2014

The Ability of Early Reading Measures Administered in First Grade to Predict Fourth Grade Reading Comprehension for Puerto Rican Students in English Immersion

Sarah Peller
University of Massachusetts - Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

Part of the Educational Psychology Commons

Recommended Citation
Peller, Sarah, "The Ability of Early Reading Measures Administered in First Grade to Predict Fourth Grade Reading Comprehension for Puerto Rican Students in English Immersion" (2014). Doctoral Dissertations. 22.
https://scholarworks.umass.edu/dissertations_2/22

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
THE ABILITY OF EARLY READING MEASURES ADMINISTERED IN FIRST GRADE TO PREDICT FOURTH GRADE READING COMPREHENSION FOR PUERTO RICAN STUDENTS IN ENGLISH IMMERSION

A Dissertation Presented

By

SARAH PELLER

Submitted to the Graduate School of the University of Massachusetts in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2014

Education

School Psychology
THE ABILITY OF EARLY READING MEASURES ADMINISTERED IN FIRST GRADE TO PREDICT FOURTH GRADE READING COMPREHENSION FOR PUERTO RICAN STUDENTS IN ENGLISH IMMERSION

A Dissertation Presented

by

SARAH PELLER

Approved as to style and content by:

______________________________
William Matthews, Chairperson

______________________________
Amanda Marcotte, Member

______________________________
Elizabeth Harvey, Member

______________________________
Christine B. McCormick, Dean
College of Education
ACKNOWLEDGMENTS

I would like to thank my advisor, William Matthews, for his many years of patience, support and guidance. Great thanks are due to Professor Craig Wells for his years of guiding me through complex statistical analysis even while living in a different state. A special thanks also goes to David Valade for not only allowing me to access student data, but for his hours of both advocating and searching when the data disappeared, obtaining further data and demographic records, and continuing to support the project for the several years that it took.

I would also like to extend a special thank you to both of my parents and all those friends and family members whose support and love helped me continue striving towards the goal even when it seemed impossible.
ABSTRACT

THE ABILITY OF EARLY READING MEASURES ADMINISTERED IN FIRST GRADE TO PREDICT FOURTH GRADE READING COMPREHENSION FOR PUERTO RICAN STUDENTS IN ENGLISH IMMERSION

FEBRUARY 2014

SARAH PELLER, B.A., SARAH LAWRENCE COLLEGE
M.A., UNIVERSITY OF MASSACHUSETTS AMHERST
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor William Matthews

The present study examined the relationship between children's early literacy-related abilities and their reading achievement in late elementary school in the context of a Puerto Rican community in Massachusetts. The researcher examined four years of student achievement test data from the public elementary schools of Holyoke, Massachusetts. The city’s particular sociologic history makes it an interesting and fruitful case for investigating issues around language, culture, and test performance that, while uniquely expressed in Holyoke, share aspects in common with many locales across the United States. The researcher sought to determine the extent to which literacy assessments administered to Hispanic children in Holyoke under Reading First grants and NCLB's high-stakes testing requirements measured the constructs they intended to measure, the value of the testing in terms of its ability to predict the students' performances on the state language arts exam in the fourth grade, and the way that
children’s English language proficiency may have influenced the measures’ predictive ability as well as the students’ progress in acquiring age-appropriate literacy abilities. Confirmatory factor analysis techniques were used to assess the construct validity of literacy-related subtests for this sample. Finally, structural equation modeling was utilized to identify and test a model quantifying the relationships between subjects’ home language, early decoding ability, reading fluency rate and reading comprehension scores. Results suggest that although students whose home language is Spanish perform more poorly that those whose home language is English on both first and fourth grade reading tests, there was no group difference found in the degree to which the early literacy measures predicted their fourth-grade reading comprehension. A single factor comprised of a variety of first grade reading measures was able to account for 56% of the variance of the students’ performance on a reading comprehension state exam in the fourth grade.
CONTENTS

Page

ACKNOWLEDGMENTS ......................................................................................................................... iv

ABSTRACT ......................................................................................................................................... v

LIST OF TABLES ................................................................................................................................. ix

LIST OF FIGURES .............................................................................................................................. x

CHAPTER

1: STATEMENT OF THE PROBLEM ................................................................................................. 1

1.1. Race, language, and educational attainment ......................................................................... 1
1.2. Language policy in the United States ...................................................................................... 4
1.3. How children learn to read ...................................................................................................... 10
1.4. Language development and language proficiency ............................................................... 22
1.5. Language use, proficiency and policy .................................................................................... 30
1.6. Measurement and language difference .................................................................................... 41

1.6.1. Language proficiency testing for classification ............................................................... 41
1.6.2. High-stakes testing .............................................................................................................. 42
1.6.3. Reading disabilities and language difference ................................................................. 45

1.7. Assessment of early skills to predict reading achievement ................................................... 47
1.8. Models with English Language Learners and predictive bias ............................................. 72
1.9. Present study ............................................................................................................................ 80

1.9.1 Purpose ................................................................................................................................... 80
1.9.2. Research questions ............................................................................................................. 80
1.9.3. Design ................................................................................................................................... 81

2. METHODS .................................................................................................................................... 83

2.1. Setting ........................................................................................................................................ 83
2.2. Structured English Immersion program .................................................................................. 86
2.3. ELL classification system ....................................................................................................... 87
2.4. Reading instruction .................................................................................................................... 89
2.5. Participants ............................................................................................................................... 89
2.6. Measures ................................................................................................................................... 91

2.6.1. Descriptions of DIBELS measures .................................................................................... 91
2.6.2. Descriptions of GRADE measures .................................................................................... 91
2.6.3. Description of MCAS fourth grade English Language Arts test ........................................ 92
2.6.4. Tests’ reported reliability coefficients .............................................................................. 92

2.7. Statistical Procedures for Analysis ........................................................................................... 93
3. RESULTS.................................................................................................................................................. 95
   3.1. Test Performance of Entire Sample .................................................................................................. 95
   3.2. Group Differences in Performance Based on Home Language ...................................................... 97
   3.3. Research Question 1: Results of Confirmatory Factor Analysis .................................................... 98
   3.4. Research Question 2: Results of Hybrid Model .............................................................................. 103
   3.5. Research Question 3: Correlations between Measures and Group Differences .......................... 104
   3.6. Research Question 4: Between-Group Comparison on Hybrid Model Fit ....................................... 105

4. DISCUSSION ....................................................................................................................................... 108

5. LIMITATIONS .................................................................................................................................... 128

APPENDICES

   A. MEAN RAW SCORES AND GROUP DIFFERENCES ........................................................................ 135
   B. CORRELATION MATRIX .................................................................................................................. 136
   C. BETWEEN-GROUP COMPARISONS OF CORRELATION COEFFICIENTS .............................. 137
   D. MEASUREMENT MODELS .............................................................................................................. 139

BIBLIOGRAPHY ....................................................................................................................................... 143
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. DIBELS reported reliability coefficients</td>
<td>92</td>
</tr>
<tr>
<td>2. GRADE reported reliability coefficients</td>
<td>93</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hybrid Model: Predicting 4th grade advanced reading comprehension</td>
<td>104</td>
</tr>
</tbody>
</table>
CHAPTER 1

STATEMENT OF THE PROBLEM

1.1. Race, Language, and Educational Attainment

Public schools' education of Hispanic children is strikingly unsuccessful (Lesaux, 2006). Of racial minority youth, Hispanic teenagers are the most likely to drop out of high school. While eight percent of White teenagers drop out, the rate is twelve percent for blacks and 30% for Hispanics (Garcia & Cuellar, 2004). The drop-out rate is one of the more deeply meaningful measures of schooling relevance and effectiveness; it tells us that nearly a third of Hispanic youth ultimately reject public schooling as their path to economic productivity in adulthood.

Severe racial and economic inequality in the American educational system—both in terms of opportunities and outcomes—is as old as the system itself (Bowles & Gintis, 2002; Kozol, 1991, 2005). De facto racial segregation in neighborhoods and schools today is nearly as severe as when integration attempts began (Charles, 2003; Massey & Denton, 1993; Zhou & Logan, 2003). Until relatively recently, Latinos were comparatively well-integrated geographically. But while Black-White segregation has decreased gradually from 1980 to 2000, Hispanics have experienced increasing isolation and segregation (measured with five indices including an index of dissimilarity; Wilkes & Iceland, 2004) and huge population increases due to constant immigration. Part of the decrease in Black isolation is in fact attributable to Hispanics increasingly sharing neighborhoods that previously were dominated by Blacks (Charles, 2003). Latinos are
now more likely than ever before to live in concentrated, isolated neighborhoods, and in several cities they have become "hypersegregated"—a term describing scores above an intensity threshold established by Wilkes & Iceland (2004) in their research on American segregation.

Although race is apparently more salient than socioeconomic class as a cause of residential segregation (Charles, 2003), as Kozol (2005) points out, "racial isolation and the concentrated poverty of children in a public school go hand in hand (p.20)." In 1997, Latino students experienced more isolation from whites and more concentration in high-poverty schools than any other racial group (Baker & Hakuta, 1997). Baker and Hakuta (1997) additionally noted an "exceptionally strong" link between school segregation by race and by poverty, reporting that Latino students are 16.3 times more likely to attend a concentrated poverty school than a student in a mostly-white school. In 1999, the average Hispanic student attended a school that was 57% Hispanic (Zhou & Logan, 2003). High-poverty schools vastly under-perform more affluent schools and are much less likely to send their students to college (Baker & Hakuta, 1997).

With exceptions in areas of California and the Southwest where Mexican-Americans have lived for generations, many Hispanic children in the United States are first or second-generation immigrants. Even after controlling for socioeconomic status and race, immigrant status predicts decreased academic achievement in the United States (August & Shanahan, 2008). Speaking a language other than English is also a risk factor for being unsuccessful in school, even though in the labor market bilingualism can be a
distinct advantage. Language minority students, the wide majority of whom are Spanish-speakers, fail to complete high school at a rate three times higher than the national average (National Center for Educational Statistics, 2000), and those who do not speak English proficiently in high school drop out five times as often as native speakers (August & Shanahan, 2008). According to the U.S. Census of 2003, Hispanics are among the least educated racial group in the United States. Only 11 percent of Hispanics over 25 had earned a bachelor’s degree compared with 17 percent of blacks, 30 percent of whites, and 49 percent of Asian Americans (U.S. Census Bureau, 2003). Over a quarter of Hispanic adults had less than a ninth-grade education.

English language learners generally, as well as racial minority students, on average fail to meet grade-level academic expectations and fall behind average Caucasian native-English-speaking students on most measures of academic achievement (Carrasquillo, Kucer, & Abrams, 2004). According to national statistics provided by the U.S. Department of Education, the achievement gap by both race and home language is already evident by Kindergarten; while white non-hispanic students earned an average of 70.2 on letter recognition, Hispanic English-speaking students scored only 51.2 and Hispanic students who speak Spanish at home earned an average score of 38.3. Most were tested on their recognition of letters in Spanish (Schneider et. al., 2006).

In fourth grade, the gap persisted. In 2002, Hispanic fourth-graders scored 30 points lower than whites in reading, and demonstrated the same size gap in eighth grade. By twelfth grade the gap had shrunk to 18 points, suggesting a narrowing of the gap.
unless the dropout rate among Hispanic students is taken into consideration, which likely inflated the average performances of those who remained in school. At all three grade levels, Hispanics performed slightly better than Blacks, but the difference was barely statistically significant. When Hispanic reading scores are separated by subgroup, Mexican Americans and Puerto Ricans demonstrate the lowest average scores (Schneider et. al., 2006).

1.2. Language Policy in the United States

From 1998 to 2002, a political campaign called "English for the Children" convinced voters in California, Arizona and Massachusetts to ban bilingual education by proposition, with the logic that bilingual classrooms slow the process of English acquisition for students who speak a language other than English at home. Approximately one-third of the California's school-aged population are "English language learners" (ELLs) – the highest proportion of any state (Hakuta, Butler & Witt, 2000). Voters were promised that these children would become fluent in English after only one year in English immersion classes. Despite a great deal of empirical evidence suggesting the superiority of long-term bilingual instruction for higher English language and general academic outcomes (August & Shanahan, 2008; Cheung & Slavin, 2005; Cummins, 2000; Maldonado, 1994; Ramirez, 1991; Snow, Burns & Griffin, 1998; Thomas & Collier, 1997), political propaganda, not parents, researchers or educators, ultimately determined how language minority children were to be educated.

English-only proponents are not shy about their assimilationist agenda. Ron Unz
(2000), the wealthy California real estate magnate who authored the laws, writes, "...our public schools and educational institutions must be restored as the engines of assimilation they once were....Our public schools should provide a single, unifying American culture rather than encouraging ethnic fragmentation (retrieved from onenation.org on November 4th, 2009)". As Gutierrez, Ba quedano-Lopez and Alvarez (2000) put it, the English-only movement is "imbued with racist discourse" (p. 223); many writers have pointed out the thinly veiled hostility towards racial and linguistic diversity in the rhetoric of Unz's political campaigns (see Corson, 2001; Crawford, 2000; Espinoza-Aguilar, 2001; Gonzales, 2001; Hartman, 2001; Horner, 2001).

The United States has always has been a multi-lingual society. New languages have arrived with voluntary immigrants, with slaves, and with people living in lands won from competing colonial powers. Today, the United States is the fifth-largest Spanish-speaking country in the world, after Mexico, Spain, Argentina and Columbia (Gonzalez, 2001). The Spanish were the largest colonizers of North America; much of the Spanish-speaking population of California can trace its history to pre-Anglo times. In South Texas, too, the majority of citizens are of Mexican descent. Many can trace their ancestors to pre-annexation times, and most still use Spanish even though they are also fluent in English (MacGregor-Mendoza, 2000). The state of New Mexico was officially annexed from Mexico in 1848; however, Spanish continued to be the majority language spoken in the territory for many years. In the late 1800's its state legislature was still operating in Spanish while other government proceedings were split about equally
between English and Spanish (Gonzalez, 2001). New Mexico's school language policies have shifted many times (Freeman, 1998).

Language policy that restricts the language use of a colonized group, and/or enforces the use of a new language, is a traditional tactic for assimilating (and controlling) colonized populations (Bratt, 2007; Crawford, 2000; Gutierrez, Asato, & Baquedano-Lopez, 2001; Hakuta & Fillmore, 1998; Skutnabb-Kangas, 2002). Language policies have been at times quite draconian (Crawford, 2000; Freeman, 1998). An early example in the United States is the story of the Cherokees, who had achieved a high rate of literacy in the 1850's by teaching in both English and Cherokee, until the U.S. government imposed an English-only policy, resulting in 40% illiteracy and a 75% dropout rate (Gonzalez, 2001). In 1898, the United States won control over Puerto Rico and declared English the official language there, where the entire population spoke Spanish (Gonzalez, 2001; Macedo, 2000). For the next 50 years, Puerto Rico had an English-only education policy imposed by the United States government. According to historian Juan Gonzalez (2001), "the result was a near-total breakdown of the education system as thousands of students stopped attending classes (p. 211)."

Around the turn of the century language use in schools became a national political issue. According to historian Rebecca Freeman (1998), "changing expectations about the role of schools in society radically transformed the discourses surrounding bilingualism and bilingual education in the United States (p. 33)." The imminent war signaled an intensification of nationalism and paranoia, and massive waves of poor immigrants from
Europe flooded into the country. Education began to be viewed as a powerful vehicle to teach patriotism and common values to bring a vast and diverse society together under one cultural roof (Freeman, 1998; Tyack, 2003). In the 1890's, laws began appearing which restricted German language education. During and after World War I most of the numerous German schools were shut down, and series of legislative actions restricted Chinese and Japanese immigration. In 1906, Congress passed the Nationality Act requiring all new immigrants to demonstrate proficiency in English. A special provision had to be added regarding Puerto Ricans since they were already U.S. citizens who spoke Spanish (Freeman, 1998). After the war, the Immigration Act of 1924 imposed strict quotas reducing immigration for the time being.

With Brown v. Board of Education in 1954 and the powerful movements for equality for African-Americans, equal opportunity became the major issue for education. Ten years later the Immigration and Nationality Act of 1965 abolished most immigration quotas, allowing huge surges of immigrants from Latin America, Asia and Eastern Europe. A dual language program in Dade County, Florida, serving upper-class Cuban immigrants, received national attention in 1963. Throughout the 1960's, the Civil Rights Movement's litany of triumphs in the courts included victories for bilingual education. In 1965, Congress passed the Elementary and Secondary Education Act, adding Title VII, the Bilingual Education Act, in 1968. Although it was never clear whether the Act's intention was to promote bilingualism itself or English language learning, it provided funding for thousands of new experimental bilingual programs all over the country. The
factors leading up to the passage of this legislation included the Civil Rights Movement, the Chicano Movement, the War on Poverty legislation, increased federal involvement in education, and the new wave of research finding positive results for bilingual education (San Miguel, 2004).

In 1974, in Lau v. Nichols, the Supreme Court ruled in favor of parents in San Francisco on behalf of Chinese students who were not receiving supplemental instruction. Justice Douglass, writing for the majority, argued that if a class of students was not succeeding due to their language minority status, discrimination was occurring. However, he also indicated that empirical program comparisons were needed to demonstrate what type of education would guarantee the Chinese students equal opportunity. San Francisco consented to provide bilingual education to Chinese, Philippine, and Hispanic children (San Miguel, 2004) and the Department of Health, Education and Welfare (HEW) was assigned to develop specific procedures to bring school districts into compliance with the Equal Opportunity legislation. It was tasked with figuring out how to identify students who needed specialized language instruction, determining which strategies would help them to both learn English and receive comparable benefits from general education, determining when to mainstream them again, and developing professional standards for involved personnel. The resulting Bilingual Education Act of 1974 required elementary schools to provide specialized instruction in English as a Second Language and subject matter instruction in the students' native language wherever schools had large enough groups of students of a single linguistic origin. The law provided funds for both
transitional and maintenance bilingual education but encouraged the former, which minimizes attention to the first language and quickly mainstreams students into all-English instruction. The new legislative requirements, particularly the Lau ruling, prompted a surge of research on the efficacy of different types of bilingual education programs (San Miguel, 2004).

There are hundreds of studies that have attempted to judge the effectiveness of bilingual education by comparing various types of bilingual programs with structured English immersion (SEI) and other English-only programs. However, implementation quality, curricular differences and other intervening variables have proved nearly impossible to control. Many studies either have found either no differences for language of instruction or seemingly contradictory conclusions. However, several large studies have examined only long-term effects and controlled for relevant mediating variables, such as program quality and socio-economic status, and these indicate that for Spanish-speaking Latino children, language of instruction is one of the most important variables associated with schooling success (Cummins, 1991; August & Hakuta, 1998; Lindholm-Leary, 2001; Saville-Troike, 1984; Thomas & Collier, 1997).

