>
The interpretation of Hammill’s (2004) meta-analysis is limited by the construct upon which it was founded; the simple view of reading. The results of the meta-analysis do not answer important questions such as ‘to what extent are items related to print indicative of the ability to comprehend written language?’ ‘To what extent are measures of spoken language indicative of the ability to decode words?’ The only concrete inference that can be derived from Hammill’s meta-analysis is that the broadly constructed predictor variables (superordinate ability clusters) were either related to word recognition ability, comprehension, or both constructs. While important information, the validity of the findings is poor because of the limited inferential application of the results.

Sample of Recent General Population Research

To further interpret the data presented in Hammill’s (2004) meta-analysis, and to extend this review of the literature to current research, two large-scale ($n > 450$), population-inclusive longitudinal studies published on or after 2001 that specified a criterion measure in greater detail were reviewed. One study used a composite of reading comprehension measures as the outcome criterion (Catts et al., 2001), whereas the other study (Vervaeke, et al., 2007) used a reading fluency measure, which we interpreted as a representation of word-level reading ability. Thus, the two studies attempted to predict two distinct constructs: reading comprehension and word recognition.

Catts et al. (2001) assessed a total of 604 kindergarten children, 328 with language impairment and 276 typically developing, for the purpose of identifying measures that were predictive of reading comprehension difficulty. All of the children were monolingual English language speakers. The kindergarten children were assessed at
the middle or end of the school year with a battery of pre-literacy and language measures. Of the 604 participants, 183 children were identified as having a reading impairment during second grade based on a score of 1 standard deviation or lower on a composite measure of reading comprehension. The children’s kindergarten test battery scores were examined and analyzed using logistic regression. The analysis identified five significant factors that contributed to future reading comprehension difficulty: letter identification, sentence imitation, mother’s education, phonological awareness: deletion, and rapid automatized naming of animals. Classification accuracy varied based on the cutoff scores chosen. Catts et al. recommended a cutoff score resulting in 91% sensitivity and 74% specificity, although other cutoff scores offered more balanced sensitivity and specificity classifications (e.g., 86% sensitivity and 86% specificity).

Even though Catts et al. (2001) carefully specified the construct of the criterion included in their study (reading comprehension), word recognition ability was not accounted for. Unless word-level reading is considered as a factor, especially with children in the lower elementary grades, it can be difficult to separate word recognition and comprehension ability. Thus, the five predictive measures identified in the Catts et al., study may have accounted for unique variance in decoding, word identification, fluency, vocabulary or comprehension. Ultimately, the criterion measure used to represent reading comprehension had limited evidence of construct validity because discriminant validity was not manifest. Thus, the findings of the Catts et al. study offered ambiguous results.

Vervaeneke, et al. (2007) assessed 499 children using predictor measures of phonological awareness and letter-sound understanding at the end of kindergarten and
then measured reading fluency ability at the end of third grade. Specifically, kindergarten measures included an individual phoneme identification task, a rhyming awareness task, a phoneme blending task, lower- and upper-case letter naming, and a letter-sound correspondence task. Factor analysis applied to the predictor measures resulted in the designation of two independent variable constructs; letter-sound understanding and phonological awareness. Third grade reading tasks included the Word Attack and Word Identification subtests of the Woodcock Reading Mastery Test—Revised (Woodcock, 1987) and a reading fluency assessment. The authors ultimately selected the reading fluency assessment to represent third grade reading ability. Although Vervaeke et al. clearly implied that reading fluency was reflective of reading as a general construct, in our perspective their reading fluency task was a measure of word recognition ability. Regression analysis indicated that the linear combination of letter-sound understanding and phonological awareness measures accounted for a significant amount of variance in third grade reading fluency ability ($r^2 = .23$). Sensitivity when using letter-sound understanding was 49% and sensitivity when using phonological awareness as a predictor was 49%. When a combination of letter-sound understanding and phonological awareness measures were used as predictor measures, the sensitivity decreased to 24%. Specificity using those same measures was 87%, 82%, and 94%, respectively. The classification results indicated that the predictive measures used by Vervaeke et al. had poor sensitivity, failing to identify a large percentage of children who were at risk for word-level reading difficulty. An acceptable combination of sensitivity and specificity is the metaphorical litmus test of the validity of predictive assessments, and the results of the Vervaeke et al. study were poor.
In summary, the review of research pertaining to the predictive evidence of validity of early reading measures for the general population indicated that the vast majority of studies were focused on English speaking, European American children. Such research revealed methods whereby children can be identified at-risk for reading problems with some accuracy (Elbro & Scarborough, 2004; Snow et al., 1998), although sensitivity and specificity have been less than adequate in many cases, and the construct of reading has been poorly defined in most studies. It is not known to what extent the findings of the meta-analyses and studies reviewed above are applicable to the Hispanic population.

Language Minority Research

Linguistic Interdependence Hypothesis

It is not unreasonable to suspect that different methods of assessment, distinct from those used to assess monolingual, European American children, may be more applicable to Hispanic children. Hispanic children are often learning English as a second language, and oral language abilities in Spanish may offer additional predictive information about future reading ability in English. While Spanish and English are clearly two distinct languages, there is some evidence that cross language transfer does occur. Cummins (1979; Cummins et al., 1984) proposed a linguistic interdependence hypothesis, which suggests that the acquisition of a second language is mediated by first language competence. This would imply that those language skills that are fundamental to reading ability in any language would transfer from the first language (e.g., Spanish) to a second language (e.g., English). There has been substantial evidence of cross language
transfer in multiple languages (August, Calderon, & Carlo, 2001; August & Shanahan, 2006; Cisero & Royer, 1995; Comeau, Cormier, Grandmaison, & Lacroix, 1999; DaFontoura & Siegel, 1995; Durgunoglu, Nagy, & Hancin-Bhatt, 1993; Escamilla, 1987; Geva, Wade-Woolley, & Shany, 1997; Gottardo, 2002; Gottardo, Yan, Siegel, & Wade-Woolley, 2001; Jiménez González & Haro Garcia, 1996; Lindsey, Manis, & Bailey, 2003; Miller et al., 2006; Quiroga, Lemos-Britton, Mostafapour, Abbot, & Berninger, 2002). Therefore it is important to determine the extent that previous research findings are applicable to the Hispanic population, and to determine which additional measures, including English and Spanish measures, might better identify Hispanic children who are at-risk for reading difficulty.

Meta Analysis of Language Minority Research

August and Shanahan (2006) published a comprehensive review of the literature pertaining to literacy development and second-language learners. Findings applicable to the early identification of reading difficulty (projected toward the Hispanic population) suggest that phonological awareness in English and in Spanish, along with working memory, rapid automatized naming, and measures related to letter knowledge in English may be predictive of reading ability. It was also found that oral language, including vocabulary knowledge in English, listening comprehension, syntactic skills and metalinguistic skills were closely related to reading comprehension. Other factors such as cognitive ability, word reading ability, socioeconomic status, home-language use, and literacy practices and educational instruction were factors that influenced reading comprehension.
August and Shanahan (2006) offered recommendations for future research applicable to the early identification of reading difficulty in English language learners. It was suggested that information regarding first-language background be collected (e.g., the amount of Spanish and English language exposure in the home, community and school) for the purpose of investigating the role of second-language (English) proficiency as it relates to reading fluency and comprehension. It was also recommended that oral language ability be explored more closely as a potential predictive variable for English language learners. It was emphasized that, though it has been fairly well established that phonological processing skills are more predictive of reading ability over oral language with monolingual English speaking students, such information has not been confirmed for English-language learners with varying English language ability. Additionally, it was recommended that other contextual variables be considered more closely, such as SES, instructional methods used in school (e.g., bilingual, ESL support, mainstream) and quality of language instruction.

**Context-Specific Assessment**

Many of the recommendations submitted by August and Shanahan (2006) relate to the collection of context-specific information, and there is a body of evidence, including a meta-analysis conducted by Kim and Suen (2003), that indicate that the predictive evidence of validity of certain measures vary according to testing conditions, such as gender, SES, length of prediction (e.g., kindergarten-first grade, kindergarten-second grade) or ethnicity (Kim & Suen, 2003; Tramontana, Hooper, & Selzer, 1988; Wilson & Reichmuth, 1985; Ysseldyke & O’Sullivan, 1987). The summative conclusion
of the meta-analysis by Kim and Suen was that “predictive validity of early assessment is indeed situation-specific” (p. 12) and that “a more cautious, specific, and limited description about predictive validity of early assessment is called for” (p. 12). In addition to carefully defining the construct of reading, collecting information that is pertinent to an individual, or a relatively homogenous group of individuals, is important when attempting to isolate factors that uniquely contribute to reading ability.

Often far removed from the situation-specific, meta-analyses frequently fail to consider heterogeneity in subjects, interventions, methods or even dependent variables (Pring, 2004). A synthesis of multiple meta-analyses in many ways may exaggerate these negative aspects. In the case of Hammill’s (2004) review, the unit of analysis (i.e., other meta-analyses) was so large that individual study quality, subject ethnicity, language ability, subject age, or specific dependent variables were not considered. Similarly, the review by August and Shanahan (2006) focused on language-minority populations, a more specific parameter than what was found in Hammills review, yet still considerably broad. Language-minority populations are a relatively heterogeneous group who are culturally and linguistically diverse from each other.

Although we have a fairly good idea of which composite (conceptual) measures correlate highest with reading ability (broadly defined) for the general population, and to a degree for language-minority populations, we do not know if there are any exceptions to this for Hispanic children. For this purpose it was considered important to identify and evaluate relatively recent (1990-2007) longitudinal research that specifically included the Hispanic population when attempting to ascertain which measures were most predictive of reading difficulty in English. Dependent variables were disaggregated whenever
possible and arranged so as to fit within intermediary reading component skills that are formative of reading and comprehension (word identification, decoding, reading fluency, vocabulary, and comprehension). These intermediary reading component skills were then divided into two broad constructs; reading (as defined by the narrow view of reading) and comprehension. Although Hammill (2004) used correlation coefficients (r) to represent the degree of association between variables, the relationship between the predictor and criterion measures used in the Hispanic-oriented research are reported in linear correlations coefficients (R) and in coefficients of determination (R^2), reflective of the regression analyses used in those studies.

**Research on Hispanic Children**

Lindsey et al. (2003) used criterion measures designed to represent both word recognition and comprehension—reflective of the simple view of reading. Lindsey et al. assessed 249 Spanish-speaking English-language learners at kindergarten using Spanish-only measures of letter and word identification, phonological awareness, expressive vocabulary, sentence repetition, rapid automatized naming, and concepts about print. They tested the children again at the end of first grade measuring decoding in English and word identification and reading comprehension in English and Spanish.

The researchers found that all Spanish kindergarten predictor variables showed some evidence of cross-linguistic transfer to the first grade English measures. Lindsey et al. (2003) used a hierarchical regression procedure that included letter identification as a control variable (a forced first step of the regression model) and phonological awareness as the second step. By entering letter identification and phonological awareness first in
the hierarchical regression model, the examiners were essentially attempting to measure the unique residual variance accounted for by the other predictor variables over and above those previously forced variables. This stepwise approach to regression analysis, where each measure was included in the analysis sequentially, cumulatively increased the stringency of the test for the latter variables. That is, those variables entered lastly in the equation were likely forced to account for an increasingly smaller amount of variance. This is surmised because of the degree of colinearity between the predictor measures. The results of combining this stringent, stepwise hierarchical regression procedure with variables that were significantly correlated is evidenced in the very small changes in $R^2$. Thus, Lindsey et al. found very small $R^2$ changes which may have been due to inappropriate selection of predictor variables or because each of the predictor variables accounted for essentially the same variance in the criterion measure.

Lindsey et al. (2003) found that letter knowledge accounted for 15% of the variance in first grade English reading comprehension and that phonological awareness, vocabulary, sentence repetition, concepts about print and rapid automatized naming combined accounted for an additional 7% of the unique variance, with each variable only accounting for 0% to 2% of the variance individually. Only phonological awareness, vocabulary, and rapid automatized naming yielded significant $R^2$ changes. The linear combination of the Spanish predictor variables accounted for a total of 22% of the variance in English reading comprehension. When using word identification as a first grade criterion measure, the results of the regression analysis indicated that 25% of the variance was accounted for by letter identification. Thereafter, phonological awareness, vocabulary, sentence repetition and concepts about print accounted for only 5% of
additional unique variance, and rapid automatized naming independently accounted for an additional 5% of unique variance in first grade word identification. Only vocabulary, concepts about print, and rapid automatized naming yielded significant $R^2$ changes. The linear combination of the Spanish predictor variables accounted for a total of 35% of the variance in English word identification. The regression analysis that included decoding as the criterion measure had similar results, with letter identification accounting for 13% of the variance and the remaining predictor variables accounting for an additional 4% of the unique variance and rapid automatized naming accounting for an additional 4% of the unique variance. Only rapid automatized naming yielded a significant $R^2$ change. The linear combination of the Spanish predictor variables accounted for a total of 20% of the variance in English decoding.

Two major points can be inferred from the Lindsey et al. (2003) regression analysis: (a) the Spanish letter identification measure administered at the beginning of kindergarten accounted for nearly the same degree and portion of variance in first grade English word recognition and first grade English reading comprehension as all other combined Spanish measures; and (b) forced, step-wise hierarchical multiple regression rendered a biased model in favor of the first variable entered in the analysis, meaning that, due to the colinearity of the predictor measures, it was impossible to determine whether any other Spanish predictor measures included in the study could have accounted for just as much variance in the first grade reading and comprehension measures as letter identification.

Lindsey et al. (2003) also examined the predictive evidence of validity of the Spanish measures used in their study and reported that sensitivity ranged from 63% to
77% and specificity ranged from 60% to 78%. These classification results suggest that the Spanish-only measures still failed to correctly identify a large number of children who were at risk for reading difficulty.

As a follow-up study to the Lindsey et al. (2003) study, Manis, Lindsey, and Bailey (2004) assessed the same cohort of students again at the end of second grade. Second grade measures that were included in the regression analyses were English word identification/decoding and English reading comprehension. The predictor variables included in the regression analyses were slightly different from the authors’ 2003 study. Manis et al. included Spanish letter identification, Spanish phonological awareness, Spanish rapid automatized naming and Spanish expressive language, leaving out Spanish expressive vocabulary, Spanish sentence repetition, and concepts about print in their analyses. The same method of hierarchical regression analysis used in the Lindsey et al. study was used in this study, with Spanish letter identification entered first into the regression analysis. Manis et al. reported that the linear combination of the kindergarten Spanish predictor variables accounted for 26% of the variance in second grade word identification, with letter identification accounting for 19% of that variance independently. The results also indicated that the linear combination of the kindergarten Spanish predictor variables accounted for 27% of the variance in second grade reading comprehension, with letter identification accounting for 16% of that variance independently. Individual measures (Spanish phonological awareness, rapid automatized naming, and expressive language) had similar (small) $R^2$ coefficients as found in the Lindsey et al. study.
Instead of using English-only or Spanish-only kindergarten measures, Páez and Rinaldi (2006) used a combination of English and Spanish measures to estimate risk for reading difficulty at kindergarten for a group of Hispanic children. Páez and Rinaldi included 244 low SES, bilingual Hispanic children in their study. The children were assessed at the end of kindergarten and again at the end of first grade. For regression analyses, the following measures were used as predictors at kindergarten: English vocabulary, English sentence repetition, English phonological awareness derived from three subtests of the Woodcock Language Proficiency Battery—Revised (WLPB-R; Woodcock, 1991), and a Spanish reading composite measure derived from three subtests of the WLPB-R—Spanish Form (Woodcock & Muñoz-Sandoval, 1995). The Criterion measure administered at the end of first grade was the English Letter-Word Identification subtest from the Woodcock Language Proficiency Battery—Revised (WLPB-R; Woodcock, 1991). Hierarchical regression analyses indicated that the linear combination of English vocabulary, English sentence repetition, and English phonological awareness accounted for 31% of the variance in first grade word identification. English vocabulary independently accounted for 21% of the variance in first grade word identification. The $R^2$ changes when sentence repetition and phonological awareness were included in the regression model were small but significant.

In a second regression model, Páez and Rinaldi (2006) included a Spanish word reading predictor measure in the regression analysis to measure its effect on first grade English word identification. Kindergarten Spanish word reading accounted for unique variance in first grade word identification over and above kindergarten English vocabulary and phonological awareness measures ($R^2$ change = .10, $p < .001$). The linear
combination of English vocabulary, English phonological awareness and Spanish word reading explained 40% of the variance in first grade English word identification.

Swanson, Sáez, Gerber, and Lefstead (2004) assessed 95 students in first grade and then again in second grade. The first grade predictor measures were a series of tasks in English and Spanish, including English and Spanish short-term and working memory tasks, a Spanish vocabulary measure derived from the Test de Vocabulario Imágenes (Dunn, Lugo, Padilla, & Dunn, 1986), and a composite first grade reading measure comprised of nonsense word and word identification measures. The second grade criterion measures included an English reading composite comprised of the Word Attack and Word Identification subtests from the Woodcock Johnson-III (Woodcock, McGrew, & Mather, 2001). A unique feature of the Swanson et al. study was that they included predictor measures pertaining to memory in their regression analysis. Results indicated that the linear combination of Spanish vocabulary, Spanish reading, English and Spanish short term memory and English and Spanish working memory measures accounted for 11% of the variance in second grade English word recognition ability. The English and Spanish memory tasks alone accounted for 8% of the variance and the Spanish memory tasks accounted for 7% of the variance in English word recognition at second grade. Only the Spanish working memory measures accounted for significant unique variance in second grade word recognition ability.

Hammer, Lawrence, and Miccio (2007) used a combination of predictor measures designed to represent 1 single construct: preschool language comprehension (English and Spanish). They then used several different measures to assess reading in English and Spanish (narrowly defined) and comprehension in English at the end of kindergarten.
Specifically, the researchers assessed 88 bilingual children at the end of their first and second year of Head Start preschool. They used a battery of receptive language assessments: the English Peabody Picture Vocabulary Test—III (PPVT—III; Dunn & Dunn, 1997), the Spanish Test de Vocabulario Imágenes Peabody (TVIP; Dunn et al., 1986), the Receptive Language subtest of the Test of Early Language Development—3 (TELD—3; Hresko, Reid, & Hammill, 1999) and the Spanish Auditory Comprehension subtest of the Preschool Language Scale—3, Spanish version (PLS—3; Zimmerman, Steiner, & Pond, 1992). The examiners assessed the same children's reading skills in the spring of their kindergarten year using the Test of Early Reading Ability—2 (TERA—2; Reid, Hresko, & Hammill, 1991) which assessed knowledge of contextual meaning, knowledge of alphabet and its functions, and print knowledge. Also, the Letter-Word Identification subtests of the Woodcock-Muñoz Language Proficiency Battery-Revised (WLPB—R; Woodcock & Muñoz-Sandoval, 1995), designed to assess letter name knowledge and word decoding ability, were administered in Spanish and English.

The researchers reported that the children's receptive English language scores at the end of Head Start were not predictive of English decoding or reading comprehension and that the Spanish receptive language measures were negatively correlated with those outcomes. It was found, however, that the rate of Spanish or English receptive language growth over 2 years during preschool was predictive of reading outcomes in kindergarten. Their findings suggest that static English receptive language measures assessed at preschool are not predictive of kindergarten English reading ability or comprehension, and that Spanish receptive language assessed at preschool is negatively correlated with kindergarten English reading and comprehension.
Hammer et al. (2007) tested a relatively small number of participants ($n = 88$) with a limited number of receptive language measures. It is possible that other static receptive language measures and/or that static expressive language measures are more predictive of reading for the Hispanic population. It is interesting to note that Páez and Rinaldi (2006) found that English expressive vocabulary was significantly predictive of first grade English word identification.

The finding that the rate of receptive language growth in English and Spanish over two years during preschool was predictive of reading could have several implications. Dependency on static, one-time measures of receptive language may not be sufficient to predict reading ability or comprehension. Frequent assessment and monitoring of receptive language ability in both Spanish and English during preschool could be more informative. When considered in the context of the development of a kindergarten early reading measure, compiling 2 years of longitudinal receptive language data prior to kindergarten would not be possible for many children, however, including a more brief dynamic assessment procedure (cf. Lidz & Peña, 1996) in the kindergarten screening battery may be useful, and add to the predictive evidence of validity of that screener.

The results of these Hispanic-focused studies are represented in Figure 1. These results offer some insight into potential differences in the predictive accuracy of language and literacy measures administered to bilingual Spanish- and English-speaking and monolingual English-speaking children. Perhaps the most striking findings are related to the role that first language measures (Spanish) play in second language reading (English). For Hispanic children, measures administered in English and in Spanish have accounted
for unique variance in English formative intermediary reading component skills. These research findings indicate that Spanish-related pre-literacy measures and Spanish language ability are predictive of reading in English for Hispanic children, and lend support to the linguistic interdependence hypothesis (Cummins, 1979; Cummins et al., 1984). It should be noted, however, that not all Spanish related tasks were found to positively correlate with intermediary reading component skills. Hammer et al. (2007) and Swanson et al. (2004) found that Spanish receptive language was negatively correlated with English word-level reading and reading comprehension, while Páez and Rinaldi (2006) found that expressive vocabulary in English and Spanish was predictive of word-level reading.