Nearly two hundred empirical studies have reported positive findings of long-term, enrichment bilingual education and strong evidence indicating its superiority over other models (see August & Shanahan, 2008; Carlo, 2007; Cummins, 2000; Lindholm-Leary, 2001; Snow et al., 1998; Thomas & Collier, 2003, Thomas & Collier, 1997). The question of which language-of-instruction program model yields better academic
outcomes is no longer controversial among second language experts and English language educators. A consensus has been reached within the relevant scientific community: Quality, long-term bilingual education consistently produces superior long-term academic outcomes for language minority students. Thomas & Collier's (1997) influential study and many others suggest that shorter-term versions of bilingual education and instruction only in English yield deflated academic outcomes on the average in the long run, mediated by many other variables such as program quality, the amount and quality of native language support, the amount and quality of English language instruction, and the academic expectations of the school setting. The only times that these results have not been replicated in language-of-instruction studies have been when researchers define "immersion" to mean any foreign language education for young children, without distinguishing between majority and minority languages or socio-economic differences (Cummins, 2000; Rossell & Baker, 1996)

1.3. How children learn to read

There are various approaches to the teaching of reading to young children which have been popular during different time periods in the United States, and have even been controversial as researchers, theorists and practitioners have disagreed about what "literacy" really means and what instructional techniques work best to bring children towards the goal of advanced literacy achievement. In 2000, the National Reading Panel, tasked by the federal government to review all of the empirical studies published on the topic, released its report identifying five components of effective reading instruction.
They were: phonemic awareness; phonics; reading fluency; vocabulary; and reading comprehension. The panel recommended that each of these broad skill areas be explicitly addressed in the early literacy curricula of all public schools (National Reading Panel, 2000). Some of these areas develop simultaneously throughout the early years of learning, and others occur in a hierarchy in which one skill is a prerequisite for another.

In 2008, the Federal Department of Health and Human Services funded another meta-analysis of the scientific research, this time focused on best practices for teaching early literacy skills. Results supported the importance of direct instruction in phonemic awareness and phonics for teaching children as young as preschool specific skills that would facilitate their acquisition of basic reading ability (National Early Literacy Panel, 2008).

Written English is achieved using phonetic coding. In other words, in English (as in Spanish and most languages, but not all) words are encoded into print sound-by-sound using symbols (letters) to represent sounds. In English, unlike in Spanish, sometimes groups of letters are required to represent one single sound. Prior to being introduced to print, a child's phonological awareness, or awareness of the sound structure of the language, is relevant to his or her later development of reading ability (McBride-Chang, 2004; Liberman, 1973; Torgeson et al., 1999; Wagner, 1999). It stands to reason that in order to decode the printed symbols that stand for individual sounds and combine them into recognizable parts to then sound out known words, children must first be able to distinguish these sounds in their minds.
Even before birth, children begin processing speech, most very easily. Yet they, and even many adults, are not consciously aware of the individual units that make up speech sounds. Rather, auditory linguistic input for children can be thought of as simply a string of undistinguished sounds making up recognizable phrases used for communicative functions. For example, there is not yet any reason for infants beginning to process spoken language to realize where one word ends and another begins. Over time, infants become skilled at differentiating words with finer and finer distinctions up to those with only single and slight differences in sound, for example distinguishing “pen” and “pan” (McBride-Chang, 2004). The same is said to be happening to their linguistic abilities in terms of word meaning; as children learn more words they are able to use and comprehend increasingly specific words which express finer and finer distinctions between concepts (Vygotsky, 1962).

Prior to or without any targeted instruction to build phonological awareness, very young children demonstrate individual differences in terms of their general phonological processing ability and ability on phonological awareness tasks. Additionally, all children's phonological awareness increases with age. And, there are different levels of awareness that can be measured. The first, or easiest, level typically measured is a child's awareness of syllables, and the last or most difficult level is awareness of phonemes, the smallest sound units that make up words. This latter level of phonological awareness is often called phonemic awareness. Phonemic awareness is the most difficult type to achieve and often requires instruction. Phonological measurement tasks also differ by
whether they measure a child's perception of these sounds, or whether they ask the child to manipulate the sounds (McBride-Chang, 2004).

Phonemic awareness, in particular, is highly relevant to later reading ability because letters and groups of letters represent phonemes. Phonemes are the aural building-blocks of words and putting phonemes together is how basic decoding begins. In written English, phonemes can be represented by a single letter or a group of letters. The visual symbolic representation of sound (letter or group of letters) is known as the grapheme (Uhry, 2005).

Long before children are able to recognize printed symbols, they can be explicitly guided, by caretakers and teachers, to recognize that clauses are comprised of individual words, words are comprised of syllables, and syllables are comprised of phonemes. Techniques for guiding children towards this important discovery include a wide variety of “sound play” aural activities in which children are made aware of syllables and sounds and guided to isolate, identify and/or manipulate onset, middle or ending sounds in words. The easiest and earliest-mastered of these abilities is rhyming, or recognizing and then manipulating the vowel and the final sound in words. Later, children can be explicitly taught to identify and manipulate onset and middle sounds. The development of this awareness then assists children to later acquire the alphabetic principle once printed letters are introduced; they are already aware of the individual sounds that make up morphological units which make up words and can pair the sounds with the visual representation of corresponding letters and letter combinations (McBride-Chang, 2004;

A child's phonological awareness overall, and phonemic awareness specifically, are significant predictors of early reading ability (Torgeson et al., 1999, 1997; Wagner, 1999; NELP, 2008). Phonological abilities have also been identified as an area of deficit for the majority of students experiencing reading failure (Lyon, Shaywitz, & Shaywitz, 2003), and, fortunately, can be remediated (Uhry, 2005; McBride-Chang, 2004; Torgeson & Wagner, 1987; Wagner et. al, 1991; NRP, 2000; Torgeson & Hudson, 2006; NELP, 2008).

To become readers, the next skill that children learn explicitly is letter-sound correspondence, also known as the alphabetic principle. First, typically in kindergarten, children learn the name of each letter of the alphabet, and the most common sound that each letter makes. When the child sees a letter he or she should be able to say both its name and one or more sounds that it makes. Conversely, when hearing a sound the child should be able to write or say the letter corresponding to the sound (Carreker, 2005).

Next, children learn to “blend” the sounds together upon seeing a simple word or nonsense word composed of a set of letters in a consonant-vowel-consonant pattern. At a basic level, they are now decoding, or breaking the code. This aspect of reading instruction, in which the sounds of the language (phonemes) are associated with written symbols (graphemes, or orthography) is known as “phonics” (Carreker, 2005; NRP, 2000). In English, children typically begin with consonant-vowel-consonant words (real or nonsense) such as C-A-T. As an interesting and potentially relevant side note, Spanish
phonics is traditionally taught in syllables ("ma", "ti") which children then put together like building blocks to form longer words. This difference is not just a difference of teaching strategy; Spanish spelling lends itself to using syllables as building blocks rather than the consonant blends and other more complex grapheme chunks that form the building blocks of written English.

English is a large language with roots in many other language systems. English per se derives from Anglo-Saxon linguistic roots, one of the branches of Germanic language pedigree. However, Modern English also contains thousands of words and parts of words borrowed from other modern language roots, mainly from Greek and Latin (Henry, 2005). Each language system contributed its own pattern of orthography, making modern English a comparatively complex and difficult language to learn to read. Although English is fairly regular (there is a “code”, or list of phoneme-grapheme relationships, that apply with few exceptions), English contains many more phonemes and graphemes and different relationships among them than many other modern languages, particularly Latin-root languages such as Spanish (Cardenas-Hagan, 2005). Additionally, in English, single graphemes correspond to many possible phonemes and can often only be “decoded” (pronounced accurately), in the context of the particular word. For example, the grapheme “-ough” can be pronounced several different ways depending on the word in which it is found. Thus, decoding accurately in English requires knowledge of many more grapheme-phoneme relationships than required in Spanish and other Latin languages, as well as recognition of many graphemes and whole
words aided by lexical context, rather than phoneme-by-phoneme blending of sound correspondences (Henry, 2005).

Because of this issue specific to English, children in kindergarten and first grade are also asked to learn, by a combination of phonics cues and rote visual memorization, a long list of “sight words”. These are words considered to be exceptions to the “code”, or the letter-sound correspondences and more complex phoneme-grapheme relationships the children are being taught. They are not totally unrelated to expected letter-sound correspondences; however, they vary enough to make it difficult-to-impossible to sound them out armed only with phonics knowledge. Unfortunately, many of the shortest and most common words in the language are “sight words”. They include to, was, the, and as.

The task of learning, and teaching, phonics goes far beyond the beginning, the C-V-C words, particularly in English. Typically in the first grade, teachers must introduce a large number of consonant digraphs, consonant blends, vowel digraphs, and then more complex grapheme structures, in addition to “sight words”, and help students put these building blocks together until students are able to “decode” novel words independently. Because English, unlike Spanish, is not completely phonetically transparent (some familiarity with the particular word is often required in order to be able to pronounce it correctly), students' ability to decode novel words is dependent on some combination of knowledge of phonics and their memory bank of sight words. This ability to correctly identify a novel word is commonly referred to as “word attack (Carreker, 2005).” Early
on, Perfetti (1985) commented that children can only comprehend what they read to the extent that they become very skilled at decoding words.

Once young children begin to be able to match sounds with their printed symbols, letters and combinations of letters, and to decode (sound out) these symbols to read whole words, the next skill is to be able to read real words that occur in connected text in an automatic, or fluent, manner (Shaywitz, 2003; Ehri, 1993; Torgeson and Hudson, 2006). Fluency is related to the automaticity with which a person can recognize more and more, and bigger and bigger words (Ehri, 2002). It has three components: accuracy, rate and prosody (Torgeson & Hudson, 2006).

Neuroscientists using functional-MRI technology discovered that good readers, those who could read fluently and automatically, used a different neural pathway to decode than those who read in a halting, labored fashion (Shaywitz, 2003). This discovery has been termed the “dual-route” or “dual-pathways” theory of word-reading. Phonological coding parts of the brain were implicated in initial word identification in all children. But after being exposed to the same words or orthographic patterns, normally-developing readers gained automaticity and recognized the words and/or orthographic chunks of words using areas in the back of the brain, no longer implicating phonological processing areas (Shaywitz, 2003). Poor, or dyslexic, readers did not gain this automaticity and continued to use phonological areas of the brain to identify words.

Developing the ability to decode fluently is a major reading goal for readers from first grade until reading rate flattens out in late elementary school. Typically children are
practicing decoding and gradually improving their automaticity and reading fluency throughout early elementary school. Normally-developing readers develop automaticity and improve their fluency through simple practice reading; the more times a word is viewed, the more likely it is to be recognized automatically rather than decoded phoneme-by-phoneme.

Some degree of fluency is required in order for reading comprehension to be cognitively possible. Dysfluent reading diverts attention from the meaning of the text as the reader is utilizing attentional resources to decode the word using phonological processes (Carreker, 2005). Only when word-level processes are mastered and words are recognized automatically, thus a certain level of fluency is reached, can children be expected to comprehend what they are reading (Perfetti, 1985; Shaywitz, 2003; Marzola, 2005). Yet around third grade, the expectation of our schooling culture is that children will move from “learning to read” to “reading to learn”. In order to acquire expected content knowledge throughout the schooling process, it is important that children become sufficiently fluent readers to keep up with reading demands and fully comprehend what they read by this point in middle elementary school (Torgeson, 2005; Torgeson & Hudson, 2006; Proctor et al., 2005). Early reading struggles are highly predictive of academic difficulty in later years; dyslexic readers who do not acquire fluency but continue to read in a slower, more labored fashion have difficulty meeting academic demands in middle school, high school and beyond (Shaywitz, 2003).

Vocabulary instruction is perhaps is not entirely intuitive as a component of
explicit reading instruction, even though it is obvious that it would have a direct and strong bearing on the ultimate ability to comprehend what one reads. An individual's knowledge of word meanings, or vocabulary, develops from birth and continues throughout adulthood, given new linguistic input. Children learn to understand and speak their native language naturally through interactions with caregivers from the moment of birth (Vygotsky, 1962; Carlisle & Katz, 2005). Beginning with the lexicon of basic functional communication, children absorb word meanings rapidly and understand and use progressively finer shades of meaning for known words. Vocabulary knowledge is related to ultimate reading ability achievement, and is also highly predictive of overall academic achievement (Phythian-Sence & Wagner, 2007). Socioeconomic status and parents' education level, as well as other variables, predict a child's vocabulary knowledge (Hart & Risley, 1995). Vocabulary knowledge also aids in reading fluency because familiarity with words in spoken language helps children to recognize known words while reading (Torgeson & Hudson, 2006). Particularly in English, familiarity with each word based on exposure in spoken language prior to seeing the word in text facilitates correct recognition and pronunciation of the written word, since English phonics is not entirely transparent.

Vocabulary has been often neglected as part of reading curricula for at least two reasons. First, children learn most vocabulary words naturally and implicitly without direct vocabulary instruction as a result of communicating with other speakers, so direct vocabulary instruction has been sometimes considered unnecessary as part of literacy
curriculum, at least until fluency has been established. Secondly, teachers often believe that there are far too many words in the language to be explicitly taught, and any attempt would make such a small impact on overall lexical knowledge that it would be a waste of time (Beck et. al, 2002). However, researchers in vocabulary instruction have shown otherwise. Isabel Beck and colleagues, Catherine Snow and colleagues, among many others, have shown that explicit instruction in vocabulary— including targeting “tier II” words specifically (Beck et. al, 2002), morphological instruction for vocabulary-building, and various strategies to apply to learning novel words while reading, improve overall vocabulary knowledge significantly and reading comprehension ability as well (Snow et. al, 1991; Carlo et. al, 2005; Beck et. al, 2002).

Finally, children put all of these parts together and read. However, the undisputed goal of reading is reading comprehension— the ability to gain meaning from written text. While it is unusual for children who can decode fluently to comprehend nothing of what they read, children’s degree of comprehension varies tremendously. As texts grow in complexity, comprehension begins to depend less on decoding ability and fluency, and increasingly on the individual's linguistic knowledge, verbal intelligence, and background knowledge about the topic and the genre. Reading comprehension is not a discreet single ability that children can or cannot do; rather, it is a continuum variable ranging from the ability to decode and understand a single word to understanding complex specialized text within a difficult literary genre (Marzola, 2005). According to the phonological definition of dyslexia (Shaywitz, 2003; Lyon et. al., 2003), dyslexic
readers have difficulty with comprehension because their cognitive resources are still being used for word identification; they are not recognizing words with enough automaticity to free up attention for comprehension. Other children sometimes defined as dyslexic may have more pronounced language-based deficits affecting their ability to comprehend language more broadly, not limited to written language (Marzola, 2005; Aaron et. al., 2008). Additionally, many young developing readers who are not dyslexic are overly focused on decoding and fluency, or identifying all of the words correctly and rapidly, missing the true purpose of reading and failing to monitor their own comprehension (Marzola, 2005).

In recent decades reading comprehension per se has become viewed as an important aspect of reading instruction. Researchers examined differences between “good” and “poor” readers, finding that readers with good comprehension engaged in metacognitive strategies as they read to monitor their comprehension. Research has supported the efficacy of teaching students self-monitoring strategies as a component of reading instruction (NRP, 2000; Marzola, 2005).

Chall and colleagues published their Developmental Model of Reading in 1990, and its basic principles continue to hold up to scrutiny (Chall, Jacobs, & Baldwin, 1990). They described stage 0 comprising pre-reading literacy skills such as phonemic awareness and letter-sound correspondence knowledge. Then, stages one and two are decoding stages in which children are learning to read, while stages three and four are vocabulary and linguistic development stages in which children are “reading to learn”.
Finally, stage five involves acquisition of more advanced linguistic ability applied to reading comprehension, in which the student is able to use his or her own knowledge to expand and analyze what he or she reads.

1.4. Language Development and Language Proficiency

Cummins (2000) argues that misunderstandings about the way that language develops and the nature of language proficiency lie at the heart of confusing and contradictory policies affecting language minority students. For example, whenever an individual is considered to be "proficient" in a new language, the reality is probably dangerously over-simplified. Language is extraordinarily complex and multi-faceted. Both sociolinguistic and psycholinguistic research is relevant to conceptions of “normal” language development in children and the effects of particular educational policies and environments on children's linguistic, cognitive and academic development.

Sociolinguistics is the study of the way that social context affects language, while psycholinguistic studies focus on the cognitive processes involved in language learning. One important focus of sociolinguistics is the way that a single language changes among different groups of people. A dialect of a language, for example, varies from place to place, while varieties differ by social variables such as ethnicity, religion, or socio-economic status. A linguistic register varies based on the social context of the language event, for example the different language used to speak during a job interview versus with a close friend. While dialects and varieties of a language vary by individual, a single language user can usually control many different registers (Solano-Flores, 2006).
Cummins' (1981) hypothesized two main aspects of language proficiency: basic interpersonal communication skills (BICS) and cognitive academic language proficiency (CALP). In his model, BICS represents the language of social interaction that practically all human beings learn naturally within their linguistic communities. Young children are famously able to acquire communicative-level proficiency in both their native and even several new languages in which they are immersed at a very early age. Children acquire basic interpersonal communication skills when they begin to understand and use oral language to communicate about everyday topics. Children learn cognitive-academic language, on the other hand, very gradually throughout their schooling careers. Its lexicon consists of more sophisticated vocabulary words that make up the academic registers of a language and are found much less frequently in oral language than in written (Beck, McKeown, & Kucan, 2002.)

Russian psychologist Lev Vygotsky (1962), most well-known for his theory of language development through social interaction, also used a dichotomous model for distinguishing aspects of linguistic proficiency. Vygotsky’s “spontaneous” concepts are those words and aspects of language that develop naturally through use in the home for communicative and functional purposes, while “scientific” concepts and words are usually learned in school (or could be taught explicitly by caregivers). Other theorists have made a similar distinction, including “theoretical” vs. “concrete” and “embedded” versus “disembedded” language (Cummins, 2000). The distinction is broadly based on the degree of linguistic autonomy of the language item, or its degree of
“contextualization” (Snow, 1991). Communicative language is highly contextualized since much of its meaning can be derived from the context of the communication, and topics are concrete. The "academic" register of a language is characterized by language that can be understood autonomously from context, is more abstract, and is often more planned (Cummins, 2000; Snow, 1991). "Academic language" as a construct is currently under development by education researchers. Similar conceptions include the "language of schooling" (Schleppegrell, 2001) and "academic English language (AEL)" (Bailey, 2007; Bailey & Butler, 2003; Snow & Uccelli, 2009). Academic language proficiency is a strong predictor of long-term academic achievement (Snow, 1991; August et al., 2006; Carlo, 2007); in fact, so is vocabulary knowledge alone (Beck et al., 2002).

One of the points of controversy between formal and functional linguistic theories relates to the degree of independence that language development is granted from cognitive development. Functional and interactional theories of language development assume that language development converges with the rest of developing cognition (Bialystok, 2001). Vygotsky (1962) is credited with the first psychological explanation of how language and cognition are interdependent in their development. He wrote that “the structure of speech is not simply the mirror image of the structure of thought. It cannot therefore be placed on thought like clothes off a rack. Speech does not merely serve as the expression of developed thought. Thought is restructured as it is transformed into speech. It is not expressed but completed in the word. Therefore, precisely because of the contrasting directions of the movement, the development of the internal and external
aspects of speech form a true unity” (p.251). This interaction between language
development and cognitive development becomes quite relevant when we think of
children learning all of the things that they learn in school in a new language that they are
just acquiring.

Children exposed to two different languages from birth acquire both languages
without any loss of meaning or ability in either, and follow a similar language
development pattern as a monolingual child acquiring one language (Bialystok &
Cummins, 1991). Research on the vocabulary of these "balanced" bilinguals suggests that
while their vocabulary size is more limited in each language compared with
monolinguals, their word knowledge is spread across two languages and is comparable to
that of monolinguals, when their total conceptual vocabulary is considered (Snow & Kim,
2007).

Psycholinguistic researchers have documented certain cognitive-linguistic
benefits that seem to characterize nearly-balanced bilinguals. Primarily, bilingual
children and adults appear to enjoy greater "metalinguistic awareness," or awareness of
the differences and similarities between two or more language systems. According to this
line of research, bilingual children are better able to understand language as a system of
signs and symbols representing meaningful things and ideas, and that this system can be
manipulated to achieve communicative ends. These studies report that even very young
bilingual children exhibit superior performance on certain comparative linguistic tasks
requiring them to replace linguistic labels with new nonsensical ones, for example, while
holding the referent common. Monolingual children were unable to separate the word from its referent; they required more development than bilingual children to begin to understand language to be a system for referring separate from the object of reference itself. Early and/or superior metalinguistic awareness has also been associated with superiority on academic tasks, most of which are highly language-dependent (Bialystok, 2001).

Although balanced bilingualism seems to be a theoretical ideal, two languages are never truly balanced within an individual. According to Bialystok (1991), bilinguals do not simply replicate their abilities across languages. Rather, children learn the language items that they need to perform life's various communicative functions in the contexts in which they are carried out (Fillmore, 2001). Children who speak Spanish at home will know more "home" words, phrases and concepts in Spanish, while they may be much more articulate in English when the topic is related to a concept learned at school.

When a new language is added after the first has begun to develop, new linguistic input is filtered through a linguistic system which has already developed, making second language learning qualitatively different from that of the first language (Bialystok, 2001). For an adult or older child who is fully proficient in his or her first language, learning a new language mostly involves learning new words for known concepts that were assimilated, developed, mapped and labeled during childhood. In other words, the adult language-learning process is usually more intentional and analytical (Krashen, 1981). It is likely that a young language minority student suddenly immersed in a new language upon
arrival to school uses some combination of these two ways of learning language (Filmore, 1991).

Evidence demonstrates that a high level of first language (L1) proficiency assists in second language (L2) acquisition, and conversely, high proficiency in L2 has positive effect on L1 development (Cummins, 1979; BournotTrites & Tallowitz, 2006; Duursma et al., 2007). Many aspects of linguistic ability transfer between languages. These include both "bottom-up" underlying cognitive-linguistic abilities such as phonological ability (Durgunaglu, Nagy & Hancin-Bhatt, 1993) and rapid automatic naming (Geva & Zadeh, 2006) as well as "top-down" abilities such as vocabulary knowledge (Proctor, August, Carlo, & Snow, 2006) and reading comprehension (Cummins, 1979).

According to Cummins (1979), the degree of transferability of skill depends on how much cognitive-academic language development has taken place in the first language. This would explain why students who arrive later with a few years of first language instruction are able to become more proficient in English and perform better academically over the long-term, even though their communicative proficiency may not develop as quickly as a younger child's (Cummins, 2000; Hakuta, Butler & Witt, 2000; Proctor et al., 2006).

Studies of cross-language transfer of vocabulary breadth have found low to moderate correlations between vocabulary size in bilinguals' two languages. Many have generated contradictory results (Snow & Kim, 2007), suggesting perhaps a mediating variable. However, most of these were cross-sectional— looking at cross-linguistic
relationships at the same age. Longitudinal studies suggest a more consistent relationship between L1 and L2. For example, early Spanish vocabulary positively predicted later English vocabulary (Proctor et al., 2006).

English language learners typically demonstrate not only limited breadth of English vocabulary knowledge, but also lack comparable depth in their word knowledge. Words carry subtle shades of meaning, alternative meanings, and different usages; in fact, the ability to use a word appropriately depends greatly on the context of use (Miller, 1999; Ordonez, et al., 2002; Verhallen & Schoonen, 1993). Some empirical educational studies have incorporated this concept of word knowledge depth as a variable, attempting to measure it in various ways such as counting aspects of meaning reported (Ordonez, Carlo, Snow, & McLaughlin, 2002; Verhallen & Schoonen, 1993). Ordonez and colleagues (2002) assessed two aspects of depth of word knowledge by asking children to provide both paradigmatic and syntagmatic information about a word. Paradigmatic information identifies the word's place in a taxonomic hierarchy, while syntagmatic information describes the object's appearance, location, or use. For example, if one were defining a “dog”, syntagmatic information may include that a dog is furry, has four legs, is domesticated, and is a popular pet. Paradigmatic information would explain that a dog is a type of animal, a mammal, and one of many domesticated animals. This latter information would clear up any confusion between the word “dog” and the word “animal” by placing the dog under the category of “animal”, inferring the existence of other furry four-legged animals which are not “dogs”. A formal definition of a word
contains both, with just the right amount of syntagmatic information to target those features that make the word unique, and no more. Superordinate knowledge is less commonly provided or understood by young children and is part of a developmental trajectory, surely influenced by literacy-based instruction, of the development of verbal intelligence, assessed also within measures of intellectual ability. Also, paradigmatic knowledge is one of many proxies signaling a much higher level of proficiency in a particular language than the ability to name a few qualities about the target word (syntagmatic information). This type of deeper linguistic knowledge is likely an aspect of what Cummins’ named CALP. It is linguistic knowledge, yet providing formal definitions is also a skill that easily transfers between languages in the sense that formal definitions work the same way across the languages. Thus, Ordonez and colleagues (2002) found that the ability to provide superordinates, or paradigmatic information, in Spanish predicted that ability in English after controlling for breadth of vocabulary knowledge. Simpler definitions, on the other hand, were more language-specific. In other words, Spanish-speaking children who had not reached this higher level of understanding about their first language were less able to apply it to understand English words and how they fit into a conceptual hierarchy.

Although later criticized for oversimplification (Snow & Ucelli, 2009), the basic distinction between at least two general aspects of language proficiency is one of the few insights from bilingual education and psycholinguistic research that has, albeit only recently, managed to influence policies affecting language minority students. English
basic communication skills alone have little predictive value regarding children's ability to learn academic subjects in English. This has been a notorious problem for the classification of English learners—teachers assume that children who appear to be fluent in conversational English are linguistically equivalent to native speakers, or have “finished” with the language issue and are prepared for regular academic demands. This continues to be the assumption driving policy. In reality, language proficiency measures that target CALP skills, in either language, are far more predictive of when children are capable of surviving academically in an English-only program (Cummins, 2001). In response to the criticism, language assessments required by No Child Left Behind for English Language Learners now usually include CALP proficiency. Unfortunately, the result of raising the proficiency thresholds required for re-designation is that students simply take a longer time to reach the new threshold, remaining classified as English language learners often for many years (Kim & Herman, 2009). If we require some degree of cognitive-academic language proficiency to mainstream second language learners, we cannot then rationally expect a student that enters kindergarten with little English ability to reach exit criteria before grades five or six (Hakuta, Butler & Witt, 2000; Thomas & Collier, 1997).

1.5. Language Use, Proficiency and Policy

Educational policy documents demonstrate little if any awareness of the empirical research on language acquisition and proficiency. For example, No Child Left Behind states that language minority children should learn English "as quickly as possible" (No
language is an all-or-nothing achievement, and one that takes only a matter of months.
Students' incomplete mastery of English is conceived as a deficit that must be remediated
quickly so that the child can move on to learning content.

In reality, learning to use the specialized aspects of the English language within
each academic discipline is a life-long venture for all students. Moreover, academic
"content" and the language with which it is built and taught cannot even theoretically be
disentangled (Vygotsky, 1962). However, NCLB's accountability system ignores the
extensive evidence showing that it takes many years for people to become truly proficient
in a first or a second language (Hakuta et. al., 2000; Krashen, 1991), that language
learning is not a linear process but varies across different language situations and
functions (Filmore, 2001), that proficiencies vary by type as well as by degree (Cummins,
2000), and that the length of time needed to learn various aspects of English depends on a
host of individual and contextual factors (Abedi, 2004; Hakuta, Goto, Butler & Witt,
2000).

The “English-only” voter propositions present an even less accurate portrayal of
language proficiency. The promise of the proposition legislation was that children in
English immersion settings would become proficient in English within one year, ready to
be mainstreamed and to handle academic tasks with no further specialized assistance
(Massachusetts General Laws c. 71A, as amended by Chapter 386 of the Acts of 2002).
To quote the Massachusetts proposition (2002): "Immigrant children can easily acquire
full fluency and literacy in a new language, such as English, if they are taught that language in the classroom as soon as they enter school.... Therefore it is resolved that: all children in Massachusetts public schools shall be taught English as rapidly and effectively as possible.... Children who are English learners shall be educated through sheltered English immersion during a temporary transition period not normally intended to exceed one school year.... Once English learners acquire a good working knowledge of English and are able to do regular schoolwork in English, they shall no longer be classified as English learners” (General Laws c. 71A, as amended by Chapter 386 of the Acts of 2002, Massachusetts).

A dichotomous decision regarding a students' English proficiency can be very misleading. While it may take only a few years for English language learners to acquire native-like communicative proficiency on everyday topics in English, it takes at least 4-7 years to perform on par with native-speakers in terms of advanced reading comprehension and academic vocabulary and proficiency (Hakuta et. al, 1999; Thomas & Collier, 1987). Gandara (1999), for example, collected data in several California schools demonstrating a large discrepancy between language minority students’ oral abilities in English and their knowledge of academic English vocabulary and ability to comprehend written English. Practitioners are presented with the dilemma of complying with the new law and providing English learners the extra academic support they still need, since after one year of English immersion students are not typically able to do "regular school work in English".

32
English-only advocates point to the positive outcomes reported for language immersion programs in which native English-speaking students are immersed and taught almost entirely through a second language in elementary school (e.g., Rossell & Baker, 1996), such as the widely-publicized results reported by French immersion schools in Montreal, Quebec, credited for popularizing foreign language immersion programs for young native English-speaking children. English-speaking students educated in French became proficient and literate in a foreign language and ended up with even better first language literacy skills than peers (Swain & Lapkin, 1991). However, the relevance of these examples to the education of Spanish-speaking Latino youth in U.S. schools appears minimal (Cummins, 1981, 2000).

The reason that some bilinguals show apparently negative effects while others demonstrate heightened metalinguistic awareness and other special abilities (Diaz & Klinger, 1991) hinges on the different linguistic and sociocultural contexts in which students live and study (Cummins, 1979, 2000). Paradoxically, early schooling conducted in a second language (i.e., language immersion) seems to produce vastly different effects for students who speak a minority language and are immersed into the majority language, compared with students who speak the majority language and are immersed into a minority language (Bialystok & Cummins, 1991; Cummins, 1981; Lambert, 1981). The distinction primarily has to do with the statuses of the languages in the society, as well as the economic and educational level of the community of speakers (Cummins, 1991). When a language is dominant and prestigious in society, there is little danger for its loss
in individuals (Francis, 2005; Lambert, 1981; Fillmore, 2000). "Additive" bilingual environments are those in which students are meant to learn a new language while maintaining and continuing to develop their first language. In an additive bilingual environment such as the French immersion programs, there continues to be sufficient support for native language development in the larger society, at home, and even usually at school. This leaves children with a comparatively balanced bilingualism, the type associated with positive cognitive-linguistic outcomes such as heightened meta-linguistic awareness (Bialystok & Cummins, 1991; Diaz & Klinger, 1991), not to mention the social benefits of bilingual proficiency.

However, in the case of language minority students who receive little or no academic support in their first language, the effects of total immersion in the dominant language appear to be subtractive rather than additive (Cummins, 2000; Fillmore, 1991; Garza & Crawford, 2005; Lambert, 1981). A "subtractive" bilingual educational environment has the explicit goal to replace the students' home languages with English (Lambert, 1981; Valenzuela, 1999). These include English-only environments, as well as short-term bilingual programs that only use Spanish to transition students to English quickly.

Ruiz (1984), in his discussion of language planning for institutions and nations, described three "language ideologies" regarding minority heritage languages: "language-as-problem"; "language-as-resource"; and "language-as-right". The United States, compared to most other countries, implicitly adopts the first ideology. Tse (2001) has
pointed out the disparity between the ways that the heritage languages of minority students are treated in policy, versus the proliferation of foreign language programs for English-speaking students. Accordingly, English-only propositions leave loopholes to allow established dual language programs to continue; these tend to serve native English speakers whose parents would like them to receive foreign language immersion at a young age. Even though they also enjoy the highest academic outcomes for language minority students, but unfortunately are few and far between in the current anti-bilingual milieu.

The stages of bilingualism in which language minority students find themselves during English-only schooling include long periods of time in which they demonstrate a lack of normative competence in either language (Filmore, 1991; Cummins, 2000; Droop & Verhooven, 2003). They occupy a between-languages place that is common in multilingual communities where different languages are used for different functions. In a subtractive schooling situation, English becomes the language of school for the individual child, and the language of sophistication and intellect, while Spanish (or perhaps a code-switching variety in which Spanish is mixed with English) is the language spoken at home on topics discussed there. One finding is that as Spanish-speaking children progress in an English-dominant educational setting, their relative language proficiency in English and Spanish move toward a period of more equal use in middle childhood, and then towards English dominance in late childhood and early adolescence (Hakuta & Diaz, 1985).
There is a strong correlation between a student’s literacy skills in their first language and the ease and success with which they learn to read and write in English (August et al., 2006; Carlo, 2007; Proctor, August, Carlo, & Snow, 2006; Shanahan & Beck, 2006). Students who arrive in the United States in mid-childhood or later, already able to read and write in their first language, may pick up conversational English more slowly, but are on average more successful in school in the long-term than earlier arrivals and native-born ELLs, even though early elementary ELLs more rapidly acquire oral fluency (Cummins, 2000; Valenzuela, 1999). However, the majority of young public school language minority students acquire literacy skills only in English while simultaneously acquiring oral English abilities, so they are unable to benefit from the transferability of many literacy skills (August et al., 2006). Additionally, realizing that older immigrant students experience better long-term educational prognoses on the average than native-born Spanish-speaking students further breaks down the persistent assumption of a direct relationship between English oral language ability and long-term academic achievement.

Cummins' (1979) "Threshold Hypothesis" theorized a minimum level of proficiency in a first language before the benefits of bilingualism emerge. This conversely suggests that children who have acquired only basic communicative proficiency in their first language are not likely to benefit from full immersion in a second language until some undefined threshold of cognitive-linguistic development has been reached in the first language (Cummins, 1979). The theory has been controversial
because of the implication that children could become stunted cognitively by particular linguistic environments. If students are not able to fully develop their home language when placed in a subtractive linguistic environment, and they are not able to truly "catch up" in English either, where does that leave them?

Without suggesting that groups of children either have no language or inferior language proficiency, which is not accurate, it is apparent that children in a subtractive linguistic context typically have a more difficult time learning the academic aspects of the English language that successful native English speakers and children in high-quality bilingual programs learn in school, and which in turn enable the development of advanced literacy skills. However, this effect is heavily mediated by socioeconomic class; language minority students are certainly not the only students who exhibit inadequate academic vocabulary knowledge. While comparative language proficiency levels are extremely complex, particularly in bilinguals, several studies have observed a sort of "low-low" proficiency phenomenon among minority language speakers educated in subtractive environments (Droop & Verhooven, 2003). This phenomenon, at times termed "semilingualism", has excited controversy and many writers have dismissed it as part of the deficit paradigm in which speakers of other languages are viewed as deficient in English (eg. MacSwan, 2000). Certainly it is not useful or necessary to add another "disability" category to the conversation and to treat the unique linguistic position of language minority in subtractive environments as linguistically deficient. In fact, the making of a "deficit" has as much to do with the normative context as with the child
himself; English-only environments themselves appear to create linguistic deficits in individuals. While it would be counter-productive to confuse such an environmentally-produced problem with an internal deficit, it is important to recognize that students educated in English-only environments begin schooling at a linguistic disadvantage and then are often denied the language-based education that successful English-speaking students may receive without necessarily recognizing it as such—e.g., training in academic English. In that sense, these students may leave elementary school with a handicap in particular aspects of language. Also, these students typically lose the ability to speak in their native language.

The phenomenon of language loss in immigrant communities over the course of generations is called language "shift". One point of contention within the research is the degree of inevitability of intergenerational language shift, and why some communities manage to maintain their heritage language in the face of various subtractive pressures while others do not. Early researchers such as Merino (1983) reported that Spanish-speaking elementary school children were more likely to lose their heritage language than to become truly bilingual. Other research, however, points more optimistically to a high degree of transmission of Spanish to second-generation speakers (Mora, Villa & Davila, 2006). Veltman (1993) and other demographers have presented evidence demonstrating a typical three-generation shift scenario.

Identification with one's ethnic or racial group, including use of a native language or community variety, is related to healthy identity formation, particularly for minority
adolescents. Psychological research on ethnic identity has found a significant positive relationship between self-esteem and ethnic identity among ethnic minority adolescents (Umana-Taylor, Diversi, & Fine, 2002). Identity development is a major theme in adolescence psychological research, and ethnic identity is an important aspect of psychosocial adjustment for many minority youth. According to Josselson's (1994) "embeddedness" conception of adolescent identity development, "individuals construct their identity in relation to others and, in particular, within groups” (as quoted in Umana-Taylor et al., 2002). Among the many studies of relationships between self-esteem and ethnic identity, Ethier and Deaux (1994) found that higher importance ratings of Hispanic identity were associated with higher ratings of self-esteem. Researchers studying immigrant psychology from an acculturation perspective view a bicultural identity as the healthiest level of ethnic identity, incorporating both personal flexibility and adaptation with strong identification with one's ethnic roots (Kouritzin, 2006).