The individual and linear combination of predictor measures examined in the research with Hispanic children typically yielded moderate $R^2$ values when associated with decoding, word identification, and reading comprehension skills in Hispanic children. Large $R^2$ values were recorded between word identification as a criterion measure and the following predictor measures: Spanish letter identification (Lindsay et al., 2003; Manis et al., 2004), a combination of Spanish letter I.D., phonological awareness, expressive vocabulary, sentence repetition, concepts about print and rapid automatized naming (Lindsey et al., 2003), a combination of Spanish letter I.D., phonological awareness, rapid automatized naming and Spanish expressive language (Manis et al., 2004) and a combination of English expressive vocabulary, sentence repetition, and phonological awareness. A large $R^2$ value was noted between reading comprehension as a criterion measure and a kindergarten Spanish composite measure consisting of Spanish letter I.D., phonological awareness, rapid automatized naming, and Spanish expressive
Figure 1. Linear regression correlation coefficients (R) and variance ($R^2$) estimates derived from a review of the predictive reading research that included Hispanic participants.
language (Manis et al., 2004).

The results of the Hispanic-focused studies vary to some extent from the conclusions drawn by Hammill (2004). Measures that were found to correlate moderately with reading in the general population, such as written language measures and intelligence, were either not assessed or did not correlate significantly with reading. Written language, was only assessed with first grade Hispanic children using a random letter/number writing assessment (Swanson et al., 2004). More foundational indicators of written language, such as letter identification, may be more appropriate for kindergarten-age children. An examination of the Hispanic-specific research revealed that intelligence consistently did not account for a significant percentage of the variance in reading ability when combined with phonological awareness, oral language, and other predictor variables. This finding differs from Hammill’s report, where intelligence was found to moderately correlate with reading for the general population.

According to Hammill (2004), writing conventions (such as measures related to concepts of print, spelling, capitalization or punctuation) were highly predictive of reading ability for the general population. Yet the one Hispanic study that did include writing convention measures (concepts about print) as a predictor measure embedded that variable in a regression model that included four other predictor measures previously entered into the equation (Lindsey et al., 2003). Concepts about print measured at kindergarten accounted for only 2% of the variance in first grade word identification after letter identification, phonological awareness, vocabulary, and sentence repetition were entered in the analysis first.
This review revealed that additional research is needed to identify the sub-skill measures that are most predictive of reading for the Hispanic population. Sensitivity and specificity has not been high for any combination of English and/or Spanish predictor measures. Reading fluency was never used as a criterion measure in any of the studies, and given that reading fluency is an integrated measure of word identification and decoding, it would seem appropriate to include that measure as a criterion variable in any study interested in predicting reading, narrowly defined.

August and Shanahan (2006) suggested that context-specific measures be carefully examined, such as oral language (including vocabulary knowledge in English, listening comprehension, syntactic skills, and metalinguistic skills), socioeconomic status, home-language use and literacy practices, and educational instruction. These factors may be what is needed to increase the predictive evidence of validity of an early reading measure for Hispanic children. Other measures, mostly unexplored, such as dynamic assessment, may be predictive of reading ability in Hispanic children.

**Dynamic Assessment**

In accordance with suggestions by Hammer et al. (2007), and given its logical application to English language learners, dynamic assessment may be an additional measure that could improve the accuracy of an early kindergarten reading measure for Hispanic children. To date no cross-grade longitudinal study has investigated the predictive accuracy of dynamic assessment for reading ability.

Single-time, static assessments such as those used in the research reviewed above have a relatively high potential for measurement error (Bracken, 1998; McCauley &
Swisher, 1984a, 1984b; Plante & Vance, 1994; Spaulding, Plante, & Farinella, 2006).
Such assessments are also generally not well equipped to control for cultural and
linguistic bias (Demsky, Mittenberg, Quintar, Katell, & Golden, 1998; Rodekohr &
Haynes, 2001; Scheffner-Hammer, Pennock-Roman, Rzasa, & Tomblin, 2002; Valencia
& Rankin, 1985; Valencia & Suzuki, 2001). Dynamic assessment, on the other hand, has
been shown to be less culturally and linguistically biased and effective in providing
clinically useful information (Gillam, Peña, & Miller, 1999; Gutierrez-Clellen & Peña,

Dynamic assessment includes 3 distinct phases; the administration of a pretest, a
teaching phase, and the administration of a posttest. This method of assessment differs
markedly from the typical standardized assessment procedures (Lidz, 1991, 1996;
Sternberg & Grigorenko, 2002). Children are explicitly presented with material during
the teaching phase with the purpose of not only improving posttest scores, but to also
document emerging skills and strategies that the child may be using (Haywood & Tzuriel,
2002; Peña et al., 2006). Examiner judgments of a child's responsiveness to the
instruction, and of the examiner's effort to facilitate learning during dynamic assessment
teaching procedures is referred to as modifiability scoring (Peña, 2000; Peña et al., 2001).
Modifiability scores have been shown to yield excellent sensitivity and specificity. For
example, Peña et al., reported that the modifiability score used in their dynamic
assessment scoring procedures was more accurate in classifying language impairment
over story grammar and narrative episodic complexity measures. The researchers also
reported that when the modifiability score was combined with the dynamic assessment
posttest scores, classification accuracy increased to 100%.
Vellutino, Scanlon, and Tanzman (1998) made a strong case against the exclusive use of static assessment measures for assessing reading difficulty in culturally and linguistically diverse individuals. They suggested that most reading difficulties are caused by inadequate literacy experiences and inadequate instruction (see Clay, 1987 and Stanovich, 1986 for similar analyses). Vellutino et al., remarked that perhaps the best way to determine whether a child has a reading problem because of experiential or instructional deficits (or because of limited English ability) as opposed to a reading disorder caused by with-in child factors (e.g., cognitive deficits) is to assess the child using a response to intervention approach. A similar conclusion can be inferred from the reauthorization of the Individuals with Disabilities Education Act, (2004), which strongly encourages the use of a response to intervention method for diagnosing learning disabilities. Dynamic assessment can be viewed as an abbreviated response to intervention approach, and may be an excellent method to help predict reading difficulty (Fuchs et al., 2007).

A recent review of the literature related to dynamic assessment and its utility in predicting academic ability was conducted by Caffrey et al. (2008). The researchers concluded that dynamic assessment contributed significant, unique variance to the prediction of academic achievement beyond static assessment methods. It was specifically noted that dynamic assessment was more accurate when noncontingent feedback was used (e.g., an examiner responded to a child in a standardized manner, regardless of individual child errors), when used with students with disabilities rather than students at-risk or typically achieving students, and when independent measures
related to the dynamic assessment or criterion-referenced tests were used in the post-
testing phase of the dynamic assessment.

There appears to be sufficient evidence to warrant further investigation into the
predictive nature of dynamic assessment for Hispanic children, yet specific information is
severely lacking. For example, dynamic assessment is an assessment method—a
vehicle—whereby information is gathered and presented. There is little information
regarding what content to include in a dynamic assessment, and how that content
contributes to reliability and predictive evidence of validity. Fuchs et al. (2007)
conducted a short-term, 11-week longitudinal study to assess the predictive evidence of
validity of a dynamic assessment consisting of nonsense words that rhymed. The teaching
phase of the assessment included modeling of nonsense words and instruction regarding
onset and rime identification and blending (an analogous decoding strategy). By focusing
on an analogous reading strategy and including nonsense words as content, Fuchs et al.
found that their dynamic assessment accounted for a significant, unique amount of the
variance in reading ability. Additionally, the research pertaining to reading correlates
have offered insight into which measures are most predictive of reading, and measures
related to reading real and nonsense words are consistently highly correlated with reading
ability (Hammill, 2004). When evaluating measures that are most predictive of future
reading difficulty for young children who have not yet received any formal reading
instruction, (i.e., measures that lead to the prevention of reading difficulty), it would be
easy to dismiss actual reading measures as irrelevant. However, in the context of a
dynamic assessment, a skill that has not yet been acquired, such as decoding, can be
taught, and a child’s response to that instruction can be measured.
Assessments that include nonsense words have been shown to be less culturally and linguistically biased (Campbell, Dollaghan, Needleman, & Janosky, 1997; Ellis Weismer et al., 2000; Peña et al., 2006), and may be most appropriate to use with Hispanic children. Thus, the design of a dynamic assessment in which the content relates to nonsense word decoding and where an analogous reading strategy is targeted might best align with current research findings.

**Dynamic Assessment: Decoding Strategies**

According to McGuinness (1997), there are four major decoding strategies children use to read unfamiliar words: part-word decoding, whole-word decoding, phonological analysis, and analogic decoding. Part-word and whole-word decoding strategies encourage guessing based on context. Children who use a part-word decoding strategy identify familiar letters or combinations of letters and assemble them to create a word (e.g., the word *shipped* is incorrectly produced as *sheep*). Children using a whole-word decoding strategy use the initial and/or final letter of a word to shape their prediction of a word (e.g., the word *push* is incorrectly produced as *play*). A phonological analysis decoding strategy involves sounding out words in individual letter-sound units (e.g., the word *dog* is read by individually producing each graphophonemic unit *c – a – t*). Children who use an analogic decoding strategy can identify units comprised of 2 or more graphemes and transfer that knowledge to other words with the same orthographic patterns or neighbors (e.g., the child recognizes *ad* and correctly reads the words *mad* and *had*). Part-word decoding and whole-word decoding have been shown to be less successful word-recognition strategies. Children who are better readers tend to use
phonological analysis and analogic decoding strategies (Adams, 1990; Badian, 2001; Catts & Kamhi, 1999; Ehri, 1992; Gillam & Carlile, 1997; Gough, Juel, & Roper/Schneider, 1983; Laing, 2002; Stanovich & West, 1989; Wolter & Apel, 2005).

A dynamic assessment that uses nonsense words, if designed appropriately, could offer insight into children’s ability to learn to recognize and use orthographic patterns consistent with an analogic decoding strategy. If nonsense words used in a dynamic assessment were designed so that they could easily be divided by individual phonemes and by recognizable graphophonemic units that were phonologically related (e.g., that rhyme) and orthotactically consistent (e.g., tad, zad, nad, kad), then phonological and orthographic patterns could be emphasized during the intervention phase of the dynamic assessment, and a child’s use of those patterns could be measured (Apel, Wolter, & Masterson, 2006; Badian, 2001; Dollaghan, 1985; Ehri, 1992; Laing, 2002; Oetting, Rice, & Swank, 1995; Stanovich, 1992; Storkel, 2004; Storkel & Rogers, 2000; Wolter & Apel, 2005). Information regarding children’s use of strategies for word-level reading can be analyzed, and the connection between those strategies and future reading ability can be explored.

**Dynamic Assessment: Working Memory**

As outlined above, a dynamic assessment procedure using nonsense words could offer insight into children’s general ability to respond to reading instruction. Also, a dynamic assessment using nonsense words, by the nature of its design and content, could offer insight into how or why a positive response to instruction occurred for some children, and not others. Information about verbal working memory could be derived
from a dynamic assessment if designed accordingly, and those measures could account for response variance.

Working memory refers to the ability to hold and manipulate a discrete amount of information in mind for the purpose of short-term retrieval (Baddeley, 2001). According to Baddeley and Hitch (1974) the working memory system relies on a central executive system which manipulates information and two 'slave' systems referred to as the phonological loop and visuospatial sketchpad. Also, an episodic buffer system was added to the model more recently in response to recent evidence implicating the involvement of long-term memory in the working memory process (Baddeley, 2000). The episodic buffer communicates with the central executive and the long-term memory systems, and integrates phonological and visuospatial information.

Unsworth (2007) and Unsworth and Engle (2006) proposed that working memory is an interaction between two modules: a primary storage system that holds information short-term and a secondary memory system that is more long-term. The primary memory system is conceptually similar to Baddeley's episodic buffer (Baddeley, 2000). According to Unsworth and Engle, the primary storage system has a limited capacity that is capable of maintaining four chunks of information at a time. Evidence of this can be found in several studies related to working memory capacity (see Cowan, 2005 for a review). Being able to integrate phonological and visuospatial information in order to recall chunks of information would appear to directly relate to tasks such as reading. Research has investigated the extent to which memory processes are related to the development of phonological awareness (Baddeley, 1986; Gillam & van Kleeck, 1996;
Wagner & Torgesen, 1987) and reading ability (Bowey, Cain, & Ryan, 1992; Gathercole & Pickering, 2000; Shankweiler & Crain, 1986; Swanson, 1994).

Hamill (2004) and Swanson et al. (2004) reported that memory accounts for a unique amount of the variance in reading ability, and August and Shanahan (2006) suggested that memory be explored as a potential predictor of reading for culturally and linguistically diverse children. Teaching children to recode nonsense words in a dynamic assessment procedure would require that they attend to individual phonemes and graphemes, store and process that graphophonemic information using working memory, and then, during the posttest of the dynamic assessment, recall that information. Thus, a dynamic assessment that includes nonsense word decoding may offer insight into children’s working memory capacity. Specifically, working memory could be measured by accounting for variability in memory performance due to temporal distinctiveness. It is common for items at the beginning and end of a list to be recalled with better precision over items in the middle of a list. These effects, referred to as primacy and recency, are measurable and indicative of working memory ability (Gathercole, 2001). A dynamic assessment measure using nonsense words could lead to the measurement of differences in working memory temporal distinctiveness.

Even though phonological memory has evidence of correlating with early reading ability in children, several researchers have noted that other phonological processing measures such as phonemic awareness and rapid naming have yield higher correlations. Further, when phonological memory is assessed in conjunction with phonological awareness, it has often failed to account for unique variance over and above phonological awareness (Pennington, Van Orden, Kirson, & Haith, 1991; Torgesen et al., 1994;
Wagner, Torgeson, and Rashotte, 1994; 1997). These findings have not been replicated with Hispanic children.

It is not known whether data about word-reading strategies or working memory will contribute to our understanding of the variable response to intervention that is expected to occur in a dynamic assessment, or if those factors are predictive of future reading ability. Nor is it clear whether a general response to instruction measure will be predictive of future reading difficulty for Hispanic children in a longitudinal study, however there is evidence to suggest that it is a reasonable hypothesis.

**Conclusions and Purpose**

In summary, previous research, while limited, has shown that measures in English and Spanish such as letter identification, sentence repetition, expressive vocabulary, rapid automatized naming, writing conventions, and phonological awareness are moderately to strongly predictive of different component reading skills for Hispanic children. Measures such as short-term memory and working memory had a weaker correlation with reading skills. Additional measures and information that have yet to be explored, or that need further investigation about their relationship to reading development for Hispanic children include dynamic assessment, socioeconomic status, school history, history of English and Spanish language exposure and use (from birth to present day), and English and Spanish language ability (i.e., typically developing or impaired).

It is unclear whether some of the independent variables identified from previous research relate to word-level reading. Many of the authors of the research reviewed defined reading through the simple view of reading, and did not specify whether
decoding or comprehension were the target dependent variables. Nor did the authors in many cases attempt to determine whether a measure was predictive of both decoding and comprehension. Because of the ambiguous state of the dependent variable in so many cases, we felt that it was important to include each measure that had been previously identified as a predictor of reading ability for the Hispanic population in a single study that exclusively focused on word-level reading in concordance with the narrow view of reading. We also thought it prudent to heed the advice of August and Shanahan (2006) and Kim and Suen (2003) and include context-specific variables as potential predictor measures. Those measures that had been previously identified as predictive of reading (broadly defined), but that are not found to be predictive of word-level reading for the subjects in this study will need to be investigated in future studies using comprehension as a criterion measure.

The purpose of this study was to determine the extent to which context-specific measures (socio-economic status, preschool attendance, English and Spanish language dominance and language ability), static measures (letter identification, English and Spanish phonological awareness, English and Spanish rapid automatized naming, and English and Spanish sentence repetition), and a measure of dynamic assessment of nonsense-word decoding (designed to yield pretest to posttest gain score data, response to decoding strategy information, and temporally-related working memory information) administered to Hispanic children at kindergarten were predictive of reading ability (narrowly defined) at first grade. The primary research questions were:

1. What are the intercorrelations between the kindergarten context-specific, static and dynamic measures of reading?
2. What are the correlations between kindergarten predictor measures of reading and first grade reading criterion measures?

3. What is the reliability of the kindergarten predictor variables?

4. What is the validity of the results derived from the individual or combined kindergarten predictor measures?

4a. To what extent do individual or a linear combination of context specific, static, and dynamic kindergarten measures predict first grade reading?

4b. How well do the individual and combined kindergarten measures classify children at risk for first grade reading difficulty?
CHAPTER III

METHODS

Participants

A subgroup of the children currently involved in the study entitled *Diagnostic Markers of Language Impairment in Bilinguals* conducted by Elizabeth Peña, (PI), Ronald Gillam (Co-PI), and Lisa Bedore (Co-PI) participated in this study. The participants included 63 Hispanic kindergarten children who were followed longitudinally to the end of first grade. Children were recruited from two schools in Ogden, Utah, by consulting with the kindergarten teachers and the students’ parents regarding Hispanic ethnicity.

All Hispanic children were invited to participate in the study. Each participant attended a general education classroom during his or her kindergarten and first grade school years. Based on equivalently weighted yearly and daily expressive and receptive language information gathered from parent or guardian interviews (described in detail below), 93% of the participants could be considered bilingual (where a second language was spoken or heard at least 20% of the time), with 60% of the children classified sequentially bilingual (where yearly data indicated at least 1 year of a dominant first language followed by the introduction of a second language), 33% classified simultaneously bilingual, 8% classified Spanish dominant, and 2% classified English dominant. When exclusively using yearly expressive language data, 50% of the participants could be considered bilingual with 33% classified simultaneously bilingual, 17% classified as sequentially bilingual, 47% classified as Spanish dominant and 3%
Table 2

*Participant Characteristics*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (months)</td>
<td>63</td>
<td>65.3</td>
<td>4.2</td>
</tr>
<tr>
<td>Mother’s education</td>
<td>60</td>
<td>8.7</td>
<td>3.1</td>
</tr>
<tr>
<td>Father’s education</td>
<td>54</td>
<td>8.1</td>
<td>3.2</td>
</tr>
<tr>
<td>SES (Weighted Hollingshead Score)</td>
<td>63</td>
<td>14.3</td>
<td>7.2</td>
</tr>
<tr>
<td>Years in preschool</td>
<td>61</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Years expressive Spanish</td>
<td>62</td>
<td>3.3</td>
<td>2.8</td>
</tr>
<tr>
<td>Years expressive English</td>
<td>62</td>
<td>0.2</td>
<td>1.1</td>
</tr>
<tr>
<td>Years expressive English/Spanish</td>
<td>62</td>
<td>2.4</td>
<td>2.7</td>
</tr>
<tr>
<td>Hours/week expressive Spanish</td>
<td>62</td>
<td>5.2</td>
<td>5.3</td>
</tr>
<tr>
<td>Hours/week expressive English</td>
<td>62</td>
<td>2.9</td>
<td>4.1</td>
</tr>
<tr>
<td>Hours/week expressive English/Spanish</td>
<td>62</td>
<td>5.8</td>
<td>4.9</td>
</tr>
</tbody>
</table>

*Note.* Data collected one week prior to kindergarten.

classified as English dominant. Details about the amount of English/Spanish exposure for
each participant along with other demographic information are presented in Table 2.

**Procedures**

Phase 1 of this study took place in the late fall/early winter of the 2007 school
year. Sixty-three Hispanic kindergarten children were administered four assessment
measures in English that targeted phonological awareness, rapid automatized naming,
sentence repetition, and letter name knowledge. A nonsense word decoding dynamic assessment was also administered. Additionally, three assessment measures in Spanish that targeted phonological awareness, rapid automatized naming and sentence imitation were administered. Also, reading-related measures administered by the participants’ school district, socioeconomic information, and context-specific information such as English/Spanish use and exposure in the home, and language ability were collected for each subject. These variables were collected and administered at kindergarten for the purpose of exploring their utility as predictor measures for future reading ability. The first author and a clinical supervisor (certified speech-language pathologists), and several graduate and undergraduate students were trained to administer the kindergarten assessments. The majority of the examiners were near-fluent bilingual English/Spanish speakers, and those examiners who were English language dominant only administered English measures.

Phase 2 of this study took place a little over 1 year after the battery of kindergarten measures were administered, near the end of the participants’ first grade school year. This phase included the collection of school district administered measures pertaining to nonsense word fluency and reading fluency. The same team of examiners administered a standardized, norm-referenced assessment of word identification. Those three first grade measures (nonsense word fluency, oral reading fluency, and word identification) were chosen to serve as formative criterion measures representative of the construct of the narrow view of reading in English.
Measurement: Phase 1 Kindergarten

Hearing screening. Because hearing impairment has been documented to be associated with reading problems (Conrad, 1979; Waters & Doehring, 1990), the children's hearing was tested by school district personnel. None of the participants had a hearing impairment at the outset of the study.

Dynamic indicators of basic literacy skills. The Ogden City School District administers the English Dynamic Indicators of Basic Literacy Skills (DIBELS: Kaminski & Good, 1996) assessment to all students in every elementary school grade 3 times per year. The DIBELS assessment is administered in English, and is used to gage progress at the student, school, and district level. Different DIBELS measures are administered depending on whether it is the beginning, middle or end of the school year and depending on the grade level of the student. At the beginning and middle of the participants' kindergarten school year, the school district administered the Initial Sound Fluency subtest (ISF), the Letter Naming Fluency subtest (LNF), the Phoneme Segmentation Fluency subtest (PSF) and the Nonsense Word Fluency subtest (NWF). The Initial Sound Fluency measure is a phonological awareness task that requires a child to recognize and produce the initial sound of as many words as possible that are presented orally in 1 minute. The Letter Naming Fluency measure requires that children name as many upper- and lower-case letters arranged in random order as fast as possible in 1 minute. The Phoneme Segmentation Fluency measure is designed to assess phonological awareness. This task requires students to segment 3- and 4-phoneme words into individual phonemes as fast as possible in 1 minute. The Nonsense Word Fluency subtest requires children to
recode a series of vowel-consonant and consonant-vowel-consonant nonsense words as quickly as possible in 1 minute.

**Context-specific information.** A parent/guardian interview (Peña, Gutierrez-Clellan, Iglesias, Goldstein, & Bedore, n.d.) was conducted in person in either English or Spanish according to the parents’ request. The questionnaire collected information on (a) qualification for free school lunch, (b) parents’ occupation, (c) parents’ level of education, (d) history of language exposure at home and at school, (e) history of preschool attendance and (f) Spanish dialect spoken in the home (e.g., Puerto Rican, Mexican).

A portion of the information collected was used to calculate weighted Hollingshead scores, an index of socioeconomic status (SES; Hollingshead, 1975). An education score ranging from 1 to 7 (1 = less than 7th grade education) and an occupation score ranging from 1 through 9 (1 = farm labor/menial service jobs) were assigned for each of the participants’ parents/guardians. The education and occupation scores were then weighted to obtain a single score for each parent/guardian, ranging from 8 to 66. Hollingshead scores below 30 were considered to be in the lower SES range.

Information about school lunch was collected to lend further insight into the participants’ socioeconomic status. Eligibility for free school lunch is determined according to whether or not a student lives in a household that has a total income that falls below 130% of the federal poverty level, receives TANF (Temporary Assistance for Needy Families), or receives Food Stamps. Children meeting other special conditions including being homeless or who have parents that are migrant workers also qualify for free school lunch programs. Each of the participants qualified for free school lunch.
Home Language Profile/Familial Routine Survey. This survey was developed as part of the norming process for the Bilingual English-Spanish Assessment (Peña et al., n.d.). It was designed to collect information about what activities the child participated in each hour of the day during the week and during the weekend, with whom the child interacted, and in which language the child was spoken to and responded. The parent interview was conducted just prior to the child’s enrolment in kindergarten.

Parents were asked to recount a typical day during the week and during the weekend listing all of the major activities their child participated in each hour (e.g., dinner time, watching T.V.) and specifying with whom their child interacted and in which language their child was exposed to receptively, and which language their child used expressively. In addition to the daily language use and exposure accounting, the parents were asked to summarize their child’s language use at home and at preschool from birth to the present-day. This yearly expressive language information was collected to better understand the extent that the child used English, Spanish or both languages across their lifespan.

Our description of the children’s bilingual skills included equivalently weighted yearly and daily language data and expressive-only yearly language data. Assigning equal weight to the yearly and daily language data potentially inflated the extent that the participants were bilingual. Yearly expressive language data indicated that nearly 50% of the participants had primarily used Spanish up to a very recent date in their lives. Therefore, we ultimately decided that the history of language use across the lifespan was the more valid indicator of language dominance, due to its representation of cumulative language experience. Thus, we exclusively used the yearly expressive language data for future analyses of language dominance.
Using yearly language use data from the language questionnaire, the participants were classified into three language groups following similar procedures used by Hammer, Lawrence, and Miccio (2007), which have been supported by research conducted by Hammer, Miccio, and Rodriguez (2004), and Kohnert, Bates, and Hernández (1999). The three language groups were as follows: English Language Dominant (i.e., 80% of expressive language history was English), Spanish Language Dominant (i.e., 80% expressive language history was Spanish), and Bilingual (i.e., any combination of meaningful Spanish and English use [21% to 79%] from birth to the current date). This analysis indicated that, as previously reported, 50% of the participants were bilingual, 47% were Spanish-language dominant, and 3% were English-language dominant. This degree of bilingualism is likely an underestimation of the participants’ English language ability. Recall that these language data were collected 1 week prior to the participants’ enrollment in kindergarten, however, the static and dynamic measures were administered 4 or 5 months following the language survey, and the participants’ language of instruction was primarily English during those months of kindergarten instruction.

The home and school language survey along with the language profile were designed to offer insight into the external factors that contribute to first and second language acquisition and to assist in identifying a language impairment or deficit. Particularly with bilingual children, a language impairment cannot be determined without first understanding the child’s language history. If a child has a life-long history of hearing and speaking Spanish with limited English language exposure, then the results of an English language assessment could be considered invalid. In this case, the language
assessment should also be given in the first language. Yet even assessing a bilingual child in his or her first language and in the second language can lead to an erroneous diagnosis of language impairment. This is because if children are not simultaneously bilingual (i.e., learning Spanish and English at the same time from birth), but instead are sequentially bilingual (i.e., exposed to Spanish first and then English as a second language at a later time) both the first and second language can appear impaired due to attrition of the first language and limited exposure to the second language (Schiff-Myers, 1992). In this scenario, it may be that the child being assessed does not have a language impairment, but is instead in a transitional phase, losing one language while learning another. These children functionally have a language deficit, and their current language skills are below what would be expected of a typically developing monolingual child.

Although the language profile information used in conjunction with a language assessment will not allow us to differentiate between children who are sequentially bilingual experiencing language attrition and those children who have a true language impairment, the language profile should result in uncertainty when interpreting the results of the language assessment. This uncertainty can have two key benefits (a) children who do not have a language impairment will not be prematurely labeled, and (b) it should promote future language assessment and the use of alternative language assessments such as dynamic assessment (Bain & Olswang, 1995; Peña et al., 2001; Peña, Quinn, & Iglesias, 1992).

Conversely, the language profile information when used in conjunction with a language assessment can lend confidence in the diagnosis of those children who are predominately English or Spanish speaking and who have a language impairment. If the
language profile indicates that a child has primarily only been exposed to Spanish since birth, yet that child performs poorly on a valid Spanish language assessment, then the probability of there being a language impairment is high. The next section describes the English and Spanish language assessment used with the participants in this study and the resulting language classifications derived from the interaction between that assessment and the language profile information.

**BESA Screener.** The Bilingual English-Spanish Assessment: Screener (BESA; Peña, Gutierrez-Clellan, Iglesias, Goldstein, & Bedore, n.d.) was administered in both English and Spanish, and was used to assess the language abilities of the children. The BESA included 2 subtests in each language; a semantics screener and a morphosyntax screener. Children received 1 point for each item correct on each subtest. Norming procedures are underway for the BESA, and data has been collected for nearly 1500 bilingual children thus far. Raw scores from each BESA subtest were compared to the mean raw score derived from the norming sample. Scores at or below the 30th percentile were considered below average. This approach to assessing the language skills of the participants utilized a dichotomous scoring procedure, where children who scored above the 30th percentile on both BESA subtests were classified as average and those who scored below the 30th percentile were classified as below average. These dichotomous scores were all that was required to estimate language ability when used in conjunction with the language dominance information (see below). However, for other analyses where a more thorough account of the variance in language skill was desired (e.g., when calculating the correlation between English BESA language scores and English sentence repetition scores), we calculated composite English and Spanish language scores.
reflective of raw score distributions. To calculate English language composite scores and Spanish language composite scores, the BESA semantics and morphosyntax scores for each language were combined. The raw scores from both subtests were transformed to z-scores and then added together, yielding an English language composite score and a Spanish language composite score. These scores were in units of standard deviation, based on the sample mean.

**Language ability.** Information from the language profile, which yielded language dominance classifications, and information from the BESA subtests, which yielded classifications of average or below average language ability in English and Spanish, was cross-referenced to depict a more accurate account of language ability. Those children who were Spanish language dominant according to the language profile questionnaire, and who scored below the 30th percentile on the Spanish Semantics and Morphosyntax subtests were classified as having language impairment. Likewise, children who were classified as English language dominant and who scored below the 30th percentile on the English semantics and morphosyntax subtests were classified as having language impairment. None of the children who were classified as bilingual scored below the 30th percentile on the English and Spanish BESA semantics and morphosyntax subtests. Thus, the results of the BESA in combination with the language profile information resulted in two different language ability groups: those having typical language ability (TL) and those having language impairment (LI). These classifications were derived from less than comprehensive assessment measures, and were only a gross estimate of language ability.

The remaining measures were included in the kindergarten screening battery based off of the comprehensive review of literature that implicated such measures as
being predictive of reading ability for Hispanic children. Those measures that were in Spanish and English were administered in random order, with English or Spanish measures administered first about 50% of the time. If Spanish measures were to be administered first, the examiner spoke with the participant in Spanish prior to the testing session and throughout the Spanish assessment process. Likewise, if English measures were to be administered first, the examiner spoke English with the participant prior to, and throughout the administration of those tasks.

**Letter identification.** A letter identification task (LID) was designed and administered in English, however Spanish responses were accepted. The letter identification task included a total of 35 upper and lower case letters printed in various typefaces displayed in random order. For this task, the children were asked to name each letter following from left to right on the form. The task was not timed, and errors were recorded. Examiners awarded 1 point for each correctly named letter.

**Phonological awareness (English and Spanish).** English and Spanish phonological awareness deletion subtests were administered to each participant. The assessments were modifications of tasks developed by Catts et al. (2001) and Rosner and Simon (1971). In these tasks, examiners first gave the participants two practice items which contained pictures to help them understand the task. For example, the children were shown a picture of a fireman and then shown a picture of a fire and a picture of a man. During that practice item the examiner asked the child to say fireman while pointing to the picture of the fireman. Then the examiner asked the child to say fireman (as the examiner pointed to the individual fire picture and man picture) but don’t say fire. The examiner then covered the picture of the fireman and the picture of fire with his or her
hand, leaving only the picture of the man exposed. These procedures were designed to help each child understand that they were to first repeat the word after the examiner, and then only say part of that word as directed. There were no pictures used in the actual assessment. During the assessment, the examiner asked the child to say the test word (e.g., *Say doghouse*) and then say only a part of that word (e.g., *Now say doghouse, but don't say dog*). A portion of each word was underlined in order to indicate to the examiner which part of the word was to be omitted. The words were not visible to the child. The task progressed from the syllabic level to the individual phonemic level to better accommodate a wide range of abilities expected of kindergarten-age children (Catts et al., 2001). There were 10 items on the English and on the Spanish phonological awareness tasks. The examiner stopped the assessment once the child made 3 consecutive errors. Examiners awarded 1 point per correct item, and a total raw score was tabulated for the English and Spanish measures.

The Spanish version was not a direct translation of the English measure; the Spanish phonological awareness words were chosen based on two criteria: (a) the words had to be relatively familiar to kindergarten-age Spanish speaking children, and (b) some of the words had to be segmentable syllabically. And, as with the English phonological awareness task, the Spanish task required the children to omit increasingly more minute parts of the words, moving from the syllabic to phonemic level (e.g., *sacapuntas, secar, feo*).

**Phonological awareness best language score.** Phonological awareness best language scores (PA-BLS) represented the highest raw score obtained on either the English or Spanish phonological awareness subtest. This score was possible to obtain
because of the equivalent number of test items across the two different phonological awareness subtests. The phonological BLS was purposed to reflect a child's true phonological awareness ability with language removed as a confounding factor. The percentage of phonological awareness English scores and Spanish scores that contributed to the phonological awareness BLS measure are shown in Table 3.

**Sentence repetition (English and Spanish).** Sentence repetition measures were designed in English and Spanish. Examiners asked the participants to imitate sentences after them, and the examiner was not allowed to repeat the sentence. Sentences were designed to increase in length and in syntactic complexity. There were 10 English sentences and 10 Spanish sentences. The English sentences progressed from having simple syntactic and semantic content (e.g., *The dog is running*) to more complex content (e.g., *Before the trip, the kids cleaned the car. Yesterday the mouse escaped from the fearsome claws of the cat*).

The Spanish assessment was not a direct translation of the English assessment, however, sentences were composed with the intent of moving from simple to complex syntactic and semantic content, and at times, were similar in content to the English sentences (e.g., *El perro esta corriendo. No me dijiste que ibas a Mexico. Ayer el raton cayó en las garras del gato hambriento.*) The participant received 1 point for each correctly repeated sentence.

**Sentence repetition best language score.** Sentence repetition awareness best language scores (SR-BLS) represented the highest raw score obtained on either the English or Spanish sentence repetition subtest. This score was possible to obtain because of the equivalent number of test items across the two different subtests. The sentence
repetition BLS was designed to represent a child’s ability to repeat sentences with language removed as a confounding factor. The percentage of sentence repetition English scores and Spanish scores that contributed to the sentence repetition BLS measure are shown in Table 3.

**Rapid automatized naming (English and Spanish).** These tasks were based on the Rapid Automatized Naming of Animals assessment used in the Catts et al. (2001) study. Objects that would be familiar to children in both English and Spanish and that had the same number of syllables in English and the same number of syllables in Spanish were used. Simple drawings of a cat, a house, and a car (gato, casa, carro) were used in conjunction with the colors blue, black and red (azul, negro, rojo). The three colors were randomly assigned to the simple drawings, and the drawings were randomly placed in rows. There were six items in four rows yielding a total of 24 items in each language.

Instructions in English and Spanish were as follows (with the order of the adjective and noun reversed depending upon the language used) *Say the color and the name of each picture as fast as you can.* The examiner used a stopwatch to measure the time (in seconds) that it took for each participant to complete the entire task.

**Rapid automatized naming best language score.** Rapid automatized naming best language scores (RA-BLS) represented the highest raw score obtained on either the English or Spanish rapid automatized naming subtest. This score was possible to obtain because of the equivalent number of test items across the two different subtests. The rapid automatized naming BLS was purposed to remove language as a confounding factor. In almost every case, the better rapid automatized naming score was derived from the English subtest. The percentage of rapid automatized English scores and Spanish
scores that contributed to the rapid automatized naming BLS measure are shown in Table 3.

Table 3

*Percentage of English Scores and Spanish Scores That Comprised the Best Language Score*

<table>
<thead>
<tr>
<th>BLS Measure</th>
<th>English</th>
<th>Spanish</th>
<th>Equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA-BLS</td>
<td>30%</td>
<td>49%</td>
<td>21%</td>
</tr>
<tr>
<td>RAN-BLS</td>
<td>97%</td>
<td>1.5%</td>
<td>1.5%</td>
</tr>
<tr>
<td>SR-BLS</td>
<td>40%</td>
<td>48%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Note. PA-BLS = Phonological Awareness Best Language Score, RAN-BLS = Rapid Automatized Naming Best Language Score, SR-BLS = Sentence Repetition Best Language Score.

**Dynamic assessment of nonsense words (English).** This assessment consisted of three phases (a) the pretest phase, where the examiner asked the children to recode four nonsense words, (b) the teaching phase where the children participated in a brief lesson on how to recode the nonsense words using an onset-rime, analogous strategy in conjunction with whole word recognition, and (c) a posttest phase, where the examiner asked the children to recode the same nonsense words displayed in a different order. The pretest contained words that followed a consonant-vowel-consonant pattern, with the short vowel /æ/ and the final consonant /d/ consistent in each word, thereby creating rhyming words (i.e., *tad, zad, nad, kad*). If the children were able to recode 75% of the nonsense words or sounds at pretest, the test was discontinued.

Once the examiner administered the dynamic assessment pretest, the children were told that they were going to learn how to read, and the examiner gave a brief lesson
on how to decode the pretest words. The teaching phase of the dynamic assessment lasted less than 3 minutes. While the instructor pointed to an individual grapheme (onset) or a graphemic cluster (rime), he or she said “This letter says x.” “Say x.” Or the examiner would say “These letters together say x.” “Say x.” Then the examiner successively taught the child to blend the onset-rime, graphophonemic information to recode the word. A second round of teaching immediately followed the first round where the examiner simply asked the child to repeat what was read, using a cloze procedure. This second round of teaching was designed to reduce the level of examiner instruction and facilitate independent recoding and word recognition. To illustrate, if the examiner was teaching a child to read the word ‘tad’, the examiner said “This letter says /t/...say /t/. These letters together say /aed/...say /aed/. Put them together and you get the word /t/-/aed/.../taed/. What is this word?” The child was then expected to recode the word ‘tad’. Then, with minimal instruction, the examiner began the second round of instruction. During the second round the examiner pointed to the letter ‘t’, waited expectantly for approximately 3 seconds, then, if the child did not respond, said /t/ while pointing to the letter ‘t’. Then the examiner followed the same procedures for /aed/ while pointing to the letters ‘ad’, for the blending procedure /t/-/aed/, and for the whole word /taed/, with the child imitating the examiner each time.

The posttest phase immediately followed the teaching phase of the dynamic assessment. During the posttest phase, the examiner showed the children the same words that corresponded to the pretest and teaching phases, with the words presented in a different order. The children were then asked to read the words. During the posttest the examiner was not allowed to help the child, however, some relatively neutral prompts
were allowed such as “You can guess - what do you think this word says?” or “Remember what I told you?” The individual phonemes and words read correctly during the pretest and posttest were recorded using the international phonetic alphabet.

**Dynamic assessment sounds gain score.** A simple approach to scoring the dynamic assessment was to subtract the number of sounds read correctly at pretest from the number of sounds read correctly at posttest. This gain score approach, while desirable because of ease of calculation, was expected to offer limited information of value. By using a simple gain score measure, those students who already had experience with reading would be at a disadvantage because they would have less information to learn from pretest to posttest. For example, if a student read 6 out of 12 sounds correctly on the dynamic assessment pretest, he or she would have fewer sounds to learn than the student who was unable to read any sounds at pretest. If, after the intervention phase of the dynamic assessment, those same 2 students read all 12 sounds correctly at posttest, then the student with the lower pretest score (0) would get the higher gain score (posttest 12 - pretest 0 = 12), and the student who had a higher pretest score (6) would get a lower gain score (posttest 12 - pretest 6 = 6). It was expected that this simple gain score approach would not necessarily reflect a student’s response to instruction, and would penalize those students with greater initial reading ability. Nevertheless, for exploratory purposes, a simple gain score of sounds was calculated because of its traditional approach to scoring nonsense word assessments.

**Dynamic assessment residuum score.** In an attempt to account for prior knowledge, the dynamic assessment was also scored using the calculation of a percentage of residuum. The percent residuum is the percent of the remainder between the pretest
the pretest score and the ceiling. The dynamic assessment residuum gain score is calculated based on the relative amount of information yet to be learned after taking into account performance on the pretest. This process can be illustrated by referring back to the two students in the example above. Recall that the student who correctly read 6 out of 12 sounds during the pretest had 6 more sounds to learn before posttest mastery. After the intervention phase of the assessment, that student read all 12 sounds correctly at posttest. By using a relative scoring procedure, the student had six more sounds to learn, and that student learned all six sounds, resulting in a percent residuum score of 100% (6 out of 6). The student who read 0 sounds correctly at pretest and then read 12 sounds correctly at posttest also received a percent residuum score of 100%. Both students responded perfectly to the intervention, and both students received a percent residuum score of 100%.

The dynamic assessment residuum gain score was designed to account for individual sounds learned and individual words learned. Because the examiner provided noncontingent feedback during the teaching phase, the focus of the scoring was based on each child’s performance given standardized instruction. The intensity and length of instruction were held constant.

The dynamic assessment was also designed to assess the participants’ response to instruction pertaining to an analogous decoding strategy. A measure of working memory was also calculated, facilitated by the design of the dynamic assessment.

**Dynamic assessment decoding strategy score.** The ability to apply an analogous decoding strategy or to quickly form mental orthographic representations (Apel, 2009; Wolter & Apel, 2005), the principle focus of the teaching phase of the dynamic
assessment, was assessed by analyzing reading strategies used by the participants during the posttest of the dynamic assessment. We assigned different values to a whole word guessing strategy that implicated erroneous word recognition, a phonological analysis strategy that had a specific focus on the initial letter of each nonsense word, an analogous decoding strategy, and a whole-word recognition strategy that implicated the use of accurate mental orthographic representations. Whole word guessing was defined as the production of one or more words that contained two or more incorrect sounds (e.g., a child produced /tap/ for /kad/) or when two or more words were repeated (e.g., a child produced /kad/ each time when attempting to recode / tad/, / nad/ and / zad/). Evidence that the participant learned to use an onset, phonological analysis decoding strategy was defined by an increase in the accurate decoding of an initial letter-sound from pretest to posttest (e.g., a child at pretest produced /gat/ for the nonsense word / kad/ but at posttest the child produced / kat/ for the nonsense word / kad/). Evidence that the participant learned to use an analogous decoding strategy was defined by the use of an onset-rime production at posttest (e.g., a child produced / t/ and the rime cluster / ad/ to recode the nonsense word / tad/). Evidence that the participant used a whole word recognition strategy was defined by the rapid, accurate recoding of the nonsense word. The rubric used for calculating the strategy score is included in appendix A. The strategy scoring procedure was comprised of the following guidelines: Participants received 10 points for using a whole word guessing strategy and when no analogous strategy was applied and no improvement in initial phonological analysis strategy was evident. Participants received 20 points when whole word guessing was evident and there was no analogous decoding strategy applied (but there was evidence of improvement in initial phonological
analysis). Participants received 30 points when they used a combination of whole word guessing and an analogous decoding strategy. Participants received 40 points when whole word guessing was not evident.