Schools' language ideology could potentially affect many aspects of the schooling experience for Hispanic families. An indifferent or hostile attitude toward the language of students can make home-school connections very difficult to achieve (Cummins et. al., 2005). Subtractive bilingual environments have been linked to social and emotional problems among language minority students such as low self-esteem and lowered expectations for achievement (Cummins, 2000, Kouritin, 2006; Thomas & Collier, 1997).

Although most studies only consider English oral language and literacy abilities, a major consequence of not using the native language for any instruction is individual
children's loss of speaking ability in that language. The earlier children are immersed in an English-only school setting, the more severe the potential subtractive effect on their native language (Fillmore, 1991). Fillmore (2001) argues that the loss of a language in early childhood, the fate likely for children in English-only school programs, has negative effects on aspects of child development. The most obvious of these is communication problems between children and parents. Parents often cannot be involved in homework help or any part of their child's education. Children's home and school worlds become increasingly disconnected. She argues most strongly against English-only preschool programs because young children still acquiring their first language can quickly lose their emerging Spanish, jeopardizing communication with parents, which is vital for positive social-emotional development. As the major local community institution and where children learn most of their language, a school’s decision (or non-decision) regarding whether to adopt a form of bilingual education can play a large role in determining individuals' bilingual abilities and the fate of the minority language itself.

Kouritzin (2006) interviewed adult second-generation immigrants regarding their feelings about their heritage languages and the effects that they believed loss of these languages has had on their social-emotional development and ethnic identity. She found that "what appear to be the most salient are the themes which have to do with schooling and identity development at school. There is resentment, anger, and grief that school policies and programs did not encourage, did not even permit, the use of first heritage languages. There are strong, frequent, echoes of never really fitting in, of never being part
of a group. The failure to recognize languages other than English bred shame in the participants, and forced them to conceal their heritage identities at school, a practice that eventually became life-long. What is most painfully clear is that identity and language 'decisions' made by children in the face of an assimilation-oriented dominant culture, are decisions later regretted. These participants seem to ask themselves, ‘Why did I let this happen? Why did anyone let this happen?’ ...related to this idea, people who lose a first language as children understand later, when they are older, how much they have lost. As they age, familial connections become more important… (Kouritzin, 2006, p. 20)."

1.6. Measurement and Language Difference

1.6.1. Language proficiency testing for classification

English language proficiency standards set for mainstreaming language minority students once they are "proficient" are riddled with validity issues, since proficiency thresholds are arbitrary points along a vast continuum. States use a variety of different assessment strategies to determine when students are proficient in English and ready to stop receiving special language services. Many use only a standardized language proficiency assessment, others use a teacher rating scale, and still others include academic achievement assessments, or a combination of the above (Kim & Herman, 2009). This of course results in disparate definitions of an English language learner between states and even districts (Ragan & Lesaux, 2006).

As practitioners begin to understand the close relationship between higher-order cognitive-linguistic abilities and academic achievement, and conversely the lack of
relationship between conversational proficiency and academic success, there is more
demand to require language minority students to demonstrate both academic English
proficiency and normative academic achievement prior to being mainstreamed. Thus, as
mentioned previously, modern language proficiency assessments usually assess
cognitive-academic language ability as well as basic interpersonal communication skills.
However, the further the proficiency threshold moves into realm of academic language
proficiency, the longer students take to reach it. In Holyoke, for example, students often
require five years or more to move from learner status to proficiency (personal
communication with David Valade, district ELL director, 2010).

Rather than a dichotomous distinction between “proficient” and “learner”, most
school districts with a substantial ELL population use a scale with several steps between
a beginner in English and a learner who has (theoretically at least) reached a level
comparable to a native speaker of English. Holyoke uses a four-level rating system for
categorizing English learners: beginner, intermediate, advanced intermediate, and
transitioning.

1.6.2. High-stakes testing

Since No Child Left Behind (2001), it has become the norm to use a single score
on a standardized norm-referenced academic assessment to make high-stakes decisions at
district-wide, school-wide, and individual student levels. Nearly one-third of states now
require students to pass academic assessments to graduate from high school and/or to
advance into certain grades. NCLB also requires that states include special populations
such as English learners in achievement testing, even though the validity of English-
language achievement tests for use with students who learned English as a second
language has been repeatedly questioned (Abella, Urrutia & Shneyderman, 2003; Mahon,
2006; Solano- Flores, 2001).

Test validity refers to the degree to which the test score varies on the basis of the
intended construct, or measures what it intends to measure. For an English learner, a test
written in English intending to measure content knowledge is actually measuring the test-
taker’s mastery of the English language, in particular the language of the content area
within the tested genre. The Standards for Educational and Psychological Testing states
that "for all test takers, any test that employs language is, in part, a measure of their
language skills. This is of particular concern for test takers whose first language is not the
language of the test (p.91, Joint Committee of the AERA, APA & NCME, 1999)."
Numerous educational measurement studies have documented that language can act as a
"construct-irrelevant" (Messick, 1989) factor affecting children's scores on standardized
tests. For example, Solano-Flores (2001) found that a given student could perform better
in English on some items and in his native language on others. He used G-theory to
analyze the degree to which a student's dialect and language variety affected their
achievement test scores, and found that the largest variation in test score was due not to
actual academic skills, but attributable to the student-item-language interaction.

Abella and colleagues (2003) examined the scoring patterns thousands of
Spanish-speaking students in Miami schools in fourth and tenth grades and concluded
that “English language achievement tests are not valid measures of academic achievement when applied to the LEP student population. Additionally, achievement tests are often not valid measures of academic knowledge when applied to students who have recently been classified as English language proficient (Abella et al., p. 11).” These researchers took as evidence of invalidity cases in which a student responded correctly in Spanish but incorrectly in English on the same content, since theoretically that suggests that the student knew the content (as evidenced by successful expression in Spanish), but could not express his or her knowledge in the English language. However, this assumes a firm distinction exists theoretically between content and the language through which it is expressed. Since validity is defined as scoring equally or less well in Spanish, students found to be tested “validly” were also those children who could not express academic ideas in Spanish or English (or could in both); most presumably had either lost or never had the opportunity to develop academic language skills in Spanish. So even when achievement testing appears “valid” in the sense that the student could not be expected to do better in another language, it seems invalid in a more basic sense to expect comparable achievement in a second, new language.

Due to the large variability in students’ patterns of language dominance and proficiencies, including dialect variations and potentially the loss of their native language, ELL students are particularly susceptible to invalid testing, regardless of the language in which they are tested. In order to perform well on a standardized test of academic content, students must not only know the content itself, but be highly proficient
in the dialect of English known as Standard English. Furthermore, students must be proficient in the specific Academic English registers of the content areas being tested, as well as the register of "test language". It is not possible to separate the language of the content, nor the language of the test employed, from the content itself.

1.6.3. Reading disabilities and language difference

Practitioners are often tasked with making dichotomous decisions about children's abilities; children either do or do not "have" a reading disability, and they either do or do not "have" English proficiency, even though both abilities lie on continua. A major preoccupation of school psychologists working with Hispanic populations has been the quest to distinguish a reading disability from a reading problem related to language difference (Rhodes, Ochoa & Ortiz, 2005). Since reading skill and language proficiency both represent complex clusters of interdependent cognitive-linguistic abilities, this task is theoretically difficult if not nonsensical. However, there is legitimate concern that students could be diagnosed as learning disabled when in fact they are just less proficient in English, which itself causes difficulties learning to read in English and assuming inadequate or non-existent native language support in school, often results in lower reading achievement (Cummins, 2000; Thomas & Collier, 1997). If we wait for these students' English language proficiency to catch up to native-speaking peers, we could fail to recognize a reading problem that is not language-based and allow, for example, a phonological deficit to go untreated. Both of these trends have apparently taken hold. Practitioners are sometimes reluctant to test young children who speak little English upon
arrival in school for possible neurologically-based difficulties they may experience learning to read; Conversely, sometimes even older, re-classified students assumed to now be proficient in English are over-identified as having specific learning disabilities in reading since they fall behind peers in reading achievement (Rhodes, Ochoa, & Ortiz, 2005).

Using norm-referenced tests with English language learners is problematic when they do not represent a substantial portion of the norming sample, since language difference is so likely to bias achievement and cognitive scores (Rhodes, Ochoa & Ortiz, 2005). Scores from standardized tests must always be interpreted with caution, considering the target students' unique language profile and the close relationship between language learning and academic and cognitive performance. Ochoa and Ortiz (2005) have provided the Multidimensional Assessment Model for Bilingual Individuals (MAMBI), an empirical tool for examining a student’s linguistic history to choose the most fair mode of assessment—non-verbal, first-language, second-language, or bilingual, as well as the Culture-Language Interpretation Matrix (C-LIM) for determining the degree to which subtests are likely to be culturally and linguistically-loaded, so that biased scores can be avoided or interpreted carefully (Rhodes et al., 2005).

Another type of literacy assessment conducted in schools measures young children's abilities on discreet tasks thought to be precursors to achieving literacy-related skills and ultimately learning to read. As early as kindergarten, and before, children are tested on various tasks such as knowledge of letter names and sounds and the ability to
distinguish the individual sounds they hear in words. Students' progress in terms of these specific skills at face value is considered progress in the curriculum; the skills correspond to what is taught in the classroom as children are taught to read.

1.7. Assessment of Early Skills to Predict Reading Achievement

A test is only valid to the extent that it measures the intended theoretical construct, as opposed to inadvertently measuring other constructs. The extent to which a test measures what it intends to measure has been called its “construct validity” (Messick, 1989). One way to provide evidence of construct validity is to calculate the correlation between an individual's performance on the test and on other tests purporting to measure the same construct. All measures intended to measure a child's phonological awareness, for example, should correlate with each other; if they do not, our concept of a single ability called “phonological awareness” could be flawed and in fact be a group of different, independent skills.

The practical purpose of measuring children's early literacy skills is to be able to determine which children are in need of early intervention in order to improve their ultimate reading outcomes. Therefore, the utility of the measures lies in the extent to which they are able to predict how students will perform in the future in terms of a culturally-valued outcome, such as “reading comprehension”. Additionally, it is useful when measures which can be administered more cheaply, briefly, and/or formatively are able to predict student performance on high stakes assessments, whether close in time or into the future. A test's predictive validity refers to the degree to which it predicts the
same student's performance on another test, or on a related skill either close in time or
into the future (Messick, 1989; Betts et. al., 2008).

Comparative reading ability is relatively stable over the lifetime. Most
longitudinal predictive studies have demonstrated that early reading ability is the best
predictor of later reading ability (Scarborough, 1998). Reported correlations in large
samples between early reading ability and later reading achievement, even up to six years
apart, are typically fairly high (r = .63 to .86; Scarborough, 1998; Butler et. al., 1995;
Shaywitz et. al., 1992). Stanovich (1986) suggested that reading ability trajectories
mirror the biblical “Matthew Effect”, in which the rich get richer and the poor get poorer.
He pointed out that children who struggle with reading in early years read less, getting
less practice and compounding their comparative weakness, while children who read well
read more and improve more quickly as a result. Many early correlational studies
supported his hypothesis (Scarborough, 2005; Juel, 1988; McKinney & Feagans, 1984).
However, longitudinal studies with more distal outcome variables, predicting reading
outcomes in older students, have produced more mixed results and not always supported
the Matthew Effect hypothesis; some have even demonstrated a narrowing of ability
range among older students (Scarborough, 2005; eg. Baker, Decker & DeFries, 1984).
The more longitudinal research that comes out, with increasingly distal outcome variables
such as predicting reading competence at a high school level and beyond, the more
complex the picture of reading acquisition becomes.

Longitudinal and concurrent predictive research has identified many correlates of
students' academic achievements, ranging from status variables such as home environment, race, class, and parent's educational level, to instructional variables, to emotional, psychological, and cognitive variables (Betts et. al., 2008). Although innumerable variables are correlated with reading ability, no single variable can fully explain it.

Over the past several decades, due apparently to increasing pressure to improve reading outcomes among students as well as accumulation of measurement validity evidence, increased attention has been paid to developing and assessing prereading, or foundational, literacy skills in young children (Fuchs et. al., 2001; Deno, 1998; Betts et. al., 2008; NELP, 2008) Research has demonstrated that screening young students in literacy abilities is associated with improved outcomes because it facilitates early intervention to prevent problems acquiring reading ability and informs direct instruction in foundational reading skills (Deno, 1998; Fuchs et. al., 2001; Torgeson et. al., 1999; NELP, 2008). Many empirical studies have emerged to validate early screening measures by assessing their ability to predict students' later reading ability (Scarborough, 1998; Torgeson et. al., 1999; Wagner et. al., 1997; NELP, 2008). For very young children not yet expected to decode whole words, two distinct skill areas have been repeatedly identified as the strongest predictors of later reading success: phonological ability and letter-sound correspondence knowledge, also called “alphabetic principle” (Betts et. al., 2005; Wagner & Torgeson, 1987). Early on, Wagner and Torgeson (1987) published results finding phonological ability in young children to predict early reading
ability even better than word-reading skill did.

More recent reviews of results of multiple studies, however, such as Scarborough's 2005 review, have demonstrated that many other cognitive and linguistic skills that can be measured in young children are as strong or nearly as strong predictors of later reading ability (Scarborough, 2005; NELP, 2008). According to Scarborough's meta-analysis of findings from 61 predictive studies using kindergarten measures, the most powerful kindergarten predictor of future reading ability was Letter Identification (M of r = .52). The next-strongest predictor was Concepts of Print (M of r = .46), then Phonological Awareness (M of r = .46), Expressive Vocabulary (M of r = .45), Sentence/Story Recall (M of r = .45), Verbal IQ (M of r = .41). The list of similarly strong correlates goes on to include many other cognitive and language skills (Scarborough, 2005).

In 2008, an extensive meta-analysis of the correlational research on early literacy skills was released by the National Early Literacy Panel (NELP, 2008). The Panel reviewed hundreds of correlational studies. It found 69 studies examining the predictive validity of measures of “Phonological Awareness” and reported an average correlation of r = .40 with “later” reading achievement across the studies. While this correlation is moderately strong, other tasks proved much more predictive, including “Decoding Nonwords”, (average r = .72) and IQ (average r = .45), among others (NELP, 2008).

Although other abilities correlate similarly or more strongly with later reading ability, phonological awareness has been afforded a great deal of attention for several
other reasons: unlike other correlates, along with basic letter-sound correspondence (or “phonics”) ability, a minimum level of phonological awareness has been considered a necessary pre-condition required for the acquisition of early word-level reading ability (Snow et. al., 1998; Betts et. al., 2005). Also, it is amenable to instruction and intervention (Uhry, 2005; Torgeson et. al., 1999), can be measured very early, and has been implicated as a neurological cause of specific reading disability (Shaywitz, 2003; Torgeson et. al., 1999; Wolf & Bowers, 1999). Indeed, the definition of dyslexia proposed by Shaywitz and colleagues implicates a phonological deficit as the causal variable (Wagner & Torgeson, 1987; Lyon et. al., 2003; Shaywitz, 2003).

Phonological processing generally is the use of phonological information-- the sound of one's language-- in processing oral and written information (Cutting & Denckla, 2001). Phonological deficits have been named as the primary cause of reading disabilities (Torgeson, 1998; Shaywitz, 2003). Phonological awareness, phonological memory, and rapid serial naming are all basic phonological processes related to and predictive of the acquisition of decoding skills (Stanovich & Stanovich, 1995; Torgeson & Wagner, 1995; Wagner, 1999). Phonological awareness alone is a powerful predictor of decoding ability (Torgeson & Wagner, 1995; Wagner, 1999; Wagner & Torgesen, 1987). Phonemic awareness is an aspect of phonological awareness in which the subject displays awareness of the phonemes-- the smallest single units of sound-- which make up words. This type of phonological awareness task has demonstrated the strongest relationship with acquiring early phonics ability (Burke et. al., 2009; Snow et. al., 1998;

“Rapid serial naming” (RSN) is a distinct phonological cognitive task which has demonstrated predictive power in terms of predicting early reading skills. On this type of task, the child names familiar objects, colors, or numbers, and rate is recorded. The construct is sometimes called “rapid automatic naming” (RAN) or “lexical access”.

Modeling research has demonstrated that phonological awareness and RSN are separate abilities which each contribute independently to decoding skill (McBride-Chang & Manis, 1996). RSN alone has also been found to predict of word-reading skill, even when phonological awareness and IQ have been partialed out (Ackerman & Dykman, 1993; McBride-Chang & Manis, 1996; Scanlon & Veluntino, 1996). The size of its relationship to word reading appears to vary across orthographies (different written languages); in more orthographically regular languages such as German, Dutch and Spanish, RSN is more predictive of word reading ability than phonological awareness is (Lindsey et. Al., 2003).

Bowers, Wolf and colleagues (1999) proposed a double-deficit hypothesis of reading disability in which impaired readers have deficits in phonological awareness, rapid serial naming, or both (Wolf & Bowers, 1999). According to Wagner & Torgeson (1987) and as expressed in their CTOPP test, these are both aspects of phonological processing, along with phonological short-term memory, which the CTOPP also tests. However, Wolf and colleagues argue that RSN is additionally measuring non-phonological factors such as general-processing speed, which contribute independently of
other aspects of phonological skill (Wolf & Bowers, 1999).

Many models have put the double-deficit hypothesis to empirical test. McBride-Chang and Manis (1996) demonstrated that phonological awareness and rapid automatic naming contributed independently to word-reading ability in weak readers. Burke et al. (2009) examined the predictive validity of kindergarten measures predicting decoding ability. Their results support the importance of phonemic awareness, phonemic coding, and automaticity (rapid automatic naming).

Researchers have used multivariate techniques including modeling and multiple regression to understand the inter-relationships among all phonological tasks. Baddeley (1986) proposed a theory of the phonological loop, in which “rehearsal rate” (same as RSN) is directly related to working memory, or the ability to retain phonological information in immediate awareness, as well as articulation ability. Researchers have concluded that learning disabled children thus have difficulty “coding the phonological features of language (Torgeson et al., 1990).” Path analysis research on memory span has suggested that processing speed (a broader construct to include visual as well as phonological information without articulation) indirectly predicts memory span through articulation rate (Kail & Park, 1994). Ackerman et al. (1990) found that articulation is highly correlated ($r = -0.71$) with phonological awareness measures. Memory span could be related to other measures only as a limiting factor, however; in other words, a child’s memory span puts a limit on their performances on any task requiring short-term memory. Performance on phonological awareness tasks, for example, is limited by the
number of speech sounds that the child is able to hold in awareness.