In conjunction with the strategy analysis, each student was rated on a 3-point response to instruction scale based on each examiner's perception of the participants' case of modifiability, ranging from easy to difficult. Students who had a high degree of difficulty in responding to the standardized instructions of the teaching phase of the dynamic assessment were assigned a response to instruction score of 0. Those students whose modifiability was moderate received a score of 10, and those students whose modifiability was considered easy received a score of 20. These response to instruction, modifiability ratings, were based on the judgment of the examiner at the time of the testing procedure. The reading strategy score (0-40) and response to instruction score (0-20) were combined to create a dynamic assessment strategy score.

**Dynamic assessment: working memory score.** Aspects of the dynamic assessment subtest permitted the measurement of working memory. Children were asked to recode nonsense words during the posttest phase of the dynamic assessment that contained both auditory (phonological) and graphemic (visuospatial) information that had to be integrated in order to recode the nonsense words. In order to measure the children's working memory ability, the reading responses during the posttest phase of the dynamic assessment were analyzed using temporal distinctiveness measures. Measurement procedures included a recency analysis, a primacy analysis, and a delayed response analysis. The working memory analyses were only conducted when the pretest score was 0, thereby eliminating the potential confound of prior knowledge.
Children received up to 4 recency points. Three points were awarded for correctly reading the individual sounds in each nonsense word and 1 point was awarded for reading the whole word. ‘Tad’ was the first of 4 words taught during the instruction phase of the dynamic assessment, designating it the target word for the recency analysis. A range of 0 to 4 primacy points were awarded based on posttest results of the word ‘kad’, which was the last word presented during the instructional phase of the assessment. Similar to the recency scoring procedure, 4 possible points pertained to the 3 possible individual sounds (3 points) and the whole word (1 point).

A delayed memory analysis was also conducted. This analysis was possible because the first two words taught during the intervention phase were the last two words tested in the posttest (and vice-versa). Children received up to 8 points for correctly reading the first two words of the posttest, and up to 8 points for correctly reading the last two words of the posttest. The first two words in the posttest (‘nad’ and ‘kad’) were each worth up to 4 points (3 points for each sound and 1 point for the whole word). The last two words presented in the posttest (‘tad’ and ‘zad’) were valued at 4 points each (3 points for the three possible sounds and 1 point for the whole word). A composite working memory score was calculated by tallying all four of the memory subtest scores, representing the temporal distinctiveness theoretical construct.

**Measurement: Phase 2 First Grade**

Measures representing the formative, intermediary components of word recognition were used as first grade criterion measures, along with a composite of those formative measures that was purposed to represent the construct of reading, narrowly defined. First grade nonsense word fluency (decoding), word identification, and reading
fluency were considered representative measures of word-level reading ability. The nonsense word reading task measured a student’s ability to apply phonic and structural analysis skills to recode unfamiliar, nonsense words. The word identification measure reflected a student’s ability to either correctly decode or recognize a real word. We considered reading fluency to be comprised of three component skills; the accuracy of word decoding, automaticity of word recognition, and prosody of oral text reading (Torgesen & Hudson, 2006). Reading fluency was considered the summative product of accurate decoding and word recognition skills. These subtests were administered when the participants were in the first grade, approximately 1 year following the collection and administration of the kindergarten predictor measures.

**Decoding: nonsense word fluency.** The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Nonsense Word Fluency (NWF) subtest was the measure selected to represent first grade decoding ability. The DIBELS NWF is a standardized assessment designed to record the number of nonsense words a child can produce in 1 minute. The test is comprised of randomly ordered VC and CVC nonsense words (e.g., raw, sig, ov) and is scored according to the number of correct sounds produced. The test-retest reliability was reported to be .83 (Good et al., 2004). The convergent-correlational evidence of validity of the DIBELS NWF when compared to a curriculum-based measure of reading fluency administered 4 months later was .82 (Good et al., 2004).

**Reading fluency.** The Dynamic Indicators of Basic Early Literacy Skills Oral Reading Fluency subtest was used to represent reading fluency. The Oral Reading Fluency subtest required that children read a passage aloud for 1 minute. Words substituted, omitted, and any hesitations that last longer than 3 seconds were scored as
errors. The total number of words read correctly in 1 minute comprised the oral reading fluency score. The test-retest reliability ranged from .92 to .97 and the alternate form reliability of different passages taken from the same level ranged from .89 to .94 (Tindal, Marston, & Deno, 1983). Convergent-correlational evidence of validity measured across 8 different studies ranged from .52 to .91 (Good & Jefferson, 1998).

**Word identification.** The Woodcock Reading Mastery Test-Revised (WRMT-R: Woodcock, 1987) Word Identification subtest (Word ID) was used to represent word identification. The *Letter-Word Identification* subtest involved the recognition of sight words that were presented in isolation. The split half reliability for first grade Word Identification was .98. The internal consistency reliability (Cronbach’s alpha: Cronbach, 1951) was reported to be .96. Convergent-correlational evidence of validity with the Woodcock-Johnson Reading Test was .69.

**Narrow View of Reading.** The three first grade measures described above (DIBELS: NWF, DIBELS: ORF, and the WRMT-R: Word ID) were combined to represent the construct of reading, narrowly defined (NVR). The raw scores of each measure were transformed to standardized z-scores then converted to scaled scores with a mean of 10 and a standard deviation of 3. The resultant scaled scores from each measure were then added, yielding a composite reading score for each participant.
CHAPTER IV

RESULTS

Descriptive Statistics and Correlations

Table 4 summarizes descriptive statistics for the kindergarten predictor measures and first grade criterion measures. Intercorrelations among the kindergarten measures and first grade measures are shown in Table 5. Each variable included in the intercorrelation matrix was initially chosen based on theory, hypothesis, and previous research findings. Specific research questions and the degree of correlation between each kindergarten predictor variable and the first grade composite measure reflective of the narrow view of reading served as the final determiner regarding the inclusion or exclusion of those variables for future regression and classification analyses.

The Hollingshead score (SES), number of years of expressive English language use, number of years expressive Spanish language use, number of years expressive bilingual English/Spanish language use, number of years of preschool attendance, the English sentence repetition measure, the dynamic assessment sounds gain score and the dynamic assessment composite working memory score measured at kindergarten were not significantly correlated with first grade NWF, first grade ORF, first grade Word ID, or the composite narrow view of reading (NVR) score. Additionally, the kindergarten BESA English and Spanish measures (semantics and morphosyntax) did not significantly correlate with the first grade NVR score (not shown in Table 5). All remaining kindergarten predictor measures were significantly correlated with the three first grade
Table 4

*Descriptive Statistics of Kindergarten Predictor Measures and First Grade*

**Criterion Measures**

<table>
<thead>
<tr>
<th>Measure</th>
<th>$n$</th>
<th>Mean</th>
<th>$SD$</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kindergarten</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letter identification</td>
<td>63</td>
<td>28.5</td>
<td>7.9</td>
<td>32.0</td>
</tr>
<tr>
<td>Phonological awareness English</td>
<td>63</td>
<td>2.4</td>
<td>2.7</td>
<td>1.0</td>
</tr>
<tr>
<td>Phonological awareness Spanish</td>
<td>63</td>
<td>2.8</td>
<td>2.6</td>
<td>2.0</td>
</tr>
<tr>
<td>Phonological awareness BLS</td>
<td>63</td>
<td>3.3</td>
<td>2.7</td>
<td>3.0</td>
</tr>
<tr>
<td>Rapid Automatized Naming English</td>
<td>63</td>
<td>194.8</td>
<td>50.3</td>
<td>213</td>
</tr>
<tr>
<td>Rapid Automatized Naming Spanish</td>
<td>63</td>
<td>51.2</td>
<td>66.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Rapid Automatized Naming BLS</td>
<td>63</td>
<td>193.4</td>
<td>50.1</td>
<td>212.0</td>
</tr>
<tr>
<td>Sentence repetition English</td>
<td>63</td>
<td>2.4</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Sentence repetition Spanish</td>
<td>63</td>
<td>2.6</td>
<td>2.2</td>
<td>2.0</td>
</tr>
<tr>
<td>Sentence repetition BLS</td>
<td>63</td>
<td>3.6</td>
<td>2.1</td>
<td>3.0</td>
</tr>
<tr>
<td>Dynamic Assessment Sounds Gain</td>
<td>63</td>
<td>6.7</td>
<td>4.6</td>
<td>7.0</td>
</tr>
<tr>
<td>Dynamic Assessment Residuum</td>
<td>63</td>
<td>58%</td>
<td>37%</td>
<td>56%</td>
</tr>
<tr>
<td>Dynamic Assessment Decoding</td>
<td>63</td>
<td>42.4</td>
<td>19.0</td>
<td>50</td>
</tr>
<tr>
<td>SES: Hollingshead Score</td>
<td>63</td>
<td>14.3</td>
<td>7.2</td>
<td>14.5</td>
</tr>
<tr>
<td>Number years expressive English</td>
<td>62</td>
<td>0.2</td>
<td>1.1</td>
<td>0</td>
</tr>
<tr>
<td>Number years expressive Spanish</td>
<td>62</td>
<td>3.3</td>
<td>2.8</td>
<td>3.5</td>
</tr>
<tr>
<td>Number years expressive Bilingual</td>
<td>62</td>
<td>2.4</td>
<td>2.7</td>
<td>0.5</td>
</tr>
<tr>
<td>Number of years of preschool</td>
<td>61</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

*(table continues)*
<table>
<thead>
<tr>
<th>Measure</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsense word fluency</td>
<td>48</td>
<td>53.9</td>
<td>17.9</td>
<td>51.0</td>
</tr>
<tr>
<td>Oral reading fluency</td>
<td>48</td>
<td>19.7</td>
<td>12.9</td>
<td>16.0</td>
</tr>
<tr>
<td>Word identification</td>
<td>38</td>
<td>39.2</td>
<td>12.1</td>
<td>40.5</td>
</tr>
<tr>
<td>Narrow view of reading composite</td>
<td>37</td>
<td>30.9</td>
<td>8.7</td>
<td>29.3</td>
</tr>
</tbody>
</table>

*Note. BLS = Best Language Score*

Formative criterion measures and the composite narrow view of reading score with the exception of Spanish sentence repetition, which was significantly correlated only with ORF and the NVR score.

Intercorrelations between kindergarten variables revealed several significant, high correlations (> .49) and several non-significant, low correlations. Those correlations (high and low) of greatest interest were noted. For example, it was interesting that the letter identification and rapid automatized naming tasks in English were highly correlated (r = .54) considering that the letter identification task was un-timed. However, both tasks required the student to point to and accurately identify items (letters or objects) while moving from left to right and from the top of the page to the bottom of the page, tracking from one line to the next, as is generally required when reading connected text. These similarities in the tasks may account for the high correlation. It was not surprising to find that letter identification and the dynamic assessment strategy scores were highly correlated given the high grapheme content overlap. Of particular interest was the very high correlation between the phonological awareness English and phonological
awareness Spanish measures. This very large correlation lends further evidence that phonological awareness is robust cross-linguistically (for a review see Durgunoglu et al., 1993). It was also interesting that the phonological awareness English task was highly correlated with the English sentence repetition measure and the working memory composite measure ($r = .43$, not shown in the correlation matrix). This strengthens the argument that sentence repetition has a strong phonological memory component (Rathvon, 2004; Torgesen, 1996). Of all the predictor variables, the dynamic assessment strategy score had the highest number of large correlations between the other predictor variables. That is, the dynamic assessment strategy score was highly correlated with at least 1 subtest related to letter identification, phonological awareness, rapid automatized naming and sentence repetition. This finding is worthy of note because it speaks to the potentially robust nature of the dynamic assessment in accounting for several reading-related skills in English and Spanish.

Several of the extremely low, nonsignificant correlations were also of great interest. For example, none of the SES, language use, or school history variables were moderately or highly correlated with any of the other predictor variables, with the exception of a moderate, negative correlation between the number of years of expressive Spanish and the English sentence repetition task, and a moderate, positive correlation between the number of years of expressive bilingual language use and the English sentence repetition task.

Reliability

A valid measure must first be reliable, and once reliability has been established, then
### Table 5

**Correlations and Intercorrelations Among Kindergarten Measures and First Grade Criterion Measures**

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Letter Identification</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Phonological Awareness English</td>
<td>.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Phonological Awareness Spanish</td>
<td>.27</td>
<td>.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Phonological Awareness BLS</td>
<td>.29</td>
<td>.92</td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Rapid Automated Naming English</td>
<td>.54</td>
<td>.35</td>
<td>.59</td>
<td>.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Rapid Automated Naming Spanish</td>
<td>.18</td>
<td>.33</td>
<td>.29</td>
<td>.33</td>
<td>.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Rapid Automated Naming BLS</td>
<td>.52</td>
<td>.26</td>
<td>.50</td>
<td>.35</td>
<td>.35</td>
<td>.23</td>
<td>.65</td>
<td>.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Sentence Repetition English</td>
<td>.16</td>
<td>.53</td>
<td>.50</td>
<td>.27</td>
<td>-.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Sentence Repetition Spanish</td>
<td>.38</td>
<td>.28</td>
<td>.36</td>
<td>.36</td>
<td>.16</td>
<td>.45</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Sentence Repetition BLS</td>
<td>.39</td>
<td>.57</td>
<td>.47</td>
<td>.57</td>
<td>.35</td>
<td>.32</td>
<td>.65</td>
<td>.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Dynamic Assessment Sounds</td>
<td>.26</td>
<td>.36</td>
<td>.37</td>
<td>.36</td>
<td>.44</td>
<td>.12</td>
<td>.40</td>
<td>.36</td>
<td>.01</td>
<td>.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Dynamic Assessment Residual</td>
<td>.43</td>
<td>.46</td>
<td>.41</td>
<td>.48</td>
<td>.44</td>
<td>.20</td>
<td>.41</td>
<td>.40</td>
<td>.17</td>
<td>.40</td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Dynamic Assessment Decoding</td>
<td>.52</td>
<td>.46</td>
<td>.50</td>
<td>.55</td>
<td>.52</td>
<td>.31</td>
<td>.40</td>
<td>.52</td>
<td>.57</td>
<td>.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. SES: Hollingshead Score</td>
<td>.16</td>
<td>.06</td>
<td>.04</td>
<td>.05</td>
<td>.10</td>
<td>-.13</td>
<td>.05</td>
<td>.26</td>
<td>-.22</td>
<td>.12</td>
<td>.24</td>
<td>.18</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Number Years Expressive English</td>
<td>.10</td>
<td>-.09</td>
<td>-.20</td>
<td>-.10</td>
<td>.16</td>
<td>-.14</td>
<td>.16</td>
<td>-.22</td>
<td>.03</td>
<td>.02</td>
<td>-.02</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Number Years Expressive Spanish</td>
<td>.09</td>
<td>-.11</td>
<td>-.03</td>
<td>-.03</td>
<td>-.08</td>
<td>.05</td>
<td>-.04</td>
<td>.36</td>
<td>.26</td>
<td>-.02</td>
<td>-.13</td>
<td>-.09</td>
<td>.06</td>
<td>-.12</td>
<td>-.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Number Years Expressive</td>
<td>-.14</td>
<td>.18</td>
<td>.11</td>
<td>.10</td>
<td>.02</td>
<td>-.16</td>
<td>-.03</td>
<td>.34</td>
<td>-.15</td>
<td>.05</td>
<td>.08</td>
<td>-.03</td>
<td>.06</td>
<td>-.16</td>
<td>-.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Number of Years of Preschool</td>
<td>-.26</td>
<td>-.15</td>
<td>-.02</td>
<td>-.08</td>
<td>-.08</td>
<td>-.03</td>
<td>-.05</td>
<td>-.04</td>
<td>-.07</td>
<td>-.10</td>
<td>-.11</td>
<td>-.18</td>
<td>-.14</td>
<td>.01</td>
<td>-.10</td>
<td>.08</td>
<td>-.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Nonword Reading Fluency (1st Grade)</td>
<td>.49</td>
<td>.31</td>
<td>.35</td>
<td>.32</td>
<td>.42</td>
<td>.40</td>
<td>.41</td>
<td>.10</td>
<td>.16</td>
<td>.38</td>
<td>.24</td>
<td>.33</td>
<td>.46</td>
<td>.12</td>
<td>.08</td>
<td>.01</td>
<td>.01</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Oral Reading Fluency (1st Grade)</td>
<td>.37</td>
<td>.36</td>
<td>.33</td>
<td>.34</td>
<td>.31</td>
<td>.40</td>
<td>.31</td>
<td>.01</td>
<td>.36</td>
<td>.22</td>
<td>.34</td>
<td>.46</td>
<td>.15</td>
<td>.05</td>
<td>.08</td>
<td>.07</td>
<td>.00</td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. Word Identification (1st Grade)</td>
<td>.63</td>
<td>.37</td>
<td>.46</td>
<td>.46</td>
<td>.53</td>
<td>.43</td>
<td>.54</td>
<td>.19</td>
<td>.26</td>
<td>.40</td>
<td>.28</td>
<td>.29</td>
<td>.51</td>
<td>.10</td>
<td>.11</td>
<td>.10</td>
<td>.06</td>
<td>.05</td>
<td>.73</td>
<td>.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. Narrow View of Reading</td>
<td>.53</td>
<td>.41</td>
<td>.42</td>
<td>.41</td>
<td>.50</td>
<td>.44</td>
<td>.51</td>
<td>.13</td>
<td>.34</td>
<td>.47</td>
<td>.24</td>
<td>.37</td>
<td>.53</td>
<td>.16</td>
<td>.14</td>
<td>.05</td>
<td>.01</td>
<td>.03</td>
<td>.94</td>
<td>.92</td>
<td>.91</td>
<td></td>
</tr>
</tbody>
</table>

* p < .05, ** p < .01
validity can be examined (Schiavetti & Metz, 1997). In order to estimate the reliability of the individual kindergarten measures (Question 3) coefficient alpha internal consistency estimates of reliability (Cronbach, 1951), were conducted for those measures that contained more than 1 item. A reliability coefficient of 0.00 indicated the absence of reliability whereas a coefficient of 1.00 indicated perfect reliability. It is generally accepted that an internal consistency alpha of .70 is satisfactory, and that an alpha of .80 or higher is desirable. As mentioned previously, Nunnally (1967) stated that when conducting early-stage research, relatively low reliability coefficients (.5) are acceptable. Because this study is highly exploratory in nature, it was determined that reliability above .6 would be acceptable, but that future iterations of the predictor measures would be expected to have higher reliability.

The letter identification measure, the English and Spanish phonological awareness measures, and the English and Spanish sentence repetition measures were analyzed using Cronbach’s alpha. Cronbach’s alpha reliability coefficients were as follows: letter identification: .95, English phonological awareness: .89, Spanish phonological awareness: .86, English sentence repetition: .79, and Spanish sentence repetition: .67. The dynamic assessment decoding strategy measure was scored in a holistic manner in that evidence of reading strategies were noted across the four posttest measures. Individual nonsense words at posttest were not scored. In order to calculate some aspect of reliability of the dynamic assessment, we decided to examine the number of sounds read correctly per each pretest nonsense word and each posttest nonsense word. We hypothesized that this scoring procedure and the subsequent analysis of reliability would render information concerning the equivalency of the responses to the nonsense words as
used in pretest and posttest contexts. In an attempt to conform to the assumptions underlying internal consistency reliability procedures, the pretest and posttest nonsense words were grouped accordingly, and treated as separate subtests of the dynamic assessment measure. Cronbach’s alpha for the pretest words was .96. Cronbach’s alpha for the posttest words was .90.

The internal consistency estimates of reliability were high to moderately-high for all of the measures with the exception of the Spanish sentence repetition measure which was low-moderate. The English and Spanish rapid automatized naming measures were not assessed for reliability because of their content structure that generated only 1 score (number of seconds). Future analyses using test-retest or parallel-forms will need to occur to assess the reliability of those measures.

Validity

In addition to reliability, test validity was explored by examining content relevancy, convergent-correlational evidence, predictive evidence, and classification results. Some of the individual predictor measures included in this study were compared to other predictor measures used in other studies or with measures that appeared to pertain to the same construct. Pearson’s product moment correlation coefficients (r) were calculated to represent the degree that the measures used in this study were related to similar measures.

The results of the kindergarten letter identification measure were compared to the mid-kindergarten DIBELS Letter Naming Fluency measure. The results of the kindergarten English phonological awareness measure were compared to the mid-
kindergarten DIBELS English Phoneme Segmentation Fluency and Initial Sound Fluency subtests. The English and Spanish kindergarten sentence repetition measures were compared to the English and Spanish BESA Screener Semantics and Morphosyntax measures, and the language ability estimation.

The kindergarten letter identification measure was largely correlated with the kindergarten DIBELS Letter Naming Fluency measure, \( r = .65, p < .0001 \), the kindergarten English phonological awareness measure was largely correlated with the kindergarten DIBELS Initial Sound Fluency measure, \( r = .52, p < .0001 \), and moderately/largely correlated with the kindergarten DIBELS Phoneme Segmentation Fluency measure, \( r = .44, p < .0001 \).

Language specific measures (e.g., English sentence repetition) were not compared to language ability (i.e., as defined by the BESA screener English and Spanish assessments and the language dominance information). Only those constructs that accounted for language dominance (e.g., the sentence repetition BLS) were compared to language ability. This is because it was not expected that a participant who was a dominant Spanish speaker would perform well on the English-related measures (or vice versa). A Bonferroni approach to control for Type I errors was applied across the related pairwise correlations. In each case the alpha level .05 was divided by the number of correlations that were computed.

First, it was important to determine whether the English sentence repetition measure and the Spanish sentence repetition measure significantly correlated with parallel BESA English and Spanish composite language measures (BESA semantics and morphosyntax subtests combined). We hypothesized that a significant correlation
between the sentence repetition measures and the corresponding BESA language measures would indicate convergent-correlational evidence of validity. A Bonferroni approach to controlling for familywise error rate across the following four tests at the .05 level resulted in an alpha of .01. The results of the correlational analysis showed that the English sentence repetition measure was highly correlated with the English BESA composite language score, $r = .62, p < .0001$, and that the Spanish sentence repetition measure was moderately-highly correlated with the BESA Spanish composite language score, $r = .42, p = .001$. Expectedly, the kindergarten Spanish sentence repetition measures were negatively correlated with the BESA English composite measure ($r = -.18$), and the English sentence repetition measures were negatively correlated with the BESA Spanish composite measure ($r = -.20$). These results indicated that the English and Spanish kindergarten sentence repetition measures were highly related to respective English and Spanish language performance.

Next, it was important to determine whether the sentence repetition BLS score and the dynamic assessment composite working memory score significantly correlated with language ability. A Bonferroni approach to controlling for familywise error rate across the following 2 tests at the .05 level resulted in an alpha of .02. The results of the correlation analysis indicated that the sentence repetition BLS measure was moderately-largely correlated with language ability, $r = .42, p = .001$. This significant correlation suggested that sentence repetition in a student’s dominant language correlated with his or her language ability in that language.

Although the participants were not administered additional assessments specific to working memory, we decided to explore the degree that the kindergarten working
memory composite measure correlated with the language ability composite and the kindergarten DIBELS Phoneme Segmentation Fluency measure. These comparisons were driven by theory and empirical evidence implicating a strong relationship between working memory, language, and phonological awareness (Alloway et al., 2005; Gathercole & Baddeley, 1993; Gillam & van Kleeck, 1996; Oakhill & Kyle, 2000; Pickering, 2006). The dynamic assessment composite working memory measure was not significantly correlated with language ability \( r = .18, p = .26 \), however it was significantly correlated with the DIBELS phonological awareness measure to a moderate-large degree \( r = .47 \).

Content-relevancy, and inferential evidence of construct validity can also be derived from the regression analyses that follow, and are explored in greater detail in the discussion section of this manuscript. Multiple regression analysis can contribute to the construction of an assessment through exploratory means (the principal application of multiple regression in this study), where the influence of different traits or abilities can be identified and extracted from a large set of variables. This extraction of influential variables (content relevance) has bearing on the predictive evidence of validity of those measures on the construct of interest (in this case reading, narrowly defined), to the extent that specification error has been curtailed. Specification errors in multiple regression, which include the failure to verify several assumptions, can lead to errors of inference, which in turn lead to the misapplication of test results, resulting in poor validity.
Assumptions of Linear Multiple Regression

Regression analysis requires that several assumptions be met before the sample-derived results can be projected inferentially to a population. Violation of the assumptions for regression analysis can lead to bias. Bias can mean that the regression coefficient estimate derived from a sample does not reflect the true value of the regression coefficient corresponding to the population. Bias can also mean that the standard error of the regression coefficients are inaccurate, resulting in incorrect hypothesis tests and confidence intervals. Bias leads to improper inferential interpretation of the regression results - and when applied to the construction of a test such as an early reading measure, negatively affects test validity.

Visual and statistical analysis of the data (in most cases the residuals) can reveal problems with the regression model. For this study, some of the assumptions required for regression analysis were examined in a secondary manner via the extensive visual and statistical analysis of other assumptions. The assumption that there was a linear relationship between all of the Y variables (criterion measures) and the X variables (predictor measures) was assessed in part through the examination of other assumptions (e.g., detection of outliers and homoscedasticity). Had any of the analyses indicated that the relationship between X and Y was nonlinear, we were prepared to use a weighted least squares regression analysis, or transform the data to fit a linear regression model to simplify the relationship.

The assumption that the predictor measures had been correctly specified based on a specific theory and/or empirical research was carefully considered. This is reflected in
the comprehensive review of the literature that led to the inclusion of the predictor measures, the specific research questions and the methods of this study. Nevertheless, while dynamic assessment and other hypothesis-driven measures were included as predictor variables, the empirical evidence was lacking for the target population of the study, and directional hypotheses were considered insufficiently formed to conduct strong theory-driven analyses such as structural equation modeling.

At the outset of the study it was expected that the perfect reliability assumption would be violated to some degree. While the predictor measures were carefully constructed, we expected imperfect reliability. As was reported in a preceding section of this report, the reliability of the predictor measures was generally acceptable, and while not perfect, acceptable reliability coefficients are generally all that is expected with the types of measures used to assess complex human behavior. Other assumptions of multiple linear regression were tested extensively. The results of those analyses follow.

Outliers

It is possible for outliers to have a powerful effect on the results of regression analysis. Regression analysis results can reflect a small number of atypical cases instead of the general data trend. If a sample size is small, the potential for outliers to affect the regression analysis increases.

Leverage: predictor variables. To assess for the possible influence of outliers in the predictor variables, a visual examination using index plots (multivariate and univariate) and multivariate regression diagnostic indices of leverage (the centered leverage value and Mahalanobis distance; Green & Salkind, 2008) were examined.
Multivariate calculations were conducted for each set of predictor variables included in 5 different models (explained below) across the three criterion measures. The centered leverage value indicates how far the observed values for each case are from the mean values on the set of IVs. The further a case’s score is from the mean of a set of IVs, the greater the potential to influence the results of the regression equation. The Mahalanobis distance is a measure of the distance between a specific case’s values on the predictor variables and the centroid of the IVs. Mahalanobis distance provides essentially the same information as the centered leverage value, but uses a different method for determining cutoff points. Using both centered leverage values and Mahalanobis distance, in conjunction with visual analysis to investigate leverage on the predictor variables was expected to yield more reliable information.

Extremity based on leverage and mahalanobis distance was determined using the following cutoffs: cutoff points for centered leverage values were calculated as follows: 3(k+1)/N which is 3 times (#IVs +1) divided by the number of participants. Mahalanobis distance cutoffs points were calculated by computing a chi-square criterion where the number of degrees of freedom was equal to the number of predictor measures, at alpha = p < .01 (Tabachnik & Fidell, 1996). For NWF and ORF with an n of 48, centered leverage cutoff points were .25, with three predictor variables, .31, with four predictor variables, and .38 for five predictor variables. For WID with an n of 38, centered leverage cutoff points were .32 with three predictor variables, .39 with four predictor variables, and .47 for five predictor variables. For Mahalanobis distance, the cutoff value for three predictor variables was 11.34, for four predictor variables it was 13.28, and for five predictor variables it was 15.09.
Several cases appeared to have extreme values of leverage through visual examination of the index plots. Centered leverage values and Mahalanobis distance analyses identified the same cases as being outliers. Cases 4, 12, 21, 23, 37, 42, 44, 45, 58, and 62 had high levels of leverage. Because leverage relates only to the predictor variables, and does not affect the $R^2$ value of the regression results, action was reserved until discrepancy was analyzed.

**Discrepancy: criterion measures.** Externally studentized residuals (Cohen, Cohen, West, & Aiken, 2003; Cook & Weisberg, 1982) offer information concerning the degree that the predicted criterion (Ŷ) is discrepant from the parametric criterion (Y). The externally studentized residual analysis for cases whose Y values were highly discrepant ($t = +/- 2$) from the predicted values were identified. These calculations were conducted for each set of predictor variables included in the 5 different models (explained below) across the three criterion measures. Externally studentized residuals offer information about the regression analysis when outlying cases are deleted. Several cases were found to have an outlying status based on the externally studentized residual analysis. Cases 35, 37, and 62 were identified as being outliers with NWF and ORF and cases 23, 42, and 58 were identified as being outliers with Word ID.

**Influence: leverage and discrepancy.** Cook’s D (Cook, 1977) takes into account information from leverage and discrepancy, and offers a global measure of the influence of outlying cases. A cutoff of 1.0 for Cook’s D was used. None of the cases met the criterion for outlier status and significant influence.

**Conceptual analysis of outliers.** Cases 4, 12, 21, 44 and 45 had high levels of leverage only. Case 35 had a high level of discrepancy only. Cases 23, 37, 42, 58, and 62
had high levels of leverage and discrepancy. None of the cases had high levels of global influence according to a Cook's D analysis. A more conceptual, analytical approach to identifying outliers was undertaken in conjunction with regression statistics and visual analyses. While 5 cases exclusively had high leverage values, an investigation of the extremity of those values using a z score analysis revealed that only 1 of the outliers had a score greater than $z = 3.3$. A z score of 3.3 or greater is a common rule of thumb for determining outlier status (Warner, 2007). The outlying score that was above 3.3 standard deviations was the English rapid automatized naming measure. Preliminary regression analyses involving the English-only predictor variables and the three different formative criterion measures (NWF, ORF, Word ID) were conducted to determine whether deleting the outlying score on the English RAN variable would alter the significance of the regression equation. While there was an $R^2$ change in the analysis, it was small ($R^2$ change = .011), and the significance level was maintained at $< .01$. Based on the preliminary analysis, it was decided to leave the independent variable distributions intact and not remediate any of the outliers found using leverage, Mahalanobis distance or a z-score analysis.

Cases that had high discrepancy (the distance between the predicted and observed values on Y) were also analyzed further. Performance on the pretest of the dynamic assessment nonsense word decoding task was used as a gross indicator of outlying performance in reading at kindergarten. A typical child just beginning kindergarten was expected to have difficulty with nonsense word decoding, thus if participants were able to read at least 75% of the pretest nonsense words, they were flagged as a potential outlier. Additionally, a visual inspection of the data and an investigation of relative standing
based on norm referencing using local and national percentile rankings were used to confirm outlying status. Comparing the participants’ scores on the criterion measures to a normative sample was deemed appropriate when identifying and estimating the extent that a score was an outlier. It was considered preferable to use local norms instead of national norms when possible due to the unique population sample in this study. Both NWF and ORF local school district norms were available. These norms were used as a guide in determining which cases were in the extreme for NWF and ORF. Word ID norms derived from a large sample of children across the U.S., published in the technical manual of the Woodcock Reading Mastery Test-Revised (Woodcock, 1987), were used.

After identifying those cases that were able to read at least 75% of the nonsense words at pretest and after examining their relative status to a normative sample on the dependent measures, univariate analyses were conducted to identify individual outliers on the dependent measures. Scatter plots and histograms were produced graphically to aid in the visual inspection of the data. This inspection revealed that there were visible outliers within the NWF, ORF (extremely high raw scores) and Word ID data (one extremely low raw score and two extremely high raw scores). Recall that the regression diagnostic discrepancy analysis identified three cases with extreme high values within the NWF data, four cases with extreme high values within the ORF data and three cases within the Word ID data (one extremely low score and two extremely high scores). Each of the cases with extremely high scores were ranked above the 85th percentile in local or national norms, and all but one of those cases correctly read at least 75% of the nonsense words at pretest (case 37 refused to read the pretest words). Case 58 was a potential outlier due to his extremely low scores.
Remediation of outliers. High outlying data were winsorized to the next highest
data point +1 and low outlying data were winsorized to the next lowest data point -1. All
data points identified as outliers, whether identified as having high discrepancy or not,
were also winsorized in the same manner. Raw scores on NWF from cases 35, 37, and 62
were winsorized from raw scores 125, 142, and 185 to a raw score of 85 (respectively).
Raw scores on ORF from cases 15, 35, 37, and 62 were winsorized from raw scores 64,
80, 83, and 112 to a raw score of 47 (respectively). Although cases 23, 42, and 58 were
identified as having high discrepancy within the Word ID data, an inspection of the data
indicated that there were no obvious outlying data in the high range of scores. For
example, cases 23 and 42 had raw scores of 50 and 53, yet 6 other cases also had scores
at or above 50. Additionally, there was no clear point to which a cutoff could be made as
opposed to the NWF and ORF data (e.g., NWF raw scores dropped from 125 to 84). The
Word ID raw scores followed a relatively sequential descending pattern (e.g., 61, 60, 57,
56 etc.), and none of the scores registered above 3.3 standard deviations from the mean.
Additionally, cases 23 and 42 were unable to read any of the nonsense words at pretest.
For these reasons, cases 23 and 42 were not winsorized. However, case 58 received a raw
score of 0 on the Word ID subtest, and the next lowest raw score was 15. The data then
ascended relatively sequentially from 15 (e.g., 15, 17, 18, 24 etc.), thus revealing an
outlying pattern for case 58, who’s score was at the 1st percentile. Additional
investigation into case 58 revealed that this child was the only participant who was
enrolled in special education services, lending credibility to his outlying status. Case 58
was winsorized from 0 to 14 on his Word ID score.
Normality

A normal distribution of scores meets one of several important assumptions required for the inferential application of regression analysis (Cohen et al., 2003). After Winsorizing the extreme outliers from NWF, ORF and Word ID, visual examination of the data via histograms and Q-Q plots revealed that the distribution of the raw scores within NWF, ORF, Word ID and the composite narrow view of reading (NVR) score appeared to follow a relatively normal distribution. Figure 2 shows the histograms of the NWF, ORF, Word ID, and NVR data, and Figure 3 shows the Q-Q plots of the NWF, ORF, Word ID, and NVR data.

Linearity

The Levene test of homogeneity of variance (Levene, 1960) indicated that the variance of the residuals was constant with the exception of the regression equation that included the dynamic assessment strategy measure and the oral reading fluency criterion measure (Levene, p = .01). To determine whether corrective action was necessary (e.g., data transformation or WLS regression), an estimate of the magnitude of the nonconstant variance was calculated. Unless the magnitude of the heteroscedasticity is large, the nonconstant variance will not have a material effect on the results of the regression analysis (Cohen et al., 2003). The magnitude of the nonconstant variance was determined by ordering the cases from lowest to highest according to their values on the predictor variable (dynamic assessment strategy) and then by dividing those cases into 3 sets (slices) with an equal number of cases in each slice.
Figure 2. Histograms representing the distribution of nonsense word fluency, oral reading fluency, word identification, and narrow view of reading composite raw scores.
Figure 3. Q-Q plots representing the distribution of the residuals of nonsense word fluency, oral reading fluency, word identification, and narrow view of reading composite.

The variance of the residuals around the regression line within each slice was calculated, revealing a mean square error of 120.9, 172.3, and 123.0 for each slice, respectively. It was determined that if the ratio of the largest to the smallest conditional variance for the slices exceeded 10, or if the conditional variance changed in a regular and systematic
manner as values on the predictor variable increased, then a remedial procedure would be considered. Because the ratio between the highest and lowest conditional variance was less than 10, and because a systematic pattern of increase was not observe from slice 1 to slice 3, the data were not corrected.

Multiple Regression Analysis

The following analyses were conducted to determine whether the English and Spanish static measures and the dynamic assessment measures collected at kindergarten significantly predicted first grade reading ability. Reading at first grade was measured using the DIBELS nonsense word fluency measure (NWF), the DIBELS oral reading fluency measure (ORF), the Woodcock Reading Mastery Tests-Revised word identification subtest (Word ID) and a composite word recognition construct representative of the narrow view of reading (NVR). The regression models were designed to identify the contribution of the kindergarten letter identification measure, kindergarten English-only measures and the kindergarten Spanish-only measures to first grade reading. Additionally, the models were designed to identify the contribution of the kindergarten English and Spanish measures when the highest score of either language was used (best language scores: BLS). Finally, it was considered important to determine whether a single or best combination of the kindergarten dynamic assessment measures predicted first grade reading when the influence of the kindergarten static measures was partialed out in the analysis.

For the first model (Model 1), English-only measures found to be significantly correlated with the NVR construct were entered simultaneously into the regression model
(English phonological awareness and English rapid automatized naming). The purpose of the Model 1 regression analysis was to determine the extent that the English-only kindergarten measures predicted first grade reading. In Model 2, the objective was to determine the extent that the Spanish-only measures predicted first grade reading, thus, Spanish-only measures were simultaneously entered into the regression model. For NWF and Word ID, these Spanish-only measures included Spanish phonological awareness and Spanish rapid automatized naming. For ORF and NVR, these Spanish-only measures included Spanish phonological awareness, Spanish rapid automatized naming and Spanish sentence repetition (based on an analysis of correlation between those predictor measures and the criterion measures).

Secondary analyses were conducted to determine whether there was a significant change in $R^2$ following the inclusion of the Spanish measures after the English measures had been entered first in the regression analysis. The results of this analysis indicated that the Spanish measures did not account for a significant amount of variance over and above the English measures for NWF ($R^2$ change = .08, $F(2, 43) = 2.53, p = .09$), ORF ($R^2$ change = .135, $F(3, 42) = 2.17, p = .06$) or Word ID ($R^2$ change = .084, $F(2, 33) = 2.33, p = .11$).

Model 3 included the English/Spanish best language score from phonological awareness, rapid automatized naming, and sentence repetition. The purpose of Model 3 was to determine the extent that the combination of the BLS scores predicted first grade reading. Model 4 included letter ID in step 1 and the letter ID and BLS measures in step 2. The purpose of Model 4 was to determine the extent that the BLS measures were predictive of the criterion measures over and above the letter identification measure.
Prior to the Model 5 analysis, a preliminary regression analysis was conducted to determine whether the dynamic assessment residuum gain score uniquely accounted for variance in the criterion variables over and above the dynamic assessment strategy score. This regression analysis would inform whether the two different dynamic assessment measures (previously identified as being significantly correlated with the criterion measures) should be included in Model 5, or whether the dynamic assessment strategy score alone should be included. The results of the preliminary regression analysis indicated that the dynamic assessment residuum score did not account for a significant amount of unique variance within NWF ($R^2$ change = .01, $F(1, 45) = .40, p = .53$), ORF ($R^2$ change = .004, $F(1, 45) = .23, p = .64$) or Word ID ($R^2$ change = .001, $F(1, 35) = .07, p = .80$).

Based on the preliminary analysis, the dynamic assessment strategy measure was entered first in step 1 of Model 5, then the dynamic assessment strategy measure along with the static measures of letter identification and English/Spanish BLS measures were entered into step 2. The purpose of Model 5 was to determine whether the dynamic assessment strategy measure accounted for a significant amount of variance in word-level reading at first grade, and to determine whether the best combination of static measures accounted for a significant amount of unique variance over and above the dynamic assessment strategy measure.