Although heavily dependent on instruction and thus not a strictly neurological correlate, letter-sound correspondence, or basic phonics ability, has stood out as the most valuable predictor of word-reading ability and basic reading ability in young children. Large meta-analyses have repeatedly implicated direct and robust phonics instruction as the most impactful aspect of early reading instruction (NRP, 2000; NELP, 2008). And, children's individual differences in basic phonics ability have repeatedly shown to be the strongest way to predict their later basic reading ability (NRP, 2000; NELP, 2008). In fact, knowing the names of the letters in kindergarten has been repeatedly demonstrated to be the single most predictive factor in determining a child's reading health (Scarborough, 2005).

Cutting and Denckla (2001) proposed a triple-deficit hypothesis adding orthographic knowledge, or “phonics” knowledge to the equation predicting word reading. They found that RSN, phonological awareness, and orthographic knowledge all had direct effects on word reading ability. Their model also added further evidence that RSN and phonological awareness are relatively independent abilities which contribute separately to reading ability. They also found that RSN had no direct effect on memory span, and that memory span was related to word reading independently but not in the model; in other words its contribution was eclipsed by that of the other measures included. Post-hoc the authors ran a regression equation removing phonological awareness and found that memory span was again a significant contributor, meaning that
the variances of phonological awareness and memory span overlap. Indeed, their correlation was significant \((r = 0.36)\) and corroborated by earlier findings (e.g., Wagner et al. 1997) suggesting that memory span is related to performance on phonological awareness tasks.

The Dynamic Indicators of Basic Emerging Literacy Skills (DIBELS; Kaminsky & Good, 1996), one brand of curriculum-based measures of early literacy skills, are the most well-researched early literacy screening measures. This group of tasks measures phonological awareness, letter knowledge, and letter-sound correspondence knowledge in children as young as kindergarten. The DIBELS measures are not intended to constitute a comprehensive assessment of a child’s abilities; rather they serve as screening measures, in the way that blood pressure is used to screen for general health problems and then problems are examined more closely with further testing (Kaminsky & Good, 1996; Shinn, 1989). Thus they were designed to be brief and convenient to administer, and their validity is dependent on their ability to predict future reading ability and correlate with more involved measures of valued early abilities. Hintze, Ryan and Stoner (2003) found strong correlations between DIBELS and CTOPP measures, helping to validate that the DIBELS measures generally correspond to measures of phonological awareness, phonological memory, and rapid serial naming, known to predict acquisition of decoding-level skills. The DIBELS also includes screening measures targeting knowledge of the alphabetic principle and basic phonics ability. Speece et al. (2003) looked at the correlations between students' performances on DIBELS curriculum-based
measures Nonsense-Word Fluency, Letter-Naming Fluency, and Letter-Sound fluency, and found them to correlate strongly into a single factor, which they termed “alphabetic construct”.

The signature of the DIBELS measures is that they incorporate a fluency component in addition to testing the skill in question, so Phoneme-Segmentation Fluency, for example, represents a combination of a child’s phonological awareness as well as his or her “fluency” in accessing and articulating phonological information stored in long-term memory.

A large body of research has examined the predictive validity of the DIBELS measures. Gofreda and DiPerna (2010) provide a comprehensive review of the research. However, all of these predictive studies measured either concurrent outcomes on a high-stakes test administered close in time, or outcomes within two years from the predictor variable. Longitudinal studies are rarer and reviewed in further detail later.

Nonsense word fluency (NWF) is a DIBELS measure intending to measure letter-sound correspondence through a fluency framework--the student must say the sound of each letter in a CVC “nonsense word” and then, if they are able, attempt to blend the three sounds together to pronounce the whole word. The number of sounds named correctly per minute is recorded. Gofreda and DiPerna (2010) reviewed eleven studies of validity evidence for NWF. Most measured predictive validity of NWF of kindergarteners and first-graders for predicting performance on reading comprehension tests in later grades or later the same year. Coefficients ranged from .41 to .80 with too
much variation between studies to summarize.

Phoneme segmentation fluency (PSF) is a DIBELS measures intending to measure phonemic awareness, also through a fluency framework. The student is told a word and asked to break the word up into its composite phonemes. The number of phonemes correctly segmented in one minute is recorded. Gofreda and DiPerna's review included 12 studies of the validity of PSF. Mostly predictive validity was measured in relation to other reading decoding measures. Predictive validity was variable between studies but consistently low, ranging from .02 to .65 (Gofreda & DiPerna, 2010).

The sample sizes used in the studies included in Gofreda and DiPerna's review were generally small, with the exceptions of Shilling et al. (2007) with a sample of 2,588 first-graders and Riedel (2007) with a sample of 1518 first-graders. Based on their substantially larger samples, these studies' findings command more legitimacy than others. Shilling et al. (2007) reported predictive relationships between first-graders' scores on DIBELS measures and their scores on the Iowa Test of Basic Skills two years later. They reported a correlation coefficient of .32 for PSF, .57-.60 for NWF, and .69 for Oral Reading Fluency (ORF). Riedel (2007) also examined first graders and reported coefficients of .15-.26 for PSF predicting performance on a later reading test, .41-.47 for NWF, and .77 for ORF.

Although measured later, once a child is able to decode words, his or her oral reading fluency rate has proven to be a powerful indicator of overall reading ability (Shinn, 1989; Fuchs et. al., 2001). A child's reading fluency skill is influenced by both
lower-level decoding-related skills such as phonemic awareness, letter-sound knowledge, and word recognition, as well as higher-level linguistic ability such as vocabulary knowledge. In this way reading fluency ability acts as a sort of mediating variable between the distinct set of abilities that we refer to when we think of “early reading” ability versus “advanced reading” ability. Fluency in reading connected text is both a powerful indicator of overall reading ability and a prerequisite skill for comprehension to be possible (Allington, 1983; Fuchs et al., 2001). As mentioned previously, in order for sufficient cognitive attention to be freed up for comprehension during reading, a student cannot also be attending to word-by-word decoding but must have achieved some threshold level of automaticity (Carreker, 2005). Also, it is amenable to more valid and reliable measurement, both static and formative, of reading ability since it can be easily represented as rate of words per unit of time. Within a responsiveness-to-intervention model for diagnosing reading disabilities and a tiered intervention system, curriculum-based measures of oral reading fluency are used both to describe a child's global reading health compared with a normative or local sample of peers, and to sensitively track individual growth in skill over time (Deno, 1998). One body of research has demonstrated the way that average young readers' fluency rate climbs steadily, at about 1 word a week increase (Fuchs & Fuchs, 1997) for several years from first grade into late elementary school, where it flattens out.

The validity of using oral reading fluency measures to gauge a child's overall reading “health” lies in its relationship to the true goal of literacy instruction, reading
comprehension. Research has supported that reading fluency is highly correlated with reading comprehension ability (Fuchs et. al., 2001; Gofreda and DiPerna, 2010). In one study, for example, the oral reading fluency abilities of third graders accounted for 56% of the variance (correlation coefficient .75) in their individual differences in reading comprehension ability (Schatschneider et al., 2004). Gofreda and DiPerna's (2010) review of the validity research on the DIBELS revealed strong evidence for concurrent and predictive validity of oral reading fluency (ORF) probes. They reviewed seventeen studies measuring both concurrent and predictive relationships between ORF and various reading assessments. ORF concurrent coefficients ranged from .59 to .88 with most estimates between .73 and .78. Only five studies measured predictive relationships for ORF and published coefficients from .67 (Good et al., 2001) to .92 (Burke & Hagan-Burke, 2007), although predictions were two years or under. However, longitudinal studies have demonstrated that as students age and reading becomes more complex, the predictive power of oral reading fluency fades (Tilstra et. al., 2009).

Modeling research has identified many factors which significantly affect reading fluency. According to Ehri et. Al. (2002), five main factors are: the proportion of words in a passage that can be recognized by sight; variations in speed with which sight words are processed; speed of decoding processes including letter-sound correspondences, morphological knowledge, and meaning knowledge; use of context; and speed at identifying meanings (Ehri, 2002). According to Torgeson (2006) and colleagues, rapid word recognition contributes the greatest proportion of variance towards reading fluency.
This does not mean, however, that fluent proficient readers have a memory store of visual representations of entire words that they match with visual input while reading; eye-movement research has shown that even highly proficient readers look at the phonemic units inside words— they just do so at an extremely rapid, automatic rate (Rayner, 1998).

As mentioned previously, while specific phonological abilities, such as phonemic awareness and rapid naming, are valuable in predicting future problems learning to decode, other language abilities which can also be measured in young children have demonstrated similar predictive power. This fact challenges the “phonological model” of reading acquisition, or at the least adds a great deal of complexity to the picture of how children learn to read. More specifically, the predictive ability of various skills and abilities appears to vary greatly depending on the age of the child, and the specific outcome variable being predicted (Scarborough, 2005). Different types of language-based tasks predict different aspects of literacy achievement. Many studies have demonstrated that phonological awareness specifically is more closely related to acquiring early decoding skills, and that its predictive power fades out when predicting later reading comprehension ability; modeling studies have found that phonological awareness has only an indirect effect on later reading ability through its direct effect on word-level decoding skills (Leppanen et al., 2008; Scarborough, 2005). Further complicating matters is the fact that phonemic awareness is highly dependent on instruction and has been found to be practically nonexistent in students with a history of no or minimal literacy instruction, or those learning in non-phonological orthographies in
which the focus is on acquiring associations between whole words and their symbols (McBride-Chang, 2004; Scarborough, 2005). Longitudinal modeling and regression studies which have added a variety of measures of language skills and linguistic comprehension into their regression equations have found them to explain individual differences in reading achievement in increasing measure as they attempt to predict reading comprehension ability in older students (Scarborough, 2005). Later reading comprehension ability is more complex and difficult to predict. The substantial predictability of later reading ability from early reading success does not itself explicate which component skills and abilities are associated with reading success, nor which are more or less important for achieving it. To more fully understand the complex ability to read, be able to predict future reading ability and to diagnose specific problems, a decades-long body of multivariate empirical investigations has sought to identify its predictors and components using mathematical modeling techniques.

Factor analysis is a statistical procedure used to assess the degree of correlation between various factors in order to determine which variables, such as in individual test scores, co-vary, or correlate highly with each other, and which are independent of each other. For example, a child's hair color and reading ability are unlikely to correlate closely, whereas many cognitive abilities and skills are highly correlated in individuals. Observed variables, such as individual test scores, which correlate closely with each other in an individual can be theoretically grouped into latent variables based on shared correlation, and named as an independent theoretical ability, such as “decoding”, or
“listening”. Path analysis is a mathematical method used to determine which latent variables or constructs predict other variables, and to quantify the interrelationships between multiple variables. Multivariate models can then be presented which illustrate the various correlations between different variables and which, taken together, explain some portion of the variance of an outcome measure of interest.

Examples of early models accounting for individual differences in reading comprehension skill include Calfee and Drum's (1986) model which identified the following significant predictive factors: decoding, vocabulary, sentence comprehension, paragraph comprehension, and text comprehension. Curtis (1980) looked at decoding and linguistic factors across different grade levels and found that decoding made significant, independent contributions for all the grades he studied—two, three and five—and linguistic comprehension was significant only for grades three and five. For third-graders, decoding accounted for three to thirteen percent of the variance of reading comprehension ability, while linguistic comprehension accounted for 23-35 percent. Singer and Crouse (1981) used path analysis to identify components of reading skill for 6th graders. They concluded that “decoding and linguistic comprehension (assessed as vocabulary knowledge) were both causally related to reading comprehension after removal of the effects of nonverbal intelligence”. Among the sixth-graders, decoding accounted for 29 percent of the explained variance, and linguistic comprehension 71 percent. Another study described in the same article used Stanovich et al. (1984)’s data and presented a path analysis model including nonverbal intelligence and phonological
awareness in which “only decoding and linguistic comprehension made significant independent contributions to reading comprehension”. After removing effects of nonverbal intelligence, decoding accounted for 19% (in grade three) and 38% (in grade five) of variance in reading comp; linguistic comprehension for 14% (grade three) and 13% (grade five) (Singer & Crouse, 1981). Butler et. al. (1985) published an early model predicting 33% of the variance at 1st grade, 42% at 2nd grade, 49% at 3rd grade, and 44% at 6th grade using the child's sex (girls tend to read better than boys), parent’s language, and six literacy-related factors (Butler et al., 1985).

Probably the most influential model published is Hoover and Gough’s (1990) Simple View of Reading. As its name implies, this model suggested that reading had previously been inaccurately characterized as “complex” in the sense that many of the variables Calfree and Drum and others had proposed actually correlated together into only two distinct independent factors: linguistic comprehension and decoding ability. Students with low linguistic comprehension could no more comprehend what they read than could students who could not decode; both linguistic ability and decoding ability limited a child's degree of reading comprehension. Beyond showing that each of these abilities made significant independent contributions to the outcome variable, the authors also hypothesized that the product of the two would add further predictive power to explain individual variance in reading comprehension, since the absence of either ability would result in zero comprehension. They tested the model on English-Spanish bilingual children in first through fourth grade. As they had predicted, when the product of
students' decoding and listening comprehension scores was added into the regression equation, estimates of students' reading comprehension ability significantly improved. Also, the two ability areas were independent, but the size of the correlation between them depended on the skill of the reader as well as grade level. Less skilled readers demonstrated lower correlations between their linguistic comprehension and decoding ability. And in the younger grades, more variance in comprehension was related to decoding skill, while linguistic comprehension became more relevant to reading in the fourth grade, particularly for good decoders (Hoover & Gough, 1990).

Many subsequent studies have tested the SVR and have generally supported its validity. Researchers have tested both summative (Decoding + Comprehension) and multiplicative (Decoding x Comprehension) versions of the model on samples of children in second through tenth grades and have been able to account for 40-80% of variance in reading comprehension ability (as reviewed in Tilstra et. al., 2009). The SVR has proven useful in its ability to distinguish readers experiencing primary difficulties with decoding versus with comprehension (Tilstra et. al., 2009).

More recently, researchers have debated whether the SVR should actually be made more complex to account for more variance in reading comprehension, arguing that key components are missing. One ability that is notably not included in the original SVR is reading fluency. Fluency has long been identified as a theoretically separate and important skill related to reading comprehension ability (LaBerge & Samuels, 1974; NRP, 2000). Tilstra et. al. (2009) tested a model adding fluency to the SVR and focused
on older readers, hypothesizing that including fluency measures would add significantly to the ability to predict reading comprehension in the upper grades. They explained that students’ fluency specifically in reading connected text could add significant variance to the equation, citing Kirby and Savage (2008), who argued that measures of fluency may be more sensitive to reader's proficiency with larger orthographic units like patterns of words (Tilstra et al., 2009). They found that fluency did contribute independently, and that for students in grades seven and nine, the combination of decoding and listening comprehension explained less than half of the variance in students’ reading comprehension abilities, while fluency contributed significantly at those ages.

However, Tilstra and colleagues also found that a major weakness of their original model had been the use of a listening comprehension measure to represent what the SVR called “linguistic comprehension”. The authors then added an expressive measure of verbal proficiency (defining words) into the equation and were able to account for further variance. They noted that the receptive test of listening comprehension was progressively less related to reading comprehension as students aged and speculated that listening comprehension could be less related to older students’ linguistic comprehension because it poorly represents students’ CALP ability. By grades seven and nine, the authors suggested, students learn the majority of new vocabulary items from written text as opposed to expanding their linguistic repertoires by listening to oral language as in younger ages. The authors suggested that future studies include more varied linguistic comprehension measures, in particular both receptive and expressive
tasks (Tilstra et. al., 2009)

Other modeling studies with both native English-speaking students and bilingual students have supported the relationship between age and the relative importance of linguistic as opposed to decoding-related factors in a students’ normative reading development. Beron and Farkas (2004) for example, using CFA, found that “early reading” was qualitatively different from “advanced reading”. These authors found auditory processing to be the strongest determinant of basic reading skills, and that auditory processing differences accounted entirely for the effects of both race and social class on reading performance. They concluded that “auditory processing is a principal conduit by which class and race effects on reading are expressed (p.125).” Also, they demonstrated that early reading achievement, in turn, is the primary determinant of advanced reading performance. Sociolinguistic researchers using qualitative methodologies have also found oral language to play a central role in social stratification as it plays out in day-to-day interactions at school related to learning and literacy. Heath (1983) examined the oral language of two cultural groups, finding that white middle-class children were socialized into particular linguistic patterns at home that mimicked those privileged by schools, such as questioning for information.

The distinction of two factors- decoding and linguistic ability- was also tested by researchers at the Woodcock-Munoz Foundation in a study identifying the variables that affect reading comprehension in childhood, adolescence and early adulthood (Floyd, Gregg & Keith, 2004). This study also found both “reading decoding” and
“comprehension-knowledge” to be significant across ages, and that the magnitude of their effects varied as a function of age. With increased age came increased dependence on “comprehension-knowledge” to determine reading comprehension, and decreased variance by decoding-related abilities. The authors divided “comprehension-knowledge” further and found that a factor representing “general information/knowledge of culture” demonstrated significant effects across the age levels, while “listening comprehension” had an inconsistent effect and “word knowledge” demonstrated no significant effects.

Similarly, Willson and Rupley (1997) compared the variables affecting reading skill for students in grades two and three with those for students in the fifth and sixth grades. These researchers identified three distinct latent variables which explained reading comprehension: “decoding” comprised of print awareness, alphabet knowledge, and knowledge of letter-sound correspondences; “word identification” comprised of orthographic, phonemic and meaning-based processes; and “background knowledge” both general knowledge about the topic and specific knowledge of the content of the text. They examined the comparative effects of these variables on reading comprehension ability at different ages. Their data shows that the construct “word identification” continues to play a role in reading comprehension ability until around age 12. They concluded that “for Grades 2 to 3 narrative text-based reading, comprehension is primarily driven by phonemic knowledge and secondarily by background knowledge for the text (Willson & Rupley, 1997, p. 45).” In this study, background knowledge was conceived as topic-specific concrete knowledge related to the theme of the text. By
grades five and six, both background knowledge and phonemic knowledge had dropped to insignificance in explaining variance in fifth and sixth-graders’ reading comprehension ability. Comprehension success was determined by strategy knowledge, as opposed to specific knowledge of the topic of the text.

Other models have included cognitive, emotional, and/or status variables to attempt to account for more individual variance in reading skill outcomes. Lepola et. al. (2005) tested a model with cognitive abilities and emotional variables including task orientation and social dependence, finding that the emotional variables had stronger predictive power than the cognitive measures. Leppanen et al. (2008) found that letter knowledge in kindergarten was the most powerful predictor of reading skill at the end of fourth grade. But they also found significant contributions for metacognitive awareness, gender, mother's education level, and visual attention. They found that phonological awareness at kindergarten affected reading in fourth grade only indirectly through its effect on kindergarten and first grade abilities, which then strongly predicted fourth grade ability.