As shown in Table 6, Model 1 entered the English measures simultaneously into the regression equation (English phonological awareness and English rapid automatized naming). This model showed that the kindergarten English measures significantly predicted first grade performance on the DIBELS nonsense word fluency ($R^2 = .21, R^2_{adj}$
### Table 6

**Multiple Regression Analysis Models**

<table>
<thead>
<tr>
<th>Model &amp; kindergarten variable</th>
<th>$B$</th>
<th>Std error</th>
<th>Beta</th>
</tr>
</thead>
</table>

**First Grade Nonsense Word Fluency**

**Model 1 (English Only)**

- Phonological Awareness English: 1.42, .97, 0.20
- Rapid Automatized Naming English: 0.11, 0.05, .35*

$R = .46, R^2 = .21, R^2_{adj} = .18, F(2, 45) = 6.12, p < .01$

**Model 2 (Spanish Only)**

- Phonological Awareness Spanish: 1.73, 0.89, 0.26
- Rapid Automatized Naming Spanish: 0.10, 0.04, .35*

$R = .47, R^2 = .22, R^2_{adj} = .19, F(2, 45) = 6.48, p < .01$

**Model 3 (English/Spanish BLS)**

- Phonological Awareness BLS: 0.79, 1.04, 0.12
- Rapid Automatized Naming BLS: 0.10, 0.05, .31*
- Sentence Repetition BLS: 1.76, 1.31, 0.21

$R = .49, R^2 = .24, R^2_{adj} = .19, F(3, 44) = 4.57, p < .01$

**Model 4 (LID vs English/Spanish BLS)**

**Step 1**

- LID: .86, 0.29, .40*

$R = .40, R^2 = .16, R^2_{adj} = .14, F(1, 46) = 8.82, p < .01$

**Step 2**

- LID: .35, .36, .17
- Phonological Awareness BLS: .78, 1.04, 0.12
- Rapid Automatized Naming BLS: .767, 0.05, .24
- Sentence Repetition BLS: 1.31, 1.39, .15

$R = .50, R^2 = .25, R^2_{adj} = .19, F(4, 43) = 3.66, p < .01$

$R^2$ change = .09, $F(3, 43) = 1.79, p = .16$

(table continues)
<table>
<thead>
<tr>
<th>Model &amp; kindergarten variable</th>
<th>$B$</th>
<th>Std error</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 5 (Dynamic Assessment vs LID and English/Spanish BLS)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic Assessment Strategy</td>
<td>.44</td>
<td>.12</td>
<td>.46**</td>
</tr>
<tr>
<td>$R = .46, R^2 = .21, R^2_{adj} = .19, F(1, 46) = 8.82, p &lt; .01$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic Assessment Strategy</td>
<td>.19</td>
<td>.19</td>
<td>.20</td>
</tr>
<tr>
<td>LID</td>
<td>.27</td>
<td>.37</td>
<td>.13</td>
</tr>
<tr>
<td>Phonological Awareness BLS</td>
<td>.37</td>
<td>1.11</td>
<td>.05</td>
</tr>
<tr>
<td>Rapid Automatized Naming BLS</td>
<td>5.25</td>
<td>.06</td>
<td>.16</td>
</tr>
<tr>
<td>Sentence Repetition BLS</td>
<td>1.08</td>
<td>1.41</td>
<td>.13</td>
</tr>
<tr>
<td>$R = .52, R^2 = .27, R^2_{adj} = .19, F(5, 42) = 3.15, p &lt; .05$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ change = .06, $F(4, 42) = .89, p = .48$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**First Grade Oral Reading Fluency**

<table>
<thead>
<tr>
<th>Model 1 (English Only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonological Awareness English</td>
</tr>
<tr>
<td>Rapid Automatized Naming English</td>
</tr>
<tr>
<td>$R = .41, R^2 = .17, R^2_{adj} = .14, F(2, 45) = 4.66, p &lt; .01$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 2 (Spanish Only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonological Awareness Spanish</td>
</tr>
<tr>
<td>Rapid Automatized Naming Spanish</td>
</tr>
<tr>
<td>Sentence Repetition Spanish</td>
</tr>
<tr>
<td>$R = .51, R^2 = .26, R^2_{adj} = .20, F(3, 44) = 5.01, p &lt; .01$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 3 (English/Spanish BLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonological Awareness BLS</td>
</tr>
<tr>
<td>Rapid Automatized Naming BLS</td>
</tr>
<tr>
<td>Sentence Repetition BLS</td>
</tr>
<tr>
<td>$R = .46, R^2 = .21, R^2_{adj} = .15, F(3, 44) = 3.85, p &lt; .05$</td>
</tr>
</tbody>
</table>

(table continues)
<table>
<thead>
<tr>
<th>Model &amp; kindergarten variable</th>
<th>$B$</th>
<th>Std error</th>
<th>Beta</th>
</tr>
</thead>
</table>

**Model 4 (LID vs English/Spanish BLS)**

**Step 1**

LID  
$R = .37, R^2 = .14, R^{2 \text{adj}} = .12, F(1, 46) = 7.39, p < .01$

**Step 2**

| LID | .27 | .27 | .18 |
| Phonological Awareness BLS | .74 | .77 | .15 |
| Rapid Automated Naming BLS | 2.56 | .04 | .11 |
| Sentence Repetition BLS | 1.21 | 1.02 | .20 |

$R = .48, R^2 = .23, R^{2 \text{adj}} = .15, F(4, 43) = 3.14, p < .05$

$R^2$ change = .09, $F(3, 43) = 1.63, p = .20$

**Model 5 (Dynamic Assessment vs LID and)**

**Step 1**

Dynamic Assessment Strategy  
$R = .46, R^2 = .21, R^{2 \text{adj}} = .19, F(1, 46) = 12.19, p < .001$

**Step 2**

| Dynamic Assessment Strategy | .31 | .09 | .46** |
| LID | .19 | .27 | .13 |
| Phonological Awareness BLS | .35 | .81 | .07 |
| Rapid Automated Naming BLS | 2.80 | .04 | .01 |
| Sentence Repetition BLS | .98 | 1.03 | .16 |

$R = .51, R^2 = .26, R^{2 \text{adj}} = .17, F(5, 42) = 2.93, p < .05$

$R^2$ change = .05, $F(4, 42) = .69, p = .60$

**First Grade Word Identification (n = 38)**

**Model 1 (English Only)**

| Phonological Awareness English | 1.08 | .74 | .22 |
| Rapid Automated Naming English | 9.49 | .03 | .45** |

$R = .57, R^2 = .32, R^{2 \text{adj}} = .28, F(2, 35) = 8.26, p < .001$

(table continues)
<table>
<thead>
<tr>
<th>Model &amp; kindergarten variable</th>
<th>B</th>
<th>Std error</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 2 (Spanish Only)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological Awareness Spanish</td>
<td>1.70</td>
<td>.64</td>
<td>.38**</td>
</tr>
<tr>
<td>Rapid Automated Naming Spanish</td>
<td>6.22</td>
<td>.03</td>
<td>.34*</td>
</tr>
<tr>
<td>( R = .57, R^2 = .32, R^2_{adj} = .28, F(2, 35) = 8.23, p &lt; .001 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 3 (English/Spanish BLS)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological Awareness BLS</td>
<td>.89</td>
<td>.75</td>
<td>.19</td>
</tr>
<tr>
<td>Rapid Automated Naming BLS</td>
<td>8.23</td>
<td>.03</td>
<td>.39*</td>
</tr>
<tr>
<td>Sentence Repetition BLS</td>
<td>.90</td>
<td>.87</td>
<td>.16</td>
</tr>
<tr>
<td>( R = .60, R^2 = .36, R^2_{adj} = .31, F(3, 34) = 6.45, p &lt; .001 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 4 (LID vs English/Spanish BLS)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LID</td>
<td>.85</td>
<td>.17</td>
<td>63**</td>
</tr>
<tr>
<td>( R = .63, R^2 = .40, R^2_{adj} = .38, F(1, 36) = 23.93, p &lt; .0001 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LID</td>
<td>.59</td>
<td>.23</td>
<td>.44</td>
</tr>
<tr>
<td>Phonological Awareness BLS</td>
<td>.89</td>
<td>.69</td>
<td>.20</td>
</tr>
<tr>
<td>Rapid Automated Naming BLS</td>
<td>3.18</td>
<td>.04</td>
<td>.15</td>
</tr>
<tr>
<td>Sentence Repetition BLS</td>
<td>.30</td>
<td>.84</td>
<td>.06</td>
</tr>
<tr>
<td>( R = .69, R^2 = .47, R^2_{adj} = .41, F(4, 33) = 7.32, p &lt; .0001 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 ) change = .07, ( F(3, 43) = 1.47, p = .24 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 5 (Dynamic Assessment vs LID and)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic Assessment Strategy</td>
<td>.33</td>
<td>.09</td>
<td>.51**</td>
</tr>
<tr>
<td>( R = .51, R^2 = .26, R^2_{adj} = .24, F(1, 36) = 12.38, p &lt; .001 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic Assessment Strategy</td>
<td>2.50</td>
<td>.13</td>
<td>.04</td>
</tr>
<tr>
<td>LID</td>
<td>.59</td>
<td>.24</td>
<td>.44*</td>
</tr>
<tr>
<td>Phonological Awareness BLS</td>
<td>.86</td>
<td>.73</td>
<td>.19</td>
</tr>
<tr>
<td>(table continues)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model &amp; kindergarten variable</td>
<td>B</td>
<td>Std error</td>
<td>Beta</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----</td>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>Rapid Automated Naming BLS</td>
<td>2.82</td>
<td>.04</td>
<td>.13</td>
</tr>
<tr>
<td>Sentence Repetition BLS</td>
<td>.26</td>
<td>.89</td>
<td>.05</td>
</tr>
</tbody>
</table>

\[ R = .69, R^2 = .47, R^2_{\text{adj}} = .39, F(5, 32) = 5.69, p < .001 \]

\[ R^2 \text{ change} = .22, F(4, 32) = 3.25, p = .02 \]

**Narrow View of Reading Composite (n = 37)**

**Model 1 (English Only)**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Std error</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonological Awareness English</td>
<td>0.94</td>
<td>.54</td>
<td>.27</td>
</tr>
<tr>
<td>Rapid Automated Naming English</td>
<td>6.03</td>
<td>.02</td>
<td>.41**</td>
</tr>
</tbody>
</table>

\[ R = .56, R^2 = .31, R^2_{\text{adj}} = .27, F(2, 34) = 7.70, p < .01 \]

**Model 2 (Spanish Only)**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Std error</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonological Awareness Spanish</td>
<td>1.02</td>
<td>.47</td>
<td>.33**</td>
</tr>
<tr>
<td>Rapid Automated Naming Spanish</td>
<td>4.90</td>
<td>.02</td>
<td>.36*</td>
</tr>
<tr>
<td>Sentence Repetition Spanish</td>
<td>.46</td>
<td>.60</td>
<td>.12</td>
</tr>
</tbody>
</table>

\[ R = .56, R^2 = .31, R^2_{\text{adj}} = .25, F(3, 33) = 4.91, p < .01 \]

**Model 3 (English/Spanish BLS)**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Std error</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonological Awareness BLS</td>
<td>.40</td>
<td>.55</td>
<td>.12</td>
</tr>
<tr>
<td>Rapid Automated Naming BLS</td>
<td>5.11</td>
<td>.02</td>
<td>.34*</td>
</tr>
<tr>
<td>Sentence Repetition BLS</td>
<td>1.08</td>
<td>.64</td>
<td>.28</td>
</tr>
</tbody>
</table>

\[ R = .59, R^2 = .35, R^2_{\text{adj}} = .29, F(3, 33) = 6.01, p < .01 \]

**Model 4 (LID vs English/Spanish BLS)**

**Step 1**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Std error</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>LID</td>
<td>.51</td>
<td>.14</td>
<td>.53**</td>
</tr>
</tbody>
</table>

\[ R = .53, R^2 = .28, R^2_{\text{adj}} = .26, F(1, 35) = 13.73, p < .001 \]

**Step 2**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Std error</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>LID</td>
<td>.25</td>
<td>.18</td>
<td>.26</td>
</tr>
<tr>
<td>Phonological Awareness BLS</td>
<td>.42</td>
<td>.54</td>
<td>.13</td>
</tr>
<tr>
<td>Rapid Automated Naming BLS</td>
<td>2.99</td>
<td>.03</td>
<td>.20</td>
</tr>
<tr>
<td>Sentence Repetition BLS</td>
<td>.81</td>
<td>.66</td>
<td>.21</td>
</tr>
</tbody>
</table>

(table continues)
Model 5 (Dynamic Assessment vs LiD and

Step 1

<table>
<thead>
<tr>
<th>Dynamic Assessment Strategy</th>
<th>0.25</th>
<th>0.07</th>
<th>0.53**</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R = 0.53, R^2 = 0.28, R^2_{\text{adj}} = 0.26, F(1, 35) = 13.90, p &lt; .001$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Step 2

<table>
<thead>
<tr>
<th>Dynamic Assessment Strategy</th>
<th>7.14</th>
<th>0.10</th>
<th>0.16</th>
</tr>
</thead>
<tbody>
<tr>
<td>LID</td>
<td>0.23</td>
<td>0.18</td>
<td>0.24</td>
</tr>
<tr>
<td>Phonological Awareness BLS</td>
<td>0.31</td>
<td>0.57</td>
<td>0.10</td>
</tr>
<tr>
<td>Rapid Automated Naming BLS</td>
<td>2.00</td>
<td>0.03</td>
<td>0.13</td>
</tr>
<tr>
<td>Sentence Repetition BLS</td>
<td>0.68</td>
<td>0.69</td>
<td>0.17</td>
</tr>
<tr>
<td>$R = 0.63, R^2 = 0.40, R^2_{\text{adj}} = 0.30, F(4, 31) = 4.11, p &lt; .01$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ change = 0.12, $F(4, 31) = 1.48, p = .23$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$= 0.18, F(2, 45) = 6.12, p < .01$, DIBELS oral reading fluency ($R^2 = 0.17, R^2_{\text{adj}} = 0.14, F(2, 45) = 4.66, p < .01$) and the WRMT ($R^2 = 0.32, R^2_{\text{adj}} = 0.28, F(2, 35) = 8.26, p < .001$)
accounting for approximately 18%, 14%, and 28% of the variance respectively. Model 1
significantly predicted first grade reading as defined by the narrow view of reading
composite measure, $R = 0.56, R^2 = 0.31, R^2_{\text{adj}} = 0.27, F(2, 34) = 7.70, p < .01$, accounting for
approximately 27% of the variance.

Table 6 also shows the results of Model 2. The results indicated that the Spanish
measures (Spanish phonological awareness and Spanish rapid automatized naming with
NWF and Word ID, and Spanish phonological awareness, Spanish rapid automatized
naming and Spanish sentence repetition with ORF and NVR), when entered into the
regression equation simultaneously, were significantly predictive of word recognition
performance at first grade. The results of Model 2 showed that the kindergarten Spanish measures significantly predicted first grade performance on the DIBELS nonsense word fluency, $R^2 = .22$, $R^2_{adj} = .19$, $F(2, 45) = 6.48, p < .01$, DIBELS oral reading fluency, $R^2 = .26$, $R^2_{adj} = .20$, $F(3, 44) = 5.01, p < .01$, and the WRMT, $R^2 = .32$, $R^2_{adj} = .28$, $F(2, 35) = 8.23, p < .001$, accounting for approximately 19%, 20% and 28% of the variance respectively. Model 2 significantly predicted first grade reading as defined by the NVR composite measure, $R = .56$, $R^2 = .31$, $R^2_{adj} = .25$, $F(3, 33) = 4.91, p < .01$, accounting for approximately 25% of the variance.

The results of Model 3, which included the BLS values of the phonological awareness, rapid automatized naming and sentence repetition measures, indicated that those measures that reflected the higher score in either English or Spanish significantly predicted first grade word recognition performance. The BLS measures significantly predicted first grade performance on the DIBELS nonsense word fluency, $R^2 = .24$, $R^2_{adj} = .19$, $F(3, 44) = 4.57, p < .01$, DIBELS oral reading fluency, $R^2 = .21$, $R^2_{adj} = .15$, $F(3, 44) = 3.85, p < .05$, and the WRMT, $R^2 = .36$, $R^2_{adj} = .31$, $F(3, 34) = 6.45, p < .001$, accounting for approximately 19%, 15% and 31% of the variance respectively. Model 3 significantly predicted first grade reading as defined by the NVR composite measure, $R = .59$, $R^2 = .35$, $R^2_{adj} = .29$, $F(3, 33) = 6.01, p < .01$, accounting for approximately 29% of the variance.

Because the letter identification task was considered neither an English-only or Spanish-only task, it was decided to assess the individual, relative contribution that variable had on the regression analysis. We originally considered placing the letter identification measure in with the BLS measures, but we determined that in so doing we
would have effectively weighted the BLS measures, precluding a fair comparison to the language-specific measures. Thus, model 4 was a multiple regression analysis with two ordered sets of predictors. One analysis included the letter ID measure, while the second analysis included the letter ID measure and the three BLS measures (PA:BLS, RAN:BLS, SI:BLS). The linear combination of the measures was significantly related to the four different criterion measures. For nonsense word fluency, $R^2 = .25, R^2_{adj} = .19, F(4, 43) = 3.66, p < .01$, for oral reading fluency, $R^2 = .23, R^2_{adj} = .15, F(4, 43) = 3.14, p < .05$, for Word ID, $R^2 = .47, R^2_{adj} = .41, F(4, 33) = 7.32, p < .0001$, and for the NVR composite, $R = .62, R^2 = .29, R^2_{adj} = .21, F(4, 32) = 5.10, p < .01$. The BLS measures did not significantly predict NWF, ORF, Word ID or the NVR reading composite over and above Letter ID, $R^2$ change = .09, $F(3, 43) = 1.79, p = .16$; $R^2$ change = .09, $F(3, 43) = 1.63, p = .20$; $R^2$ change = .07, $F(3, 43) = 1.47, p = .24$; and $R^2$ change = .11, $F(3, 32) = 1.88, p = .15$, respectively.

In an attempt to determine the extent that the dynamic assessment strategy measure was predictive of first grade reading, it was compared to a linear combination of all of the static measures previously identified as accounting for the greatest amount of variance in the criterion measures. Model 5 was a two-step procedure in which the first step included the dynamic assessment strategy measure and the second step included the dynamic assessment strategy, letter identification and the BLS measures (PA, RAN, SR). For NWF, the results indicated that letter ID and the BLS measures did not account for a significant amount of variance over and above the dynamic assessment strategy measure, $R^2$ change = .06, $F(4, 42) = .89, p = .48$. Similar results were found with ORF, $R^2$ change = .05, $F(4, 42) = .69, p = .60$, and the NVR composite, $R^2$ change = .12, $F(4, 31) = 1.48$, respectively.
\( p = .23 \). In contrast, the results of the regression analysis indicated that letter ID and the 
BLS measures accounted for a significant amount of variance over and above the 
dynamic assessment strategy measure when using Word ID as the criterion measure, \( R^2 \) 
change = .22, \( F(4, 32) = 3.25, p = .02 \). The dynamic assessment strategy measure 
accounted for 26% of the variance of Word ID while the dynamic assessment strategy, 
LID and BLS measures combined accounted for 47% of the variance of Word ID.

Because the letter identification measure (model 4) and dynamic assessment 
strategy measure (model 5) accounted for significant variance over and above all other 
predictor measures, we conducted a post-hoc analysis to compare those two measures to 
determine whether they accounted for the same or unique variance in the criterion 
measures. When we entered the dynamic assessment measure first in the regression 
equation and then entered the dynamic assessment and the letter identification measures 
second, the results were nearly identical to model 5, where the letter identification 
measure did not account for significant, unique variance over and above the dynamic 
assessment strategy measure for NWF, ORF, and the NVR composite first grade criterion 
measures, \( R^2 \) change = .03, \( F(1, 45) = 1.93, p = .17 \); \( R^2 \) change = .02, \( F(1, 45) = 1.26, p = 
.27 \); \( R^2 \) change = .08, \( F(1, 34) = 3.98, p < .05 \), respectively, yet the letter identification 
measure did account for significant, unique variance over and above the dynamic 
assessment strategy measure for the Word ID first grade criterion measure, \( R^2 \) change = 
.18, \( F(1, 35) = 10.87, p = .002 \). When we entered the letter identification measure first in 
the regression equation, and then entered letter identification and dynamic assessment 
measures second, the dynamic assessment measure accounted for significant, unique 
variance over and above the letter identification measure with NWF, ORF, and the NVR
composite first grade criterion measures, $R^2$ change = .08, $F(1, 45) = 4.91, p = .03$; $R^2$ change = .00, $F(1, 45) = 5.41, p = .03$; $R^2$ change = .08, $F(1, 34) = 4.12, p > .05$, respectively, yet not with the Word ID criterion measure, $R^2$ change = .03, $F(1, 35) = 2.03, p = .16$.

**Classification Analysis**

One of the main purposes of the kindergarten measures was to accurately identify Hispanic children who were at-risk for reading difficulty. Classification analysis as described by Lichtenstein and Ireton (1984) was used to determine the sensitivity and specificity of the measures. Kindergarten children were classified as at risk if they performed at or below the 20th percentile (Rathvon, 2004) on 1 or more of the kindergarten measures or if they met the at risk criterion on the dynamic assessment decoding strategy measure, which was a score of 20 or lower. Poor reader status in first grade was determined by scoring at or below the 20th school district percentile on the NWF or ORF measure or at or below the 20th percentile on the WRMT-R word identification measure, using the test's published norms. We considered including the composite NVR score in the classification analysis as a criterion measure, however, the criterion for at risk on the NVR measure would be, in our opinion, commiserate with a score at or below the 20th percentile on any 1 of the 3 formative criterion measures (NWF, ORF, Word ID). That is, if NWF, ORF, and Word ID are representative, formative measures of reading, narrowly defined, then difficulty with any one of those measure would be sufficient evidence of word level reading difficulty.
Participants who were at or below criterion on a single or group of kindergarten predictor measures and also at or below the 20th percentile on a first grade criterion measure were classified as true positives. Participants who were above the criterion on both the kindergarten and first grade measures were classified as true negatives. False positives included students who met the kindergarten criterion but were not identified on first grade measures as having reading difficulty, whereas false negatives included students who did not meet the kindergarten criterion but scored at or below the 20th percentile on the first grade reading measure.

Sensitivity, the ability of a kindergarten measure to identify those first grade students who scored at or below the 20th percentile on the reading measure, was calculated by dividing the number of true positives by the sum of the true positives and false negatives. Specificity, the ability of the kindergarten measure to identify students who were reading above the 20th percentile on a reading measure, was calculated by dividing the number of true negatives by the sum of the true negatives and false positives. The results are presented in Table 6.

Prior to analyzing the classification accuracy of the predictor measures, a potential confound to the findings was explored. In this era of early identification and early intervention, it is difficult to obtain a sample of participants that received equivalent reading instruction from kindergarten to first grade. The Ogden City School District, from which the sample of students in this study were derived, used several reading measures to examine reading ability from the beginning of kindergarten through high school. Based on the results of reading assessments administered three times per school year, students were given intensive reading instruction, strategic reading instruction, or
general reading instruction. These levels of reading instruction differed in intensity and in the precision of treatment targets. Differential instruction is a potential confound for predictive research. Differential instruction can disrupt (in a positive direction) the course of reading development if evidence-based practices are applied in sufficient intensity (Torgesen et al., 2001). Predictor measures administered at kindergarten that were highly reflective of reading difficulty cannot be validated at a later time if the reading outcome was altered through intensive reading instruction.