Also muddling potential conclusions, studies have adopted differing ways of defining “linguistic comprehension”. Clearly, there is something that begins to control a large portion of the variance in reading ability sometime between late childhood and adolescence that has been identified and at times quantified by a constellation of related labels—verbal intelligence, linguistic comprehension, linguistic ability, vocabulary knowledge, fluid reasoning, and strategy knowledge. Stanovich (1984) has argued that
intelligence plays an increasing role in predicting reading achievement with age; however, Willson (1991) countered that intelligence plays no role in determining reading ability until after elementary school. More recent research has identified language-based variables such as vocabulary knowledge to be valuable predictors of advanced reading skills (Scarborough, 2005; Joshi & Aaron, 2012; Chall et. al., 1990; Proctor et. al., 2005). Researchers have realized, too, that the linguistic abilities needed to succeed in the upper grades are far more complex and extensive than vocabulary knowledge alone, and have begun to define and understand “academic English language” as a linguistic register transmitted socially within some families of particular cultural background and/or provided through exposure to advanced linguistic forms in school, and which manifests as a hidden determinant of the ability to manage the literacy demands of higher-level academic pursuits (Bailey, 2007). This variable appears to mediate the relationship between intelligence and reading achievement, just as phonemic knowledge within a certain language system eclipses the predictive relationships between socioeconomic status, race, and early reading achievements.

As students age it becomes more difficult to identify all of the specific component skills accounting for the variance in their academic performance due to the “increasing diversity of cognitive function” in development within individuals (Williston & Rupley, p.61, 1998)”, and the diverse nature of outcome measures considered. In other words, “reading comprehension” can mean many things, particularly in higher grades. How complex is the text? What type of text? What genre? What field of study was the text
taken from? As students grow older their success is increasingly based on their mastery of a select set of speech and literary genres—those privileged by the dominant cultural forces in society. On the other hand, prediction becomes statistically more reliable because of the “fan effect”—older children demonstrate a larger spread between low and high ability, making predictions more reliable for them than for younger children whose abilities all lie closer to the mean (Williston & Rupley, 1998).

In sum, longitudinal modeling research seeking to predict reading comprehension has demonstrated that when a linguistic variable is included to help explain variance, it accounts for an impressive portion of the variance in the dependent variable, even among native English speakers learning to read in their own language, and increasingly as students age, decode automatically, and texts become more complex. “Reading” and even “reading comprehension” mean something very different for young children versus older children. Early reading ability is more about decoding, word identification, and then developing fluency as a bridge to basic comprehension, while later reading ability is best predicted by including variables such as linguistic knowledge, verbal proficiency, verbal intelligence, and background knowledge. The “linguistic variable” is a loose construct which many have interpreted as “listening comprehension” but which is apparently better represented by a mix of measures to include both receptive and expressive aspects of verbal proficiency (Tilsrta et. al., 2009; Scarborough, 2005).

Models accounting for individual difference in reading ability in terms of advanced reading comprehension have also led to new ways to conceive of reading
disability. The concept of weakness in phonological awareness as a cause of reading
disability has been called the Phonological Model, and has become incorporated into
many widely-accepted definitions of dyslexia, which claim that dyslexia “typically
results from a deficit in the phonological component of language”, considering further
problems such as decoding deficits, comprehension deficits, and vocabulary deficits to be
“secondary consequences” of the original and causal phonological problem
(Scarborough, 2005; Lyon, Shaywitz, & Shaywitz, 2003, p.2). However, models
demonstrating equal or greater influence of a multitude of other factors, including various
linguistic skills, have challenged the assumption of a causal trajectory beginning with
phonological weakness causing a phonics deficit and then leading secondarily into
Matthew Effect results (Scarborough, 2005; Joshi & Aaron, 2012). The original SVR
model, for example, was able to account for a significant amount of variance in students'
reading comprehension ability (40-80%; Hoover & Gough, 1990; Catts et. al., 2006;
Joshi & Aaron, 2012) by including both decoding ability and linguistic comprehension,
and could also identify in which area a student was weakest in order to guide intervention
efforts. Aaron and colleagues (1999) found that in their sample of 198 third-, fourth-, and
sixth-graders, 7% experienced adequate decoding and poor comprehension, 8% adequate
comprehension with poor decoding, and another 8% experienced both deficits, amounting
to about 25% with a reading disability if so defined, as well as identifying how to
intervene. Adding fluency as a separate factor seems promising for further intervention
information (Tilstra et. al., 2009).
Aaron and colleagues (2008) proposed an extension of the SVR for describing and defining reading disability called the Component Model of Reading (CMR). In their model, reading ability is affected by many variables within three domains: cognitive; psychological; and ecological. The cognitive components include word recognition (decoding and sight word knowledge) and linguistic comprehension; the psychological components include motivation, teacher expectancy, and interest among others; and the ecological components include instructional factors, dialect, home environment factors, and English-as-a-second-language status (Aaron et. al., 2008; Joshi & Aaron, 2012). Of course, while accounting for much of the variance in reading comprehension ability in individuals, this broad inclusion of status and environmental variables challenges more than just the definition of SLD, but the larger way that we think of “disability” as a primarily neurological problem.

1.8. Models with ELLs and predictive bias

Although ELLs demonstrate an achievement gap compared with native English speakers of all demographic groups, studies with outcome measures of basic reading decoding ability (e.g., word recognition, reading fluency on controlled passages, literal comprehension of simple text) have generally found more similarities than differences in the way that second language students and native English speakers learn to read (Carlo, 2007; Cheung & Slavin, 2005; Geva & Zadeh, 2006; Lindsey, Manis & Bailey, 2003; Slavin & Cheung, 2004; Leseaux, Rupp & Siegel, 2007). Additionally, research has demonstrated strong similarities in the underlying cognitive and psycholinguistic abilities
that predict young English-speaking and ELL’s facility in learning to decode written language (Geva & Zadeh, 2006; Lindsey et al., 2003; Proctor et al., 2006). Geva and colleagues (2006) studied the component processes that contribute to word- and text-reading efficiency in L1 (first language) readers compared to L2 (second language) students. They found that rapid automatic naming, phonological awareness, letter-sound knowledge, and sight word knowledge played similar roles in predicting word and text reading efficiency in L1 and L2 learners.

Many of the cognitive processes that predict and facilitate learning to decode basic text transfer across languages. Branum-Martin et al. (2006) found that English and Spanish phonemic awareness overlap highly in bilingual children. Several studies have demonstrated that phonological abilities measured in either language predict how easily ELLs learn to decode basic texts in English (Durgunoglu et al., 1993; Geva & Zadeh, 2006; Lesaux, Rupp & Siegel, 2007; Lindsey, Manis & Bailey, 2003). Generally speaking, the bottom-up cognitive processes related to early decoding skill acquisition are the same regardless of a child’s home language, and in the case of Spanish, phonological abilities transfer readily to English decoding (Durgunaglu, Nagy, Hancin-Bhatt, 1993).

Modeling studies on language minority students learning to read in English have focused on examining the additional importance of linguistic knowledge as a determinant of literacy acquisition. If linguistic comprehension ability, as distinguished from decoding, plays a progressively larger role in determining age-appropriate reading comprehension skills for all children as they age (Scarborough, 2005), it stands to reason
that language minority students' comparative reading abilities would suffer more in later years, when linguistic comprehension replaces decoding as the main cognitive task of reading. However, this effect could logically also be offset by ELL’s continual development of further language.

Proctor and colleagues (2005) tested a predictive model of reading comprehension for English language learners at several ages using equation modeling methodology, and found English vocabulary knowledge to be the greatest predictor of reading comprehension, both indirectly and directly. As with native speakers, their model also suggested that for ELLs English vocabulary knowledge plays a steadily increasing role in reading comprehension ability with age. Similarly Gough and Hoover (1990)’s SVR was originally tested on Spanish-English bilinguals, finding that linguistic comprehension plays an increasingly important role in determining reading comprehension ability with age. These results suggest that even after being re-classified as proficient in English, a linguistic disadvantage could continue to affect ELLs’ academic performance, particularly reading comprehension, into later grades.

Geva and Zadeh (2006) have presented data suggesting a threshold level of text difficulty, at which children learning to decode in a second language reach a ceiling in potential skill development due to the limitations of their linguistic comprehension.

Geva & Farnia (2011) found that word and text reading fluency in ELLs was single factor in 2nd grade but separate factors in 5th grade, when fluency was more related to linguistic comprehension. In other words, ELLs were slowed-down by
linguistic limitations in a way that showed up on a later reading skill. They examined several factors at grade 2 -- cognitive, language, word-reading, and reading fluency -- and found that phonological awareness and vocabulary predicted most of the variance in those students' reading comprehension ability in the fifth grade. As other researchers had discovered, for ELLs, adding linguistic assessments added further predictive power. Measures of ELLs' English syntactic skills and listening comprehension accounted for further variance in their reading ability at grade 5, while those measures did not add to the predictive power of the model for native English speakers.

One dissertation study from Florida Reading First schools examined the predictive power of DIBELS measures administered in kindergarten to predict ELLs' oral reading fluency in the first grade, finding that kindergarten English letter-naming fluency was the best predictor, with English vocabulary knowledge coming in second (Ummahan, 2010). Ummahan also found that race, gender and eligibility for free and reduced price lunch overshadowed ELL status to predict impaired reading fluency in the first grade, with the average white female not eligible for free and reduced lunch to read as fluently as average native English speakers by the end of that grade. She was also able to account for further variance in ELLs' oral reading fluency ability by adding a measure of English vocabulary knowledge in kindergarten into the equation.

Another dissertation study found that concurrent correlations between DIBELS Oral Reading Fluency scores and TerraNova-Reading (measuring reading comprehension) scores increased for ELLs from first to third grades, with third grade
predictions similar to estimates on native speakers found by other research. Additionally the researcher found that adding an English receptive vocabulary measure successfully explained further variance for ELLs and in fact overshadowed the predictive validity of oral reading fluency for first grade ELLs (Millett, 2012).

Several studies including Shapiro & de Ramirez (2006) have found oral reading fluency progress-monitoring to be reliable to measure growth in reading skill in ELLs. In other words, ELLs' reading fluency grows steadily with growth in skill.

But these findings do not address the question of whether their oral reading fluency rate is as predictive of comprehension abilities as for native English speakers, or whether oral reading fluency demonstrates “predictive bias” based on ELL status. Predictive bias is not a new concept (Cleary, 1968). It basically asks whether a single regression model holds true for different groups, and linear regression models are typically used to examine the predictive bias of tests (Betts et. Al., 2008). “Test bias” is a more general term examining whether tests are psychometrically “fair” across groups, especially for students with particular differences hypothesized to affect test performance. A test’s predictive bias specifically has to do with whether the test under- or over-predicts certain groups’ performance on other assessments at a concurrent or future administration (Betts et. al., 2008; Hintze et. al.,2002; Hosp et. al., 2011). Longitudinal studies of predictive bias focus specifically on predicting achievement over time.

A few studies have examined whether curriculum-based measures of reading
exhibit predictive bias based on race or ethnicity. Kranzler and colleagues found evidence of bias by race combined with grade level (Kranzler et. al., 1999). However, Hintze and colleagues (2002) identified problems with their design and did not find evidence of predictive bias for oral reading fluency for African-American students versus Caucasian for the predictive validity of oral reading fluency (Hintze et. al., 2002).

Only a small number of studies have assessed predictive bias based on home language or ELL status, and results have varied widely between studies (Hosp et. al., 2011). Catts et. al. (2009) attributed at least part of the lack of consistency across predictive validity studies on the potential of severe floor effects in early literacy measures. Floor effects occur when a measure’s scale does not go low enough to capture a student’s ability and many scores are bunched at the lower end of the scale. This negatively affects predictive ability of the measure. And, if one group exhibited more severe floor effects than the other, the result will show differential prediction between groups for the measure (Hosp et. al., 2011; Catts et. al., 2009). Catts and colleagues (2009) examined a large sample in Florida and found severe floor effects for Kindergarten measures, less severe floor effects for first grade measures, and no effects for second or third grade measures. They showed how these effects caused problems for the measures’ predictive validity estimates, and called for more sensitive early literacy measures.

Klein and Jimerson (2005) found predictive bias as a function of ethnicity (Hispanic and Caucasian), home language and SES for oral reading fluency and reading
comprehension. Riedel (2007) found that DIBELS ORF and reading comprehension measures were more highly correlated in ELLs than non-ELLs, and that ELLs experienced lower predictive accuracy for Phoneme-Segmentation Fluency at .37-.39. However, their study had a very small sample of ELLs compared with native speakers.

Wiley and Deno (2005) examined the DIBELS Maze and ORF tasks with ELLs and native English speakers in third and fifth grades and found differences in predicting their performance on a concurrent high stakes state exam. However, Roehrig et al. (2007), found no evidence of predictive bias for ORF predicting performance on the state exam based on SES, race (African American vs. Hispanic), or ELL status.

Fien et. al. (2008) conducted a review of 24 studies comparing ELLs and English speakers. Seven of these found differential prediction. However, five used assessments from winter of kindergarten and potentially were affected by the differential floor effects at the predictor variable (Catts et. al., 2009).

Grant et. al. (2011) predicted 3rd grade reading comprehension among several language groups and native speakers, finding differences in predictive validity for receptive vocabulary, decoding, and print exposure, mediated by language proficiency. They additionally pointed to need to divide sequential and simultaneous bilinguals into different groups.

Schieffel et. al. (2012) examined the sensitivity and specificity of all DIBELS measures to predict students at risk for reading problems. They found that for all students, the DIBELS is better at predicting reading success than failure. They also found that
DIBELS demonstrated greater sensitivity (higher true positive rate) for ELLs than non-ELLs but better specificity (higher true negative rate) for non-ELLs than for ELLs.

Betts et. al. (2008) examined whether fluency-based measures of early literacy skills included in the MKA exhibited predictive bias for different subgroups of students in terms of predicting their reading ability two years later, in the second grade. They found bias for Hispanic American students, but not for ELLs. However, they noted that their ELLs were divided between two major languages, and that the Spanish-speaking students performed lower on average in Kindergarten and could be studied as a separate group.

Hosp and colleagues (2011) conducted a predictive bias study with a large sample of over 3000 students, examining differences across various demographic categories in terms of using DIBELS Nonsense-Word Fluency (NWF) and Oral Reading Fluency (ORF) probes at first grade to predict performance on a state criterion-referenced test in the third grade. They found evidence of widespread predictive bias based on subgroup membership. They noted that evidence of predictive bias or lack of bias varies greatly across studies with no clear pattern evident, thus no conclusion about the true degree of predictive bias presented by each predictor variable onto each outcome variable can be drawn by synthesizing the current research. Rather, it is likely that inconsistency between other variables, such as instruction in between the data points or the timing or type of outcome variable, makes it very difficult to compare results.

Other predictive research on ELLs has put to empirical test Cummins' (1979)
linguistic interdependence hypothesis, which suggests that acquisition of a second
language is mediated by a child’s level of competence in his or her native language.
Kieffer (2012) found that oral language variables in both Spanish and English predicted
English reading although only English language proficiency was uniquely predictive. For
ELLs, author found that English productive vocabulary was a better predictor than
Listening Comprehension. Proctor (2006) found that Spanish vocabulary strongly
predicted English reading achievement for bilingual students, with the effect stronger for
more fluent readers. August et. al. (2006) found that first-language reading skills are
related to second-language reading skills, but that for this relationship to occur there must
exist some reading skill in the first language.

1.9. Present Study

1.9.1. Purpose

This study seeks to contribute to the small body of scientific literature examining
the role of race and English language proficiency in potentially mediating the relationship
between young children’s literacy-related abilities and later achievement in English
reading comprehension, specifically in the context of an English-only subtractive
linguistic school environment.

1.9.2. Research Questions

The following empirical questions structured the design of the present study:

1) How do the subtests of the various early literacy and language assessments
administered to Holyoke children load onto the latent constructs they are hoping to
measure? How many factors emerge? Can all of these subtests be described accurately as measuring three distinct constructs: phonological ability, basic decoding ability and English oral language ability? To what degree are the latent constructs correlated in this population?

2) What degree of predictive relationships do the latent constructs which emerged from the factor analysis have with later measures of reading achievement? Do they have a statistically significant direct effect on the outcome variable? Are there indirect effects? How much variance in the outcome variable is explained by the model?

3) Are there statistically-significant differences in the predictive correlations between the individual measures and the outcome variable based on the native language of the child?

4) Are there statistically-significant differences between the fit of the overall model predicting fourth grade reading comprehension based on the native language of the child? Are there statistically-significant differences among any of the specific parameters tested within the model based on the native language of the child?

1.9.3. Design

The researcher examined four years of student achievement data from Holyoke, Massachusetts, a small city whose public school population mostly has mostly come from Puerto Rico within one to two generations. The wide majority of students in the present study's sample are Hispanic (97%). Over half were identified upon school entry as English Language Learners (ELLs). The district is one of the lowest-performing in the state in terms of performance on state academic assessments, and is one of the poorest
districts as well. The Reading First grants won by Holyoke’s elementary schools incorporated the use of curriculum-based measures to monitor students’ progress towards reading goals. The question begs, then, how early indicators measuring underlying cognitive abilities such as phonemic awareness, and early literacy knowledge such as letter names and letter-sound correspondences, will be useful in screening for children likely to need help learning to read if those children are also learning a second language and losing their first in the process.

This study provides evidence of the predictive validity of early literacy measures to predict fourth grade reading comprehension ability for a sample of Puerto Rican Hispanic students. Additionally, factor analysis is used to identify the component skills comprising their early reading ability in the first grade and the degree to which the combined first grade measures predict the fourth grade one. A structural equation model is tested describing the concurrent and longitudinal relationships between the first and fourth-grade skills. Then the group was divided based on home language, to determine whether home language made a difference in their achievement at both ages, and in the predictive validity of early the literacy measures.

In order to answer the research questions, the researcher analyzed a subsection of this district’s extant student achievement data, which includes data on all students in the district, beginning with the 2004-2005 school year to the 2008-2009 school year.
CHAPTER 2

METHODS

2.1. Setting

The subject town of Holyoke, Massachusetts was chosen partly for convenience. However, the city also represents a prototypical example of a sociological phenomenon in notable occurrence in schools all over the United States—an influx of Puerto Ricans has gradually made this small city in Western Massachusetts one of the most Hispanic places on the East Coast. At the same time, different racial groups are mixing less in the public schools; Holyoke's Irish population, previously the majority, more frequently send their children to the town's many parochial schools.

Holyoke is nicknamed “Paper City” and is best known for its once-flourishing paper industry which peaked early in the twentieth century. Working-class Irish immigrants dominated the city for most of its history, and Germans, Poles, and Jews had a strong presence as well. Prior to incorporation, Holyoke was known as “Irish Parish”. In the 1950’s Puerto Ricans began arriving, often via New York City. In 2004, approximately 37% of Holyoke’s population was of Puerto Rican descent, the largest percent Puerto Rican population of any city in the United States outside of Puerto Rico. According to the census of 2000, 41% of the population was Latino. Interestingly, though, 77% of children in Holyoke Public Schools are Hispanic as of the school years studied, according to the districts’ data (retrieved from

http://www.hps.holyoke.ma.us/rc_dean.htm on May 10th, 2010).
Holyoke is one of the poorest cities in Massachusetts. As of the census of 2000, about half its school children lived below the poverty line. Median household income was $30,441, and median per capita income was $15,913. Poverty is known to have a large negative impact on educational achievement. With many children experiencing both low socio-economic status during elementary years and learning in a second language, it is not altogether surprising to find them struggling to perform at a level comparable to children without these obstacles.

In 1981 Holyoke Public Schools settled a school desegregation lawsuit. The settlement agreement included both bilingual education and special education plans (Am.1981. Hispanic Parents Advisory Council v. City of Holyoke, Civil Action No.80-0172-FHF).

Manuel Frau-Ramos and Sonia Nieto (1991) published a program evaluation-style research paper about the Holyoke Public Schools in which they investigated the causes for students dropping out of high school, a significant problem for the community. They determined that 68% of Puerto Rican youth in Holyoke Public Schools left high school without graduating. The rate was significantly smaller among students born in Holyoke compared with those born in Puerto Rico, and students enrolled in Transitional Bilingual Education were half as likely to drop out as their counterparts in Sheltered English Immersion. In fact, one of their recommendations to the district was to expand the bilingual program, although any efforts in that direction were soon to be mooted by the passage of English-only legislation state-wide.
Based on many informal conversations with ESL teachers in Holyoke and web searches of blog entries and other web archives from the time, the passage of Question 2 in 2002 certainly did not go unnoticed, although the public reaction was not formally organized into a protest (while in California the law had led to various community protest actions) and drew no media coverage. According to several accounts, when the law passed, Holyoke’s bilingual programming was abruptly dismantled. Puerto Rican teachers, many of whom had been recruited in Puerto Rico to come to Holyoke and teach Spanish content areas within a bilingual program, were called into principals' offices and informed of their new titles as Instructors of English as a Second Language. They were told that they must immediately switch from teaching content areas in Spanish to teaching English as a Second Language. They were for the first time required to pass an English language proficiency exam; many could not pass and had to leave their careers in education because of their inadequate command of the English language.

The ESL teachers that I saw at the elementary level were in assisting positions within classrooms, although their education levels were comparable to the lead teachers. In conversation, several mentioned that prior to the passage of the legislation they had run their own autonomous classrooms teaching subjects and language lessons in Spanish. Since the passage of Question 2, ESL teachers deliver ESL lessons to pull-out groups, provide support for second language students mainstreamed into regular classrooms by translating instructions when necessary, and helping ESL students with academic work.

The wide majority of school staff with whom I spoke had staunchly opposed
Question 2. One ESL teacher in Holyoke, affected by the law, wrote afterwards that, “Almost all of the Hispanics and most of the Anglos I work with were strongly against Question 2. Those serviced by bilingual education voted strongly against Question 2 (retrieved from http://bilingualeducationmass.wordpress.com on May 20th, 2010)”

According to an exit poll survey of 1,200 Latinos, 92% of Hispanics voted against Question 2. A nationwide People En Espanol survey of 6,000 Hispanics found that 95 percent of Hispanic respondents back bilingual education. Many of those that wish to have their children in bilingual education programs may not be able to (retrieved from http://bilingualeducationmass.wordpress.com on May 20th, 2010). One first-grade Anglo teacher with whom I spoke expressed favor for the change, explaining that in English immersion, students seemed to acquire English reading skills more quickly.

2.2. Structured English Immersion program

Holyoke Public Schools include eight K-8 schools, two high schools, a charter school, and a pre-school. Structured English Immersion (SEI) is the model used for educating English language learners, who form a substantial portion of the school-aged population. This model is more or less explicitly mandated by Question 2. Children identified as “English Learners” are assigned an English-language proficiency level according mainly to their score on the Massachusetts English Proficiency Assessment (MEPA) and assigned a corresponding number of hours per day or per week to be pulled for specialized English language classes, taught by a teacher certified in teaching of English as a second language. For example, for students in levels 1 and 2 of English
proficiency, the district guidebook calls for “2.5 hours to a full day of direct ESL instruction.” At level 3, 1-2 hours per day is required along with 1-2 hours of specialized reading instruction; at level 4 and 5, a minimum of 2.5 hours per week of “direct ESL instruction” is mandated.

SEI has two components: “sheltered content instruction” and “English as a second language (ESL) instruction”. The term Sheltered English Immersion (SEI) was coined by Keith Baker and Adriana de Kanter (1983), who opposed bilingual education. The program is modeled explicitly on French immersion programs for English-speaking children in Canada. Its features include “significant amounts of the school day…. dedicated to the explicit teaching of the English language, and students are grouped for this instruction according to their level of English proficiency (Clark, 2009; p.3).” Secondly, academic content plays a subordinate role to the English language per se, the main content of SEI instruction. Clark (2009) states that the principle is that “students must have a strong understanding of the English language before they can be expected to learn grade-level content (Clark, p.4).” SEI instruction includes direct instruction in English grammar. SEI programs, according to Clark, are designed to last only one academic year. Additionally, all teachers are subject to multiple trainings regarding how to adapt lessons to incorporate language objectives and to make instruction comprehensible to second language learners (Clark, 2009).

2.3. ELL Classification system

Students in Holyoke Public Schools are placed into the English Language Learner
program based on their performance on either the Massachusetts English Language Proficiency Assessment (MEPA) or the Massachusetts English Language Assessment-Oral (MELA-O), for pre-literate children. Upon school entry, the assessment is administered only in cases in which the parent identifies the student's home language as other than English. The MEPA assesses four skills traditionally considered the pillars of learning a second language: speaking, listening, reading, and writing. The resulting level (out of five) represents a composite of the four skills. The MELA-O is related but only assesses speaking and listening, and is used in Holyoke in place of the MEPA for children too young to read and write.

Students are divided into five levels of English proficiency. Level 1 is the most non-proficient beginner, and Level 5 represents native-like proficiency. Only one level can be assigned to a student, since each level incorporates speaking, listening, reading and writing ability designations, assuming accordance between an individual’s literacy-based skills and his or her oral language skills.

A students’ classification is a decision made by a school-based team based upon a variety of considerations and data, not solely a MEPA or MELA-O score. However, generally a MEPA score of “4” is equivalent to re-classification as English proficient. According to the director of the district’s services for English language learners, teams also look for a passing score on MCAS English Language Assessment. He stated that students in Holyoke typically take many years to become “English proficient”.

2.4. Reading Instruction
During the years that data was collected, Holyoke Public Schools participated in Reading First, a federal Department of Education grant program focused on putting scientifically-based methods of early reading instruction into needy schools, with the stated goal that all children read well by the end of the third grade (http://www2.ed.gov/programs/readingfirst/index.html, 2013). In Holyoke, the Reading First grant included focus on the five aspects of reading instruction identified by the National Reading Panel (National Reading Panel, 2000). Additionally, the Reading First grant funded the collection of formative reading data including systematic screening (three times annually) using the Dynamic Indicators of Basic Early Literacy Skills (DIBELS, Kaminsky & Good, 1996) coordinated with the implementation of a three-tiered system of reading intervention delivery. Students scoring under threshold scores on screenings were given reading interventions using evidence-based intervention programs, and DIBELS progress-monitoring was used. Spring DIBELS screening results were used in the present study.

2.5. Participants

Student data to include all target measures (including DIBELS data collected during Reading First) was available beginning in the fall of 2004 until the Reading First grant was discontinued in 2010. However, data availability was further limited by students moving into and out of the district, as well as irregular collection and saving of much of the data collected during Reading First. In order to be included in the study, students needed to have all DIBELS and GRADE data from first grade available, as well
as their score on the MCAS in fourth grade. This restriction limited the sample size substantially as the district is characterized by high mobility and much of the Reading First data was lost or not collected and/or saved in the database. Additionally, students who were retained were not included as they did not appear in the same cohort in the fourth grade.

First grade was chosen because it is an important year for reading instruction and for predicting future reading ability. Under the Reading First grant, children in the first grade focused on developing phonics knowledge, learning more “sight” words, and gaining some automaticity in word recognition. At this grade level, children's abilities range widely from struggling to identify all letters and achieve basic decoding of CVC words, to students reading up to 50 words per minute on a novel passage. According to DIBELS assessment procedures, first-grade students are administered assessments measuring their phonemic awareness, basic decoding, and reading fluency, so a wealth of literacy-related assessment information was available at that grade level.

The outcome variable chosen was performance on the fourth grade language arts assessment which is part of the battery of educational assessments administered by the state to all fourth graders, the Massachusetts Comprehensive Assessment System (MCAS). This was considered a measure of grade-level reading comprehension.

The final sample contains 232 students in two cohorts-- one beginning the first grade in the fall of 2004 and another which began the first grade in the fall of 2005. No differences were predicted between cohorts. Participants attended all elementary schools
in the district. 97% of the sample was Hispanic; a few students were African American and a smaller number Caucasian. Forty-six percent (107) were native English-speakers, according to a question about dominant language spoken at home included in student enrollment data; 54% (125) were native Spanish-speakers.

2.6. Measures

The following are the first grade measures included in the present study. Three measures of the Dynamic Indicators of Beginning Emergent Literacy Skills (DIBELS; Kaminsky & Good, 1996) and five measures of Pearson’s Group Reading Assessment and Diagnostic Evaluation (GRADE) were included. Only spring administrations were used. These assessments were administered to participants during April and May of 2005 and April and May of 2006.

2.6.1. Description of DIBELS Measures

Oral reading fluency (ORF) - student reads aloud for one minute from connected text; words correct/minute

Nonsense Word Fluency (NWF) - student names letter sounds then attempts to blend non-words in CVC form; sounds correct per minute

Phoneme segmentation fluency (PSF) - student names separate phonemes after hearing a word aloud; phonemes correct/minute

2.6.2. Description of GRADE Measures

Word reading (WR)- Student hears a word in a sentence and then repeated in isolation and chooses correct written word from four choices
Word meaning (WM)- Student reads word and matches with picture depicting meaning
Sentence comprehension (SC)- Written cloze activity with single sentences- student
chooses correct answer to fill in blank from 5 choices
Passage comprehension (PC)- Student reads short paragraph then chooses correct answer
from five choices
Listening comprehension (LC)- Student listens to one-two sentences read aloud and
chooses picture that best describes meaning

2.6.3. Description of MCAS Fourth Grade English Language Arts Test

On the MCAS fourth grade English Language Arts section, the student reads texts
(silently) in several genres and responds to multiple-choice questions; Also, the student
reads texts in several genres and provides open-response answers (ranging from a
sentence to a short paragraph).

2.6.4. Tests' Reported Reliability Coefficients

The reliability coefficients reported below for DIBELS measures are all median one-
month alternate-form reliability coefficients.

Table 1: DIBELS reported reliability coefficients

<table>
<thead>
<tr>
<th>Test</th>
<th>Coefficient</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phoneme-Segmentation Fluency</td>
<td>0.67</td>
<td>First grade (Good et. al., 2004)</td>
</tr>
<tr>
<td>Nonsense-Word Fluency</td>
<td>0.83</td>
<td>First grade (Good et. al., 2004)</td>
</tr>
<tr>
<td>Oral Reading Fluency</td>
<td>0.95</td>
<td>(Dynamic Measurement Group, 2008)</td>
</tr>
</tbody>
</table>

The following Cronbach's alpha reliability coefficients were reported for GRADE
subtests (Williams, 2001).

Table 2: GRADE reported reliability coefficients

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Reading</td>
<td>.81</td>
</tr>
<tr>
<td>Word Meaning</td>
<td>.90</td>
</tr>
<tr>
<td>Sentence Comprehension</td>
<td>.90</td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>.88</td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td>.76</td>
</tr>
</tbody>
</table>

2.7. Statistical Procedures for Analysis

Confirmatory Factor Analysis (CFA) was used to answer the first research question. The LISREL software program was used to realize the analyses. In an exploratory mode, several CFA’s were tested using LISREL software to determine the arrangement of factors with the best overall fit describing how the measures related to each other and how many latent variables they described.

The second research question was answered by fitting a hybrid model using structural equation modeling methods and LISREL software. The best-fitting CFA measurement model was arranged so that first grade latent variables predicted fourth grade MCAS performance, conceived as a latent measure of fourth grade-level reading comprehension. The goodness-of-fit of the model was then tested and quantified using LISREL software.

The third research question was answered by calculating correlation coefficients for each pair of measures. Then, significance-testing was performed on each difference score. A generic online calculator was used (http://www.vassarstats.net/rdiff.html), which used Fisher r-to-z transformation to calculate a z-value in order to assess the
statistical significance of each difference between correlation coefficients.

The fourth research question was addressed using multiple group confirmatory factor analysis (MGCFA). MGCFA is a method used to evaluate formal hypotheses of parameter invariance across groups (Rios & Wells, in press). The goal was to compare the fit of the hybrid model, including comparing each estimated parameter, across the two groups of subjects as previously defined (English-speakers versus Spanish-speakers), to answer whether a subject's group membership affects that subjects' predictive estimates. A series of hierarchical steps is involved in conducting MGCFA. First, a baseline model is established separately for each group. Next, hierarchically nested models are tested systematically by constraining each parameter and then comparing the fit of each nested model. Change in the Comparative Fit Index (CFI) is used to measure the degree of invariance presented by each newly constrained nested model, until all parameters have been tested for invariance by group membership (Wells & Rios, in press).
CHAPTER 3

RESULTS

3.1. Test Performance of Entire Sample

On the fourth grade language arts assessment of the Massachusetts Comprehensive Assessment System (MCAS), raw scores from 0-38 fall into the “warning/failing” category, 39-47 defines “needs improvement, 54-62 is “proficient”, and 63-72 is “advanced”. The 232 students in the present sample performed significantly more poorly than average performances of students across the state. Students in the present sample obtained a mean raw score of 34.38, with a standard deviation of 12 points. This mean score falls into the “warning/failing” category. Scores among the subjects in the sample were fairly regularly distributed, suggesting that approximately 95% of scores fell between 10.38 and 58.38 and approximately 77% of scores fell between 22.38 and 46.38. See Appendix 1.

The group overall also performed below-average on most first grade early literacy measures. According to the DIBELS benchmarks established by normative research conducted by University of Oregon, a student who reads a median of 40 words per minute (WPM) or more on a first-grade level DIBELS Oral Reading Fluency (ORF) passage in the Spring of first grade is considered “low-risk”; scores from 20 to 39 WPM are considered “some risk”; and scores from 0 to 19 WPM are “at risk”. The mean score of students in the present sample was 24.24 words read correctly in one minute, a “some risk” score, with a standard deviation of 21.76 WPM. However, the distribution was
highly irregular, making it difficult to summarize using a single mean score. Importantly, 45.5% of the sample only read between 0 and 15 words correct in a minute, and 74.4% of the sample read 30 or fewer words correct in a minute.

Nonsense-Word Fluency (NWF) is scored as letter sounds correctly identified in a minute. According to DIBELS benchmark threshold scores, first graders who can say the sounds for 50 or more letters in a minute at the end of the year are “established” in their letter-sound correspondence ability. Students naming between 30 and 49 letter sounds per minute are “emerging”, while students naming fewer than 30 sounds are experiencing a “deficit” in this skill. The mean score of students in the present sample was 48.92, in the high end of the “emerging” range, with a standard deviation of 24.95, although the distribution was quite irregular, with a third of the sample scoring around 30 sounds per minute.

Scores on Phoneme-Segmentation Fluency (PSF) are calculated as the number of phonemes correctly identified in a minute. According to DIBELS benchmark suggested thresholds, first graders in the Spring who can segment between 0 and 9 phonemes in a minute are performing in the “deficit” range, scores between 10 and 34 are “emerging”, and a score of 35 or more phonemes per minute is considered “established” phonemic awareness. The mean score of the present sample was 42.42 with a standard deviation of 14.63, suggesting that on average students in the group demonstrated established phonemic awareness.

3.2. Group Differences in Performance Based on Home Language
The sample was divided into two groups based on students' first or home language, in order to compare the performances of ELLs with those of non-ELLs on each measure included in the study. This information originated from a question parents were asked upon registering their children for school. 107 (46%) of the 232 students in the sample were identified as native speakers of Spanish upon school entry, and 125 (54%) as native English-speakers. The mean score on the fourth grade MCAS ELA test for just the Spanish-speakers group was 32.12, while the mean for English-speakers was 37.44. This 5.32-point difference was statistically significant (p= 0.0006; two-tailed).

On Oral Reading Fluency, English-speakers obtained a mean of 27.58 WPM, while Spanish-speakers obtained a mean of 21.77 WPM, a difference of 5.81 WPM. This difference is statistically significant (P=0.0429; two-tailed).

On Nonsense-Word Fluency, the mean score for English-speakers was 52.636 and for Spanish-speakers was 46.296. This difference of 6.34 sounds per minute is not statistically significant (P=0.0535; two-tailed).

On Phoneme-Segmentation Fluency, English-speakers' mean score was 44.103 while that of Spanish-speakers was 41.304. The difference is not statistically significant (p=.0146; two-tailed).

Statistically significant differences in mean scores were identified for native English-speaking students versus ELLs on most measures: oral reading fluency; word reading, word meaning, sentence comprehension, paragraph comprehension, listening comprehension, and fourth grade MCAS scores. The sole exceptions were Nonsense-
Word Fluency and Phoneme-Segmentation Fluency. Nonsense-Word Fluency demonstrated nearly significant group differences; had a one-tailed test been used predicting Spanish-speakers earning a lower score the difference would have been statistically significant.