To explore the possible confound of differential reading instruction, a series of sensitivity and specificity plots were drafted using reading instruction as the predictor measure and the NWF, ORF, and Word ID measures as first grade criterion. The results indicated that only 53% of the participants who were identified as having reading difficulty at first grade received intensive reading instruction during kindergarten, and only 64% of the students identified as having reading difficulty at first grade were receiving intensive instruction at that time. These findings suggest that a large percentage of participants who were identified as having reading difficulty at first grade never received intensive reading instruction in kindergarten or first grade. These findings improved our chances of accurate classification.

The classification analyses indicated that the kindergarten dynamic assessment strategy measure was perfectly sensitive (demonstrating 100% sensitivity) in identifying those participants who would score at or below the 20th percentile on the first grade ORF measure and the first grade Word ID measure. This level of sensitivity is meaningful when considered in conjunction with specificity, which was 80% and 88% respectively.
Findings also indicated that the dynamic assessment decoding strategy measure resulted in excellent sensitivity (86%) and excellent specificity (85%) for NWF.

In contrast, a combination of the English-only kindergarten measures and a combination of the BLS kindergarten measures, including letter ID, correctly identified 57% and 86% of the participants who scored at or below the 20th percentile on the NWF measure, with 77% and 64% specificity. The same two combinations of kindergarten measures each correctly identified 75% of the participants who scored at or below the 20th percentile on the oral reading fluency first grade measure, with specificity equaling 76% and 60%, respectively. The English-only kindergarten measures had 75% sensitivity and 82% specificity on Word ID while the BLS and LID measures had 100% sensitivity and 65% specificity on Word ID.

Because we consolidated the scores of the dynamic assessment decoding strategy measure with the examiner modifiability rating, we were interested in examining the extent that the dynamic assessment accurately classified the participants using only the objective scoring procedure. Thus, we removed the examiner response to instruction modifiability scores from the dynamic assessment decoding strategy measure, used a cutoff score of 20, and calculated the sensitivity and specificity of that measure with each of the first grade formative criterion measures. Results indicated that the dynamic assessment decoding strategy measure, with the modifiability rating removed, had 100% sensitivity and 62% specificity for ORF, 100% sensitivity and 68% specificity for Word ID and 86% sensitivity and 76% specificity for NWF.
Table 7
Classification Analysis

<table>
<thead>
<tr>
<th>Kindergarten measure</th>
<th>NWF</th>
<th>ORF</th>
<th>Word ID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>English-only</td>
<td>57%</td>
<td>77%</td>
<td>75%</td>
</tr>
<tr>
<td>BLS and LID</td>
<td>86%</td>
<td>64%</td>
<td>75%</td>
</tr>
<tr>
<td>DA decoding strategy</td>
<td>86%</td>
<td>85%</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Note. BLS = best language score, LID = letter identification, DA Decoding Strategy = Dynamic Assessment Decoding Strategy, NWF = first grade nonsense word fluency, ORF = first grade oral reading fluency, Word ID = first grade word identification.*

**Results Summary**

English-only and Spanish-only kindergarten measures accounted for significant variance in the first grade criterion measures, which were English reading tasks. However, the Spanish-only kindergarten measures did not account for a significant amount of variance over and above the English-only kindergarten measures. The kindergarten BLS measures accounted for significant variance in the first grade criterion measures, however, the BLS measures did not account for significant variance over and above the individual contribution of letter identification. The greatest $R^2$ values derived from the static predictor measures were from the linear combination of letter ID and the BLS measures. In order to assess the robustness of the predictive evidence of validity of the dynamic assessment strategy measure, it was entered individually in the first step of a regression analysis and was included in a second step with the letter ID and BLS measures. This procedure weighed the dynamic assessment measure individually against
the best combination of the dynamic and static measures. Results indicated that the
dynamic assessment strategy measure individually accounted for a significant amount of
variance in all four of the first grade criterion measures. Further, the results indicated that
the kindergarten dynamic assessment strategy measure accounted for comparable
variance in three of the four first grade criterion measures (Word ID the exception) when
compared to the best combination of dynamic and static English and Spanish measures.
The dynamic assessment strategy score accounted for 19% to 26% of the variance in the
criterion measures whereas the linear combination of the dynamic assessment strategy
measure, the letter ID measure and the BLS measures accounted for 17% to 39% of the
variance.

The dynamic assessment strategy measure yielded superior classification accuracy
over any other individual static predictor measure and over any combination of static
predictor measures. The dynamic assessment strategy measure resulted in perfect
sensitivity (100%) and moderately-high specificity (80%) for ORF, perfect sensitivity
(100%) and high specificity (88%) for Word ID, and high sensitivity (86%) and high
specificity (85%) for NWF.
CHAPTER V
DISCUSSION

Study Overview

This research was designed to assess the validity of a group of kindergarten screening measures. The study was driven by questions about the content-relevancy, construct, and predictive evidence of validity of individual and combined static and dynamic measures administered at kindergarten that were directionally hypothesized to account for significant variance in intermediary, formative criterion measures in English. The criterion measures were reflective of the construct of reading, narrowly defined.

At the outset of the study, it was necessary to delineate the construct of reading in accordance with contemporary views of validity. The measurement of validity is no longer confined to correlation coefficients between a new measure and a criterion. Instead, validity is viewed as a unitary concept evidenced by carefully defined constructs (Messick, 1993) and the assessment of those constructs, which yield appropriate inferences and actions (Messick, 1989). The narrow view of reading was the theoretical foundation whereby predictor and criterion measures, with empirical abetment, were selected for this study. We hypothesized that if an early reading assessment were conceptually driven by the narrow view of reading, its results would be more likely to engender valid inference and prompt specific action. This is due to the ostensible connection between the narrow view of reading and related word recognition skills.

Having defined the construct of the criterion we wished to predict (reading, narrowly defined), we then turned our attention to the content-relevancy of the
assessment items, which, individually or in combination would be used to predict that construct. We carefully analyzed the extant early reading predictive research appertaining to the general population, culturally and linguistically diverse populations, and finally, the Hispanic population. Our ability to identify measures that were specifically predictive of the narrow view of reading was often impeded by construct irrelevant variance (Messick, 1993). A high percentage of previous research was built upon the theoretical foundation of the simple view of reading, which potentially conflated two distinct constructs -- word recognition and comprehension. Thus, the review of research yielded a large number of potential predictor measures that were ambiguously connected to the narrow view of reading. The selection of predictor measures was also driven by first and second language factors, which were necessary to explore due to the bilingual nature of the participants. Additional predictor measures were derived from hypothetical postulation, yielding subtests consequent of dynamic assessment: dynamic assessment raw gain and residuum measures, a dynamic assessment reading strategy measure, and a dynamic assessment working memory measure reflective of temporal distinctiveness.

Each of the predictor measures were administered to the bilingual participants of this study during kindergarten, whereas criterion measures appertaining to the narrow view of reading were administered during first grade. Correlational and intercorrelational analyses (questions 1 and 2) were used to identify predictor measures that were significantly correlated with the criterion measures. Intercorrelation analyses offered information concerning the potential shared variance of the measures—information useful in interpreting the results of multiple regression analysis. Once predictor measures had been selected, further investigation ensued regarding their reliability, construct
validity, and content-relevancy. Reliability (question 3) was primarily examined using Cronbach’s estimation of internal consistency. Construct validity was examined via evidence of the test items’ convergent, concurrent correlation with other measures designed to represent similar constructs. Content-relevancy was, apart from the previously referenced correlation and intercorrelation analyses, carefully considered when constructing and selecting the measures used in this investigation. Multiple regression analyses were also used to evidence content-relevancy. Multiple regression aided in the extraction of influential predictor variables from the large group of variables accumulated and assessed. In this study, multiple regression was used to estimate the effect size of each of the models tested, focusing on the adjusted $R^2$ coefficient of determination statistic. The adjusted $R^2$ statistic offered information regarding the degree that the significant linear relationship between the predictor and criterion measures was meaningful (Question 4a). Once predictor measures had been selected and analyzed, their individual and/or combined predictive evidence of validity according to classification accuracy was examined (Question 4b).

**Correlations**

**Bivariate Correlations Between Predictor and Criterion Measures**

Correlation coefficients were permitted considerable influence in this study. We used the initial correlation coefficient matrix to aid in sorting the multiple kindergarten measures into two groups; those that were significantly correlated with the first grade narrow view of reading construct and those that were not correlated. These multiple, pairwise correlations produced some unexpected results. Based on the results of their
meta-analysis, August and Shanahan (2006), suggested that context-specific variables including socioeconomic status, preschool history, and language ability may be predictive of reading ability for culturally and linguistically diverse children. We found that none of those measures were significantly correlated with first grade word-level reading ability. The SES measure was calculated using a four-factor Hollingshead scale, which has evidence of reliability and validity (for a review see Cirino, Chin, Sevcik, Wolf, Lovett, & Morris, 2002). The correlation between the Hollingshead SES scores and the criterion measures were low, ranging from $r = .12$ to $.18$. A lack of variance in the SES measure could account for the low correlation statistic. The mean weighted Hollingshead score was 14.3 with a standard deviation of 7.2, well within the criteria for low SES status even when scores were above 2 standard deviations from the mean.

Even lower positive (and higher negative) correlations were found between the first grade criterion measures and years of expressive language (English, Spanish, and bilingual), which ranged between $r = -.10$ to $.14$, and preschool attendance, which ranged between $r = -.03$ to $.05$. There was considerable range in preschool attendance (given the low number of years possible). Each participant who had attended preschool had been enrolled in a Head Start program. The number of years of preschool attendance ranged from 0 to 3, with 27 participants having never attended Head Start, 13 participants having attended 1 year of Head Start, 17 participants having attended 2 years of Head Start, and 4 participants having attended 3 years of Head Start. These preliminary results indicated that attendance of a Head Start preschool for the participants did not account for significant variance in first grade word-level reading ability.
Language ability, as measured by a combination of BESA morphosyntax and semantics screener scores and language dominance, correlated to the highest degree with the first grade measures, \( r \) scores ranging from .11 to .28, however, the results were non-significant. Language ability, to the extent that it varied within the participants of this study, may not have a strong bearing on first grade word recognition ability.

These results indicate that the context-specific measures administered at kindergarten had little influence on word-level reading ability for the Hispanic children. By disambiguating the construct of the criterion measures used in this study through the narrow view of reading, directional hypotheses regarding specific predictor measures can easily be generated when contradictions between this study and previous studies using less defined reading constructs emerge. That is, if a measure had previously been identified as predictive of reading, and the reading construct delineated was ambiguous in terms of word recognition and comprehension, our contradictory findings implicate comprehension as being the potential beneficiary of the predictor measure in question. Thus, we hypothesize that predictor measures related to language ability and SES would have considerable influence on language comprehension, which would, in turn, affect reading comprehension.

We hypothesized a significant correlation between preschool attendance and first grade word-level reading ability. However, we found that preschool attendance was correlated with the first grade criterion measures to a lesser degree than any other predictor measure assessed. These findings were troubling, and prompted an investigation on our part to determine first, whether our expectations of the Head Start program in Ogden Utah were ill conceived. Interestingly, the goal of the Ogden-Weber Community
Action Partnership, INC: Head Start program was, “to develop the overall social competence of children. This goal is achieved by recognizing that the families are the prime educators and by working closely with families to become self-sufficient, helps the child in the long term” (http://www.owcap.org/HeadStart/headstart.htm). This statement suggests that our academic expectations of the Ogden, Utah Head Start program were not commiserate with their stated goals.

Other correlations were interesting due to their significance and relative strength. Sentence repetition (BLS) surfaced as a fairly robust correlate of word-level reading \((r = .47)\), and also correlated moderately with language ability \((r = .42)\). Sentence repetition may be a unique measure in that it possibly accounts for both word recognition and comprehension ability. In accordance with previous research, letter identification, phonological awareness and rapid automatized naming were moderately to strongly correlated with word recognition ability. Of all the predictor variables, the dynamic assessment measure elicited behavior most reflective of the actual reading process, and once the participants’ response was analyzed within a framework of modifiability, the dynamic assessment measure consistently had the higher correlations with three out of the four first grade reading measures. This finding aligns with the results of prior meta-analyses, in that actual reading measures were most predictive of reading ability (cf. Hammill, 2004).

**Intercorrelations Among Predictor Measures**

A number of the predictor measures accounted for a similar proportion of variance in first grade word-level reading performance. Some of the large correlations were expected. For example, the English rapid automatized naming measure and the
rapid automatized naming best language score (BLS) were correlated to a very large degree \( (r = .95) \). This extremely high correlation was due to the fact that the majority of best language scores for the English/Spanish rapid automatized naming measures were scores derived from the English measure (see Table 3). The same explanation is not applicable for the high correlations between the English and Spanish phonological awareness measures. It appears that phonological awareness is an underlying skill that transcends specific languages. This finding is consistent with previous research that has found that phonological awareness ability is consistent across languages. The high correlation between the English and Spanish phonological awareness measures indicate that the administration of both measures may not be necessary to validly measure phonological awareness ability in bilingual children.

The letter identification and dynamic assessment strategy measures were highly correlated \( (r = .64) \). Recall that participants did well on the dynamic assessment strategy measure if they used an analogous decoding strategy to recode the nonsense words. Recoding words using an analogous procedure required knowledge of initial letter-sound correspondence. Thus the correlation between the letter identification measure and the dynamic assessment strategy measure was expected.

**Reliability**

No assessment can have sufficient validity if it does not have acceptable reliability. An assessment that is reliable produces similar scores across conditions and situations (Venn, 2000). Reliability can be measured through several means, including split-half and test-retest procedures. It is also possible to assess reliability through the
calculation of Cronbach’s alpha, which represents the internal comparison of test items. Cronbach’s alpha was conducted with those measures that had more than 1 sub-item. Reliability, when measured using internal-consistency estimates, is sensitive to the number of items included in the test and the interrelations among the test items (Pedhazur & Schmelkin, 1991). The fewer test items a measure contains, the greater the likelihood to obtain lower coefficient reliability estimates. This places a screening instrument at a disadvantage, in that the few items contained therein must have little variance. Our examination of the reliability of the predictor measures offered evidence of satisfactory reliability, ranging from .67 to .96. While not perfect, the reliability of the measures was sufficient to warrant the further investigation of validity.

Content Relevancy

Surface Evaluation of Content Relevancy

Content-relevance can be determined in part by examining the extent to which test items evoke measurable behavior that is reflective of the construct under investigation (Anastasi & Urbina, 1997). Expert opinion can also render additional confidence in the relevancy of the content of an assessment. Content-relevance was of primary concern when the kindergarten and criterion measures were constructed and selected for this investigation. We previously described in detail the contents of each context-specific, static, and dynamic measure administered to the participants, however, we have not yet expounded upon the degree to which those measures were administered successfully, nor have we discussed the extent that, based on our observations, those measures prompted
target behaviors. Here we have confined our discussion to those measures included in the regression analyses and classification analyses.

The letter identification measure included upper- and lower-case letters in a standard font type along with letters depicted in more unusual fonts. We observed that ceiling and floor effects were avoided, and that the assessment procedures were easy for the examiners and participants to understand. These factors contributed to the facile documentation of the target behavior—letter naming.

The phonological awareness measures (English and Spanish) contained common names that were expected to be familiar to the majority of English and Spanish speaking children. The words were also selected based on the ease with which they could be separated into constituent words (dog-house) or individual phonemes (/m/-ake). The phonological awareness measures required the participants to attend to phonological and phonemic components of spoken words. The participants were initially trained to perform the phonological awareness task using pictures and examiner cues. This level of training reduced the possibility that the subjects’ responses to the task were due to limited understanding of the procedure. The phonological awareness measure administered to the participants prompted a number of incorrect responses. The means and standard deviations of the test results indicated that the participants only averaged 2.4 and 2.8 correct responses (out of 10) in English and Spanish (respectively). There are several potential reasons for this low response, including the possibility that the test was discontinued too quickly (the examiners were instructed to terminate the phonological awareness subtests after a participant made 3 consecutive incorrect responses). It is also possible that the test items were too difficult for the majority of participants. A
comparable phonological awareness measure was used in a study by Catts et al. (2001), which generated similar floor effects with monolingual, English-language-dominant children.

In an unpublished dissertation, Bridges designed a dynamic assessment of phonological awareness in direct response to the floor effects encountered by Catts et al. (personal communication, November 20, 2008). Bridges reported that the dynamic assessment of phonological awareness measure greatly improved the distribution of scores, and did not evidence floor effects. Specter (1992) employed a dynamic assessment of phonemic awareness at early kindergarten to predict word recognition ability. The results indicated that the dynamic assessment had superior predictive evidence of validity over a static measure of phonemic awareness. Also, Laing, Kamhi, and Catts (1997) compared the extent that static and dynamic measures of phonological awareness predicted early reading ability. Laing et al. used graduated prompts to assist children in producing sounds in words. The assessment was scored based on the number of prompts required for successful completion of the tasks. Results indicated that the dynamic assessment of phonological awareness measure was better at predicting reading ability over static measures of phonological awareness.

It is possible that a dynamic assessment of phonological awareness would reduce floor effects and result in improved predictive evidence of validity for bilingual, Hispanic children. When considering the initial design and content of the measures used in this study, we considered the potential benefits of using a dynamic assessment of phonological awareness, however, we hypothesized that the dynamic assessment nonsense word measure would account for an overlapping proportion of variance derived
from any phonological awareness measure – static or dynamic. The correlation between the dynamic assessment decoding strategy score and the static phonological awareness score was moderately large ($r = .46$), yet, in our opinion, small enough to warrant further investigation into the use of a dynamic assessment of phonological awareness task in conjunction with the dynamic assessment decoding strategy measure. The rapid automatized naming measures administered to the participants included pictures of common objects (house, cat, car) in basic colors (red, blue, black), however, the results of this study revealed that the Spanish rapid automatized naming measure contained items that many of the participants were unable to name consistently. The rapid automatized naming measures included a pre-training sequence designed to help the participants understand the procedures and test items. Based on the results of the Spanish rapid automatized naming measure, however, it is clear that the participants could have benefited from more instruction on the objects and colors during the training phase of the test, or that the items of the Spanish rapid automatized naming measure needed to better reflect Spanish speaking children’s vocabulary. Finding even more simple Spanish vocabulary for Hispanic children who are 5-years old, however, would probably be a futile act that misses the point. The test was already designed to elicit very simple Spanish words. It is important to note that 79% (38) of the participants were unable to complete the Spanish rapid automatized naming task because they did not know the Spanish names for the colors red, blue, black (rojo, azul, negro), the objects house, car, cat (casa, carro, gato), or a combination of those Spanish words. In contrast, all but 1 of the participants was able to complete the English rapid automatized naming task. Of the participants able to complete both the English and Spanish rapid automatized naming
measures, only 5 performed better on the Spanish task over the English task (see Table 3). This preliminary analysis of the rapid automatized naming measures may indicate that the Spanish measure, in its current design, has limited value and should be significantly altered or eliminated.

The English and Spanish sentence repetition measures contained sentences that increased in syntactic complexity, with the earlier items containing fewer subordinate clauses, elaborated noun phrases and basic vocabulary. This assessment was relatively easy to administer, and the participants did not appear to have difficulty understanding the task required of them. Floor effects were not nearly as evident in the results of the sentence repetition measures as they were in the phonological awareness tasks, but the students mean scores were still only 2.4 and 2.6 on the English and Spanish measures (respectively). The median, however, was 2.0 for the English sentence repetition measure as opposed to 1.0 for the English phonological awareness measure. Regardless, further refinement of the sentence repetition measures, especially when internal consistencies of reliability are considered, is warranted in our estimation.

The dynamic assessment decoding strategy measure used nonsense words that were pseudowords, or words rarely used in English or Spanish (nad, kad, tad, zad). The procedure used during the teaching phase of the dynamic assessment utilized an evidence-based decoding strategy, and the feedback was non-contingent, in concordance with findings of a systematic review of dynamic assessment by Caffrey et al. (2008). Administration of the dynamic assessment procedures were adhered to with little difficulty by the examiners, and the participants appeared to understand the tasks required of them.
In general, the design and content of the measures seemed adequate to elicit measurable behavior in English and Spanish that was reflective of the desired construct. However, our retrospective, surface examination of content relevancy revealed that several measures would likely benefit from changes. Some of those changes might only require the alteration of 1 or 2 words in a sentence, whereas other changes might involve the transformation of the structure of an assessment from a static measure to a dynamic measure.

**Multiple Regression Analyses**

Multiple regression analysis can be considered a means whereby content-relevance is assessed. Multiple regression analysis offers evidence of a significant linear relationship between predictor and criterion measures, and provides estimates of effect sizes. In the regression analyses conducted in this study, predictor measures were grouped according to language (English, Spanish or BLS) and assessment type (static or dynamic).

The degree of multicollinearity between the predictor measures used in this study, as discussed previously, precludes a clear interpretation of the regression analysis results. While the multicollinearity among the predictor variables is probably not high enough to warrant the use of ridge regression or principal components regression (Cohen et al., 2003), the moderate degree of multicorrelation does imply that caution should be taken when interpreting the results of the regression analysis.

Several important findings emerged from the regression analyses conducted. Models 1 and 2 of the regression analysis indicated that the kindergarten English-only
and Spanish-only static measures (phonological awareness, rapid automatized naming, and sentence repetition) significantly predicted word-level reading at first grade. However, the Spanish measures did not significantly account for unique variance of first grade reading performance over and above the English measures. This finding has considerable implications for the assessment of risk for reading difficulty for bilingual kindergarten children. It appears that the inclusion of Spanish measures may not add meaningful predictive evidence of validity over and above English measures for word-level reading at first grade.

Results also indicated that the combined BLS static measures included in Model 3 did not account for a significant amount of variance over and above the individual letter identification static measure (Model 4). While preliminary, these results indicated that the letter identification measure administered at kindergarten significantly predicted first grade word-level reading as well as a combination of the best scores derived from several English and Spanish measures.

Model 5 of the regression analysis indicated that the linear combination of static BLS and letter identification measures did not account for significant, unique variance in the first grade criterion measures over and above the dynamic assessment decoding strategy measure, except when using the first grade Word ID measure as criterion. This finding indicated that the dynamic assessment decoding strategy measure significantly predicted first grade nonsense word fluency, oral reading fluency and the composite word-level reading measures as well as the combination of BLS and letter identification measures.
Results of post-hoc analyses revealed that the letter identification measure did not account for significant, unique variance in NWF, ORF, or the NVR composite over and above the dynamic assessment measure, yet inverse analyses indicated that the dynamic assessment measure did account for unique variance over and above letter identification with those three criterion measures. However, the first grade Word ID criterion measure yielded different results, in that the dynamic assessment strategy measure did not account for unique variance over and above the letter identification measure. Thus, the dynamic assessment strategy measure independently accounted for significant variance in three of the four first grade criterion measures, including the narrow view of reading composite, and the letter identification measure independently accounted for significant variance in the Word ID first grade criterion measure.

According to the multiple regression analyses, the kindergarten dynamic assessment strategy measure was predictive of those first grade criterion measures designed to assess decoding ability directly (i.e., nonsense word fluency) or indirectly (i.e., oral reading fluency), and the letter identification measure was most predictive of the first grade criterion measure designed to assess word identification, which required the ability to recall mental orthographic representations. The pairing of these predictor and criterion measures renders a logical interpretation in that the scoring procedures of the dynamic assessment focused on decoding strategies, and the letter identification measure assessed a child’s ability to access previously stored mental orthographic representations of letters.
Convergent-Correlational Evidence of Validity

Construct validity can be assessed by determining the extent that one measure correlates positively with similar measures that are administered at approximately the same time (Gall et al., 2007; Messick, 1995). The letter identification measures (LID $r = .65$) and English phonological awareness measures ($r = .52$, and $.44$) were moderately to strongly correlated with comparable kindergarten DIBELS measures. The dynamic assessment working memory composite measure did not significantly correlate with language ability, however it did correlate significantly with the DIBELS Phoneme Segmentation Fluency measure ($r = .47$). As previously mentioned, phonological awareness has been shown to account for the same, and an even greater amount of variance in reading ability over and above working memory measures in predictive models. Therefore, we were not surprised to find a relatively high correlation between the kindergarten phonological awareness measure and working memory measure. In our view, however, working memory was worthy of exploration, especially when assessed with bilingual participants.

The kindergarten measures that were expected to relate to language ability (sentence repetition) were analyzed by comparing them to the English and Spanish BESA screener. Both the English and Spanish sentence repetition measures were significantly correlated with same-language BESA scores, providing evidence of construct validity for those two kindergarten measures. Likewise, the sentence repetition BLS measure was significantly correlated with language ability as classified using the BESA results and information about language dominance. This finding indicated that the sentence
repetition BLS measure had evidence of relevance to the construct of language ability. Additionally, the sentence repetition measures administered at kindergarten were the only variables to concurrently correlate with kindergarten language ability, phonological awareness, working memory and first grade word-level reading. This finding may suggest that sentence repetition is capable of accounting for variance in language ability as well as phonological awareness and/or other latent constructs of word-level reading.

The convergent-correlational evidence of validity that we have reported have considerable limitations, and the results thereof should be interpreted with strong caution. This exhortation is based on the evidence that no single or combination of measures used in prior research have yielded acceptable, valid predictive information for later reading ability, and that construct validity, when analyzed through concurrent methods, is wholly dependent upon the validity of the measures of comparison. Each of the correlation coefficients described above offers information regarding the degree of shared, reciprocal association, but little more. That is, the validity of the kindergarten DIBELS and BESA measures have not been assessed in longitudinal research to any meaningful degree, hence, the resulting correlation coefficients carry little meaning other than a representation of the relationship between largely unvalidated measures.

**Predictive Evidence of Validity**

The purpose of this longitudinal study was primarily predictive, and while seemingly counterintuitive, studies of this nature are concerned more with the degree that an assessment is predictive as opposed to why it is predictive. Thus, the preponderant
verification of construct validity was evidenced inferentially in the measures' individual or combined classification accuracy.

The classification analyses indicated that the measure of dynamic assessment decoding strategy had excellent predictive evidence of validity. The dynamic assessment strategy measure yielded 100% sensitivity and 80% specificity for first grade oral reading fluency, 100% sensitivity and 88% specificity for first grade word identification, and 86% sensitivity and 85% specificity for first grade nonsense word fluency. To our knowledge, this is the first prospective, across-grade, early reading study to report these high levels of sensitivity while maintaining acceptable specificity involving monolingual or bilingual participants. Furthermore, classification resulting in 100% sensitivity with 80% or higher specificity was replicated across two first grade criterion measures of distinct design. The dynamic assessment strategy measure was the better classification measure over any combination of English and Spanish static measures. It consistently had superior sensitivity and specificity over the other individual or combined kindergarten measures. This finding is remarkable given that the dynamic assessment procedure required less than 5 minutes to administer and was only administered in English.

There are several possible reasons why the dynamic assessment decoding strategy measure used in this study resulted in high classification accuracy. Previous research has indicated that measures closely resembling authentic reading tasks were most predictive of reading ability for the general population (Hammill, 2004). As mentioned earlier, this finding is not particularly surprising—it is logical that actual reading tasks would best reflect reading ability. Nonetheless, using actual reading tasks in an assessment designed to identify children at risk for reading problems before reading difficulty emerged would
be impossible using static measures. However, because of the test-teach-retest arrangement of the dynamic assessment process, measures that children are not yet capable of doing without assistance can be a focus of investigation. Thus, the successful classification results derived from the dynamic assessment decoding strategy measure may be attributable to its suitability in measuring actual reading ability when standardized guidance is proffered.

The examination of what was modified likely contributed to the classification accuracy of the dynamic assessment decoding measure. When the dynamic assessment sounds gain score, reflective of the raw score gain in sound accuracy from pretest to posttest was calculated, the dynamic assessment was not significantly correlated with the first grade criterion measures. When a modification of that gain score was calculated using a percent residuum analysis, the dynamic assessment was significantly correlated with each criterion measure. Nevertheless, only when the dynamic assessment measure was scored by analyzing in greater detail the response to instruction proffered during the teaching phase, and the scores assigned to different reading strategies evidenced by the participants were differentially weighted to reflect those reading strategies used most frequently by proficient readers, did the measure correlate to a moderately-high to high degree with the criterion measures and yield excellent classification results. Hence it was the degree that the participants’ applied the instructions given during the teaching phase to the posttest that was most predictive of future reading ability. It is interesting to note that even when the examiners’ estimation of modifiability was removed from the dynamic assessment decoding strategy score, specificity was reduced, but sensitivity was not affected. Interpreted, this can mean that just by using the more objective measure of
the response to decoding strategy instruction, the dynamic assessment decoding measure maintained excellent sensitivity, with specificity ranging from 62% to 76%.

Some strong notes of caution, however, are warranted. Principally, the analyses in this study were exploratory with a small sample of bilingual children, and replication across a different, larger sample of bilingual children will render a robust test of the predictive evidence of validity of the measures included in this investigation. Attrition was evident; the kindergarten sample included 63 participants, yet at first-grade 48 of those participants were available for testing. The word identification first grade criterion measure was only administered to 38 of the participants.

The high degree of multicolinearity between the dynamic assessment decoding strategy measure and the English and Spanish static measures suggests that we may have only discovered a more parsimonious means of estimating the same variance in reading ability that was possible previously. This is not to say that a more parsimonious accounting of variance in reading ability is not desired, only that, as the regression analyses have indicated, the measures accounted for less than 50% of the variance in first grade reading ability, and that there is greater accuracy to be obtained.

**Summary and Conclusions**

The dynamic assessment decoding strategy measure emerged from the analyses as a potentially excellent measure for use in classifying bilingual, Hispanic children at risk for future reading difficulty. The dynamic assessment was relatively simple to administer, took less than 5 minutes to complete, and included procedures and content more sensitive
to a culturally and linguistically diverse population over static, language-specific measures.

By using the narrow view of reading as the theoretical construct of the formative criterion measures and the composite reading measure, the results of the dynamic assessment decoding strategy measure are easily interpretable, and could lead to highly specific action for those children identified at risk. In conjunction with the classification accuracy, these findings imply that the dynamic assessment decoding strategy measure has excellent validity via evidence of inferentially derived construct validity and strong evidence of predictive evidence of validity.
REFERENCES


Monograph of the British Journal of Educational Psychology.


APPENDIX
Dynamic Assessment Decoding Strategy Scoring Rubric

Dynamic Assessment Pretest/Posttest Scoring Guide

10 points: Whole Word Guess
           and
           No Evidence of Onset-Rime Strategy
           and
           No Evidence of Initial Sound Learning

20 points: Whole Word Guess
           and
           No Evidence of Onset-Rime Strategy

30 points: Whole Word Guess
           and
           Evidence of Onset-Rime Strategy

40 points: No Whole Word Guess

Student Response to Instruction Scoring Guide

0 points: Difficult

10 points: Moderate

20 points: Easy

Note: If no basic posttest was administered, then the student receives 60 points
CURRICULUM VITAE

Douglas B. Petersen

Address: University of Wyoming
Division of Communication Disorders
Dept. 3311, 1000 E. University Ave.
Laramie, WY 82071
Phone: (307) 766-6405
Email: dpeter39@uwyo.edu

CURRENT POSITION

Assistant Professor
College of Health Sciences
Division of Communication Disorders
University of Wyoming
Laramie, WY

EDUCATION

Ph.D. Disability Disciplines: Speech-Language Pathology
Utah State University 2006-2009
Logan, Utah

Dissertation: Using Static and Dynamic Measures to Estimate Reading Difficulty for Hispanic Children.
(Ronald B. Gillam, Ph.D., Committee Chair).

M.Ed. Speech-Language Pathology, 2002
Utah State University 2000-2002
Logan, Utah

University of Kansas 1999
Lawrence, Kansas

B.A. Communicative Disorders and Deaf Education, 1999
Minor - Spanish
Utah State University,
Logan, Utah
RESEARCH EXPERIENCE

Bilingual Coordinator 2006-present

*Diagnostic Markers of Language Impairment in Bilinguals* [Elizabeth Peña, (PI), Ron Gillam (Co-PI) and Lisa Bedore (Co-PI)]. Five-year, multi-million dollar grant from the National Institutes of Health to study the identification of language impairments in bilingual children. Duties include the supervision of the bilingual research assistant team comprised of several undergraduate and graduate students, training and coordination of Spanish language assessments, coordination of student, parent, and teacher consultation and collaboration and direct administration of Spanish and English language assessments.

Instrument Development 2004-present

Research and development of the *Predictive Early Assessment of Reading and Literacy (PEARL)*. The PEARL is an early literacy screening instrument comprised of English and Spanish language-based, static and dynamic pre-literacy subtests designed to identify mainstream and culturally/linguistically diverse kindergarten students who are at risk for future reading difficulty.

PUBLICATIONS

Peer Reviewed Articles


PUBLICATIONS:

Books/Book Chapters

PROFESSIONAL ACTIVITIES/CURRENT RESEARCH
In Preparation

Gillam, S., Petersen, D. B., Fargo, J., Davis, P., & Gillam, R. B. (in preparation). The comprehensive analysis of the macrostructure and microstructure features of African American children’s narratives. The purpose of this study is to obtain nationally representative developmental normative data related to the expressive narratives of African American children.

Petersen, D. B., Gillam, R.B., Lewis, G., & Reutzel, R. (in preparation). Predictive validity of a bilingual literacy screener for Hispanic children. The purpose of this study is to determine the extent that both English and Spanish measures administered at kindergarten are predictive of reading at first grade for bilingual, Hispanic children.

Spencer, T. D., Slocum, T., & Petersen, D. B. (in preparation). Narrative intervention with preschool children: Duration and Intensity. The purpose of this study is to investigate the effects of narrative intervention on the macrostructural and microstructural features of narratives produced by preschool children.


**GRANT SUPPORT**

Doctoral education funded through a grant from the *National Institute of Deafness and Other Communication Disorders* and a stipend offered by *Utah Regional Leadership Education in Neurodevelopmental Disabilities (URALEND) Interdisciplinary Training Program.*

Contributed to the planning, organization and content of an *Institute of Education Sciences, Goal Two* development grant. Major aims of the grant proposal included the use of experiment design procedures to develop a narrative intervention program for school-age children with language impairment. Submitted October, 2008.

**PRESENTATIONS**

**National and Regional**


Petersen, D.B. (2009). *Using Pre-Literacy Measures and Dynamic Assessment to Estimate Reading Difficulty in Hispanic Children.* Technical Session presented at the Utah State University Graduate Student Research Symposium, Logan, UT, April.


**PRESENTATIONS**

**School Districts**


Petersen, D.B. (2002). In-service presentation: Language, the foundation for literacy: From phonological awareness to fluency. San Diego City Schools.


**TEACHING EXPERIENCE**

**Graduate Courses**

**SPPA 4210 (5140) - Evaluation Procedures in Communication Disorders**  
Division of Communication Disorders, University of Wyoming  
(instructor, Fall 2009)

**COMD 6900 - School-Based Speech-Language Pathology**  
Department of Communicative Disorders and Deaf Education, Utah State University (instructor, Summer 2008).

**COMD 6020 - Language Assessment and Intervention**  
Department of Communicative Disorders and Deaf Education, Utah State University (co-instructor, Fall 2007).

**TEACHING EXPERIENCE**

**Undergraduate Courses**

**COMD 2500 - Language, Speech, and Hearing Development**  
Department of Communicative Disorders and Deaf Education, Utah State University (co-instructor, 2007).
COMD 2500 - Language, Speech, and Hearing Development
Department of Communicative Disorders and Deaf Education, Utah State University (guest lecturer, 2006).

PROFESSIONAL WORK EXPERIENCE

Speech-Language Pathologist 2006-2007
Ogden City School District, Ogden, UT

DUTIES: Responsibilities included the implementation and supervision of district-wide Spanish speech and language assessments, reports, and related intervention.

Speech-Language Pathologist 2004-2006
Aberdeen School District, Aberdeen, WA

DUTIES: Responsibilities and projects included speech and language assessments and intervention for elementary and high school students in English and Spanish. Funding acquisition, training and support for low-tech and high-tech augmentative communication devices. Improve classroom teacher collaboration through increased curriculum material integration in speech and language therapy. Provide multiple in-service/small group seminars to faculty on speech, language and language-based literacy issues. Collaborate with resource teachers, occupational therapists, psychologists, life skills teachers, administrators, community based teachers, Primary Developmental Program teachers and paraprofessionals to promote speech and language in an interdisciplinary model. Translate IEP and IEP related meetings in Spanish, and many of the other responsibilities listed below in previous employment.

Speech-Language Pathologist 2002-2004
San Diego Unified School District, San Diego, CA

DUTIES: Directed a collaborative effort that included phonological and phonemic awareness training, music therapy, modifications of the Visualize and Verbalize Program, the Story Grammar Marker Program, LIPS and the Seeing Stars Program. Developed the Predictive Early Assessment of Reading and Literacy (PEARL) early literacy screening instrument in English and Spanish. Used and promoted reading intervention techniques such as guided reading, shared reading, reading aloud and the Reading Recovery program. Collaborated with resource specialists, occupational therapists, physical therapists, reading specialists, school administration, teachers and parents. Used in-class and pull-out service delivery models during and after school. Directed and participated in faculty in-services
with topics related to phonological awareness, language and literacy, and receptive and expressive language development through literature and narratives. Directed and performed all Spanish language speech and language assessments and interventions. Provided Spanish language translation services for various meetings and parent contact. Wrote professional English and Spanish reports and correspondence.

Speech-Language Pathologist 2002
Primary Children’s Hospital, SLC, Utah

DUTIES: Initiated co-treatments and collaboration with other speech-language pathologists, occupational therapists, physical therapists and parents in a hospital outpatient clinical setting. Worked with clients with autism, Down’s syndrome, cerebral palsy, TBI, stuttering, aphasia, voice disorders and other speech and language disorders. Evaluated and provided therapy in English and Spanish for receptive and expressive language disorders, articulation impairments, dysfluency, voice disorders, and feeding concerns. Used sign language, PECS, TEACCH, social stories and other therapy methods for speech and language impairments. Produced professional reports and writings and communicated with other professionals and organizations to help coordinate services in both English and Spanish.

Speech-language pathologist 2000-2002
Nebo School District, Spanish Fork, Utah

DUTIES: Worked with a large, varied caseload of students with speech and language disorders at multiple elementary schools throughout the school district. Evaluated, assessed and provided speech and language services to students through individual, small group, and in-class service models. Established appropriate long and short-term goals and objectives and directed or participated in IEP/IEP related meetings. Served as a member of the Nebo School District Autism Specialist Team with responsibilities that included parent training, professional in-services, student case management, consultation and direction of services. Served as the speech-language pathologist for two severe autism classes. Performed district-wide Spanish language assessment and intervention for Spanish speaking students in need of SLP services. Diagnosed and/or implemented therapy for disfluency, voice disorders, articulation, cleft palate and oral motor difficulties. Used narrative based assessment and interventions, incorporated sign language, PECS, TEACCH, social stories and other methods in receptive and expressive language treatments. Developed proprietary narrative assessment and intervention methods as well as music based language therapy.
Speech-Language Pathologist Assistant 1998-1999
Private Residence, Logan, Utah

DUTIES: Worked individually with a young child with autism in the home. Participated in, developed, and implemented receptive and expressive language therapy using child-centered, milieu and clinician-directed methods. Collaborated with speech-language pathologists, occupational therapists and family members to create a comprehensive intervention/treatment plan. Used music as a language therapy method. Worked with parents on fundamental language elicitation/intervention methods effective for children with autism.

Speech-Language Pathologist-assistant 1998-1999
River Heights Elementary School, Logan, Utah

DUTIES: Assisted in speech-language therapy programs. Worked with Spanish and English speaking children receiving speech and language services. Performed Spanish language assessments and translation for parent meetings.

ADMINISTRATIVE POSITIONS/PROFESSIONAL SERVICE

Master’s Thesis Committee Chair (present)
Division of Communication Disorders Recruitment Committee (present)
Member of Convention Program Committee for Language Science for the American Speech, Language Hearing (ASHA) Convention. Chicago, IL, November 20-22.

COMMUNITY SERVICE

Spanish speech & language testing, Laramie School District, Laramie, Wyoming, 2009
Volunteer: Boy Scouts of America, Utah 2008 - Current
Volunteer: Cub Scouts of America, Utah 2007, 2008
Volunteer: After school speech-language services, San Diego 2003-2004
Volunteer: Boy Scouts of America Varsity Assistant Coach, San Diego 2002-2004
Volunteer: Youth Leader, Lawrence, Kansas 1999
Volunteer: Reading support provider, Logan Utah 1997
Voluntary religious/humanitarian service, Dominican Republic 1991-1993

PROFESSIONAL ASSOCIATIONS/LICENSURE/HONORS

Finalist for the Robin’s Award for Graduate Student Researcher of the Year, Utah State University, April, 2009.

Research Presentation Award at the College of Education Graduate Student Research Symposium, Utah State University, April 2009.

Awarded the 2008-2009 Graduate Student Researcher of the Year for the College of Education, Utah State University, April, 2009

Nominated for the 2008-2009 Graduate Student Teacher of the Year Award for the College of Education, Utah State University, April, 2009.

Utah State University Department of Communicative Disorders and Deaf Education Graduate Student Academic Award, April 2008.

Research Presentation Award at the College of Education Graduate Student Research Symposium, Utah State University, April 2008.

Member of American Speech-Language Hearing Association (ASHA).

Member of Special Interest Division: Language Learning and Education; American Speech-Language Hearing Association (ASHA).

Member of Special Interest Division: Communication Disorders and Sciences in culturally and Linguistically Diverse (CLD) Populations; American Speech-Language Hearing Association (ASHA).

Member of Special Interest Division: School-Based Issues; American Speech-Language Hearing Association (ASHA).


Deans List, Junior and Senior years, Utah State University (GPA 3.9).

Member Golden Key Honor Society.