3.3. Research Question 1: Results of Confirmatory Factor Analysis

LISREL software was used to perform factor analysis. First, the covariance matrix was identified for all measures for the whole sample, with group differences ignored. A group of theoretical constructs, or factors, was proposed to describe the various measures, based on findings of previous studies. The first measurement model proposed incorporated four separate factors in all—three separate first-grade constructs and fourth grade reading comprehension. The goal was to identify the same independent factors identified in the literature: “linguistic comprehension” ability separate from “decoding” separate from “phonological” ability. Therefore each measure was first considered at face value; in other words, if a subtest purported to measure “comprehension” based on its own name and description, it was placed into the “linguistic comprehension” group to test that claim. “First Grade Linguistic Comprehension” thus included the GRADE measures Sentence Comprehension (SC), Paragraph Comprehension (PC), Listening Comprehension (LC), and Word Meaning (WM). The second factor, “First Grade Decoding”, also included those measures which at their face appeared to be intending to measure basic decoding ability including word-reading, letter-sound correspondence, and fluency with reading words within connected
Thus this factor included Word Reading (WR), Nonsense-Word Fluency (NWF), and Oral Reading Fluency (ORF). “First Grade Phonological Awareness” was represented by phoneme-segmentation fluency, unfortunately the only phonological measure available. Previous studies have found phonological abilities to constitute a separate ability area from both linguistic comprehension and decoding ability (Shaywitz, 2003; McBride-Chang & Manis, 1996; Wolf & Bowers, 1999; Stanovich & Stanovich, 1995; Torgeson & Wagner, 1995).

The model demonstrated adequate fit according to LISREL software goodness-of-fit statistics, including Normed Fit Index (NFI) score of 0.93, Comparative Fit Index (CFI) score of 0.94, Relative Fit Index (RFI) score of 0.89, and a Goodness-of-Fit Index (GFI) score of 0.88. However, two latent variables were highly correlated with each other at 0.97 (see Appendix 4). Linguistic Comprehension and Decoding were not functioning as independent factors so it was not justified to separate them. However, evidence from previous modeling studies strongly supports the existence of two separate factors at this age level (Singer & Crouse, 1981; Hoover & Gough, 1990; Floyd, Gregg & Keith, 2004; Willson & Rupley, 1997).

The decision regarding which measures might load most onto which latent factors was initially based on the abilities that the subtests purported to measure, thus all measures of “comprehension” were initially included into the Linguistic Comprehension factor. However, this grouping was not supported by the results of the factor analysis; rather, results suggested that the early reading measures were all highly correlated
together (except Phonemic-Segmentation Fluency which had already been considered a separate factor), with factor loadings ranging from 0.75 to 0.86, with the notable exception of Listening Comprehension with a loading of 0.37. Additionally, it was noted that Listening Comprehension possibly demonstrated low reliability and high error at 0.86. The most likely explanation for this measure correlating so poorly with the others has to do with the face validity of the other measures used, not necessarily that a linguistic factor or Listening Comprehension specifically could not theoretically be distinguished for this sample, given more pure measures of linguistic abilities. The measures purporting to measure aspects of reading and/or linguistic “comprehension” actually required test-takers to achieve a certain minimum level of decoding prior to assessing the child’s comprehension. SC and PC both required students to independently decode the words in the prompts in order to respond correctly to comprehension-related questions, so the measures could be functioning as measures of decoding, not comprehension, for this age/ability level, up to a certain threshold of minimum decoding ability. Therefore the decision was made to move these two measures to the Decoding factor.

A second measurement model was then run after re-arranging the measures to account for the issues described above. This time, “Linguistic Comprehension” was defined with only two measures, those that seemed most closely related to linguistic comprehension with minimal prerequisite decoding ability required. “Linguistic Comprehension” would now include only LC and WM. “Decoding” included WR, SC,
PC, ORF, and NWF. The Phonological factor and the fourth grade Reading Comprehension factors were unchanged. See Appendix 4. This second measurement model demonstrated slightly reduced fit, resulting in the following goodness-of-fit statistics provided by LISREL software: NFI = 0.92; CFI = 0.93; RFI = 0.87; and GFI = 0.86. NFI and CFI dropped 0.01 points and RFI and GFI dropped 0.02 points compared with the first measurement model. On the other hand, LC’s loading increased slightly from 0.37 to 0.40, and the correlation between the latent variables Decoding and Linguistic Comprehension decreased from 0.97 to 0.91. However, the issues were not completely resolved by making this small change-- LC’s loading of 0.40 still suggested that it is measuring something very distinct from what WM is measuring and that the two together could not be considered a single variable, and that the latent variables proposed were not sufficiently independent.

A third model tested attempted to again re-arrange the indicators in a different way in the hopes of revealing independent factors. See Appendix 4. This time, 5 total latent variables were identified. The Decoding factor contained NWF, ORF, and WR since these three all measure very basic decoding-level word-reading ability. A separate factor called “Comprehension” was identified incorporating WM, SC, and PC, the three measures which required decoding but based scores more on comprehension ability. The MCAS and Phonological factors were kept intact using their single indicators. LC, which demonstrated low correlation with all other measures, was separated and considered its own factor, now named more narrowly “Listening Comprehension”. Because no
reliability estimate for LC would be produced by running the model, it was necessary to estimate the reliability of the subtest based on previously published studies, as done with other single indicators. Results of reliability testing published by Pearson, Inc. were used to identify LC's internal reliability estimate of 0.78 (Williams, 2001). The same procedure for dealing with single indicators within CFA models was again used to specify an error estimate for LC to be included in the regression equation (Brown, 2006). The third measurement model demonstrated slightly improved fit over the second one (NFI=0.93; CFI=0.94; RFI=0.87; GFI=0.89). However, the two latent indicators Decoding and Comprehension were still correlated at 0.97, so they could not theoretically be expressed as independent variables despite the slightly improved fit of the overall model.

The results of these three exploratory factor analyses strongly suggested that all of the first grade measures included in the study, with the exception of LC and PSF, are highly dependent on decoding skills even though some purport to measure comprehension per se. WR, WM, SC, PC, ORF, and NWF are all highly correlated together and therefore must be considered a single latent variable at the first grade level. To account for the inclusion of measures including comprehension components, the name of the factor was changed to the broader “Early Reading” in place of “Decoding”. LC was not closely correlated with any other measure, so it had to be separated and considered its own independent factor, which was simply named “Listening Comprehension”.

102
The final measurement model reflected these findings (see Appendix 4). Goodness-of-fit statistics revealed identical model fit to the second measurement model. However, this one represented an improvement since it resolved the problem. LC now represented its own factor so that all of the latent variables were now functioning as independent factors. All other decoding-related first-grade measures loaded successfully onto a single latent variable.

3.4. Research Question 2: Results of Hybrid Model

Once having identified the best measurement model through factor analysis, a hybrid model was fit in which the fourth grade latent variables predicted fourth grade achievement on the MCAS Language Arts test. Phonemic Awareness (defined by PSF) and Listening Comprehension were not highly correlated with the other early reading measures or with fourth grade MCAS scores. Therefore they were hypothesized to affect early reading ability directly and later reading only indirectly through their effects on first-grade reading ability. This hypothesis was supported by the final hybrid model. The final model placed Phonemic Awareness and Listening Comprehension predicting the latent variable Early Reading, with Early Reading in turn predicting fourth grade Reading Comprehension represented as MCAS scores.
Figure 1: Hybrid Model: Predicting 4th grade advanced reading comprehension

All paths were statistically significant at .05 one-tailed analysis. 57% of the variance in fourth grade MCAS scores was explained by the model. Goodness of Fit statistics demonstrated adequate fit: NFI = .912; CFI = .924; RFI = .879.

3.5. Research Question 3: Correlations Between Measures and Group Differences

Appendix 2 lists the correlations found between each measure, including the predictive power of each first grade measure in isolation to explain the individual variation in fourth grade MCAS performance. The first-grade measure most predictive of fourth-grade MCAS ELA performance for the whole group was GRADE's Sentence
Comprehension ($r = 0.644$). This measure alone accounted for $41.47\%$ of the variance in the students' MCAS performance. Continuing on in order of predictive power demonstrated, Paragraph Comprehension came next ($r = 0.599$), accounting for $35.88\%$ of the variance in MCAS performance. DIBELS Oral Reading Fluency ($r = 0.581$) accounted for $33.75\%$ of the variance, GRADE Word Reading ($r = 0.551$) $30.36\%$, GRADE Word Meaning ($r = 0.536$) $28.73\%$, DIBELS Nonsense-Word Fluency ($r = 0.475$) $22.56\%$, GRADE Listening Comprehension ($r = 0.291$), and finally the least predictive measure was DIBELS Phoneme-Segmentation Fluency ($r = 0.242$) accounting for only $5.86\%$ of the variance in MCAS scores.

Appendix 3 lists results of statistical significance testing comparing correlation coefficients found for the ELL students versus native English-speakers. All of these discreet between-measure relationships were non-significant for differences based on language, with the exception of Word Reading versus Paragraph Comprehension, which was significantly more correlated for native English-speakers than for ELLs.

3.6. Research Question 4: Between-Group Comparison on Hybrid Model Fit

The next research question addressed was whether the predictive power of each measure and ultimately the latent variables differed between the two groups of subjects, divided according to home language. As described in Methods, MGCFA (using LISREL software) was used to conduct measurement invariance testing to determine whether group membership had statistically-significant effects on each predictive parameter estimated within the hybrid model.
Most importantly, no statistically significant interaction was found between group membership and the predictive coefficients between 1st Grade “Early Reading” and 4th Grade “Reading Comprehension”.

Additionally, no statistically significant interactions were found between group membership and the predictive coefficients describing the degree of relationships between Phonemic Awareness and 1st Grade Early Reading, nor between Listening Comprehension and 1st Grade Early Reading.

Most other parameters estimated within the model also demonstrated no statistically-significant differences based on group membership.

There were two exceptions in which statistically-significant differences based on group membership were found between the parameters estimating the loading of certain assessments onto the latent variable. These differences were found for the 1st grade Grade subtests Word Reading and Word Meaning, specifically their relationship with the latent variable 1st Grade Early Reading. On these two measures, English-speaking students' performances were more closely correlated with their other performances on the group of early literacy measures. Spanish-speaking students' performances on these two measures were more varied and less correlated with their early reading ability overall, as represented by the latent variable.

Additionally, the error estimate of one measure, Word Reading, resulting from the estimation of all the parameters, was different for students in the English-speaking group compared with the Spanish-speaking students, with Spanish-speakers demonstrating
more test error on this measure than English-speakers and this difference was statistically significant. In other words, more of Spanish-speakers' performance on this measure were unaccounted for by their overall early literacy performance.
CHAPTER 4

DISCUSSION

Holyoke Public Schools has particular features which make it a both a unique and a generalizable school system to examine. In the past few decades, the population of students who attend the City of Holyoke's public schools has become strikingly racially homogenous. Families of Puerto Rican descent make up the wide majority of those using the public schools, while the town's large Anglo population seems to have deserted them. While the student body cannot be described as racially “diverse”, this phenomenon of racially homogenous populations with majority racial “minority” students served has become the norm in many areas, in particular in urban public school systems. In many urban districts, most public schools students are black and/or Hispanic; in Holyoke, a small city, the population is overwhelmingly Hispanic and in particular heavily Puerto Rican.

This demographic feature makes Holyoke an important testing ground to answer specific questions not only about language learning per se, but about many issues that arise for a multi-generational community of “immigrants” from a place which is not really a foreign country but not really a state, who are simultaneously American and not always considered American, who speak Spanish sometimes exclusively, sometimes bilingually, and sometimes not at all, and who are part of a larger racial minority (“Hispanic”), as well. A different paper could be written examining identity issues and particularities of the Puerto Rican community of Holyoke, using anthropological or other
qualitative techniques. This paper did not attempt that task. Instead, the present investigation sought to examine empirical data in the form of students' tests scores on measures of early and late-elementary literacy tasks. However, each research question, even those asked about the entire sample prior to dividing it by home language, was asked in consideration of both the particularity and the potential generalizability of the students of Holyoke.

Since the nearly the whole sample was composed of Hispanic students, dividing them by home language provided a convenient set of controls for looking specifically at the first language of the child as a variable, since racial minority status was held constant. Often, studies comparing performances of language minority students with their peers are comparing mostly Hispanic ELLs to mostly Caucasian and/or African-American students, or a mixed-race control group and then attributing differences to ELL status per se. However, when this comparison is made there is a potential confound between second language learning and being Hispanic and the proportion of variance explained by students' Hispanic status versus their actual experience as speakers of another language remains unknown. The same holds true for socioeconomic status among other variables. The opportunity was ripe in this sample for controlling for both racial minority status and socioeconomic status and examining the effect of first language in isolation, particularly since first language status divided the sample nearly in half.

Additionally, the performance of the entire group, comprised of mostly Hispanic students from poor families with any number of variations in their linguistic
backgrounds, and the ability to compare these results to those of previous modeling studies about literacy acquisition conducted with other samples, are valuable opportunities as well.

On the other hand, as described in Chapter 1, language proficiency lies on a continuum and does not neatly divide into two distinct groups—those with and those without. Bilingualism is even more complex. When Puerto Rican parents in Holyoke registered their children for school and were asked on a registration form whether their children spoke English at home or Spanish, the answer may have not been very clear to many and yet they were forced to choose “English” or “Spanish”. An unknown number of children in the sample could have grown up speaking both English and Spanish in equal measure, yet “Both” was not an available option. Others may have grown up in bilingual households with one language more dominant, and those children at school entry age were perhaps dominant in one or the other but still had some command of both. Many likely could understand both languages but could speak only one, while others could have truly been monolingual, both receptively and expressively, in one language or the other. These differences and every variation in between were not only possible, but very likely, due to the complex nature of multi-lingual families and language acquisition in children growing up in bilingual or multi-lingual families. Taking a continuum variable and creating two distinct groups represents a psychometric limitation in terms of research methodology in which the goal is to maximize the differences between groups in order to be able to increase the chance of attributing statistical differences to the variable
in question, never to compare two groups with questionable distinction in respect to the variable. These issues are addressed in more detail in the Limitations section. However, the reality of language is as such-- any individual's degree of exposure, knowledge, proficiency, and skill in any particular language that is used in that persons' community is a complex matter not easily measured or quantified. In fact, from a sociolinguistic perspective, all or most of the children in the sample could be considered part of a “language minority” due to the likelihood that most lied somewhere on the bilingual continuum, and their particular brands of English formed and affected by their unique backgrounds was perhaps not Standard even when dominant in comparison with Spanish ability. Language proficiency and language dominance are certainly different concepts altogether.

The first result of examining the test performances of students in this sample of 232 was the finding that the group, compared with both national norms and, when applicable, state norms, did not do well on the tests administered in the first grade or on the state language arts exam that they took in the fourth grade. As described in the Results section, most of the students performed within the lowest category and were thus identified by test authors as “at-risk”, in the first grade. MCAS results are available on the state's website, including comparisons of all different locales, so the information about Holyoke students' comparatively low performance on the fourth grade MCAS assessments in 2007 and 2008 is not new. Holyoke Public Schools historically has been one of the lowest-performing in the state, and this ranking did not change over the time
period covered in the study.

What were less easily known were Holyoke students' aggregate performances on the first grade literacy screenings. Results of this investigation showed that the group as a whole, prior to considering first language, did poorly in the first grade on all measures of early literacy-related skills compared with students in the DIBELS and GRADE tests' national normative samples.

While the entire group performed below grade-level expectations across measures and grades, there were also statistically-significant differences found in their performances based on whether their parents had identified them as English-speaking or Spanish-speaking upon school entry. Spanish-speaking students performed even more poorly than their English-speaking peers on all of the first grade measures except for DIBELS Nonsense-Word Fluency and DIBELS Phoneme-Segmentation Fluency.

The Spanish group also performed more poorly than the English-speaking group on the fourth grade MCAS. Language proficiency gains and current ELL status were not taken into consideration; no students changed groups. This means that while the Spanish-speaking students may have become much more proficient in English, and an unknown proportion may have even been reclassified and no longer considered ELL students by the fourth grade, these students as a group continued to demonstrate lower performance on a test of English Language Arts in the fourth grade than their peers who had been identified as English-speakers upon school entry.

The results of testing for group differences among the scores indicated that only
Phoneme-Segmentation Fluency and Nonsense-Word Fluency were apparently immune to the effects of first language status in terms of student performance. These results suggest that first-grade Spanish-speaking students may perform more poorly on measures of first-grade literacy skills than Hispanic students of similar socioeconomic class (all students in the sample were eligible for free or reduced lunch) who speak English upon school entry, including DIBELS Oral Reading Fluency and GRADE Word Reading, Word Meaning, Sentence Comprehension, and Paragraph Comprehension. In other words, when practitioners are interested in determining whether a Spanish-speaking student is experiencing a specific learning disability or whether his or her relatively low performance can be attributed to second language learning, it would be prudent to compare his or her performance to normative samples comprised of Spanish-speaking students such as the present sample. Otherwise, low performance could be the result of language difference per se. The finding that the Spanish-speaking student's average fourth-grade MCAS performance was significantly impaired compared with that of the English-speaking students is particularly important information for practitioners to be aware of, because these students had already had four or more years of English immersion and surely demonstrated fluent English Basic Interpersonal Communication Skills (BICS). Practitioners sometimes assume that older students are no longer affected by language difference and expect grade-level performance by late elementary school. The present results challenge this assumption.

It is worth noting that the findings of multiple statistically-significant differences
in performance on the measures based on first language status is not akin to finding bias or unfairness inherent in the measure. Rather, these results simply show that the Spanish-speaking students as a group performed more poorly on average than English-speakers did on the measures. Predictive bias was addressed and will be discussed later.

The next set of research questions addressed by the present investigation had to do with the correlations among the individual measures, including those administered more or less concurrently, in the first grade, and the correlation between each first grade measure and that student’s score on the fourth grade MCAS ELA section. Generally speaking, and in support of previous concurrent validity research on DIBELS and GRADE measures, correlations between the individual measures of various early reading skills administered in the first grade (excluding PSF and LC) were moderate to high (r = 0.5 to 0.78; see Appendix 2). Not surprisingly, lower correlations were found between those measures more focused on comprehension and those more related to decoding skills. The lowest correlation reported was between Phoneme-Segmentation Fluency and Sentence Comprehension (r = .109 to .248). Although the comprehension-related measures were less correlated with decoding measures than more obviously similar measures, they were moderately correlated. Nonsense-Word Fluency for example was moderately correlated with Word Meaning (r = 0.56 to 0.527), although less correlated with Paragraph Comprehension (r = 0.514 to 0.305).

One important finding with various potential implications was that for this sample, Phoneme-Segmentation Fluency was only minimally correlated with all of the
other measures of early reading ability. For example, a correlation of only \( r = 0.272 \) was found between PSF and ORF, and a correlation of only 0.259 was found between PSF and Word Reading (WR). Its strongest relationship was with Nonsense-Word Fluency (NWF), which makes sense intuitively since both purport to measure a student's ability to correctly isolate and name a sound. Still, the correlation was relatively low (\( r = .401 \) to .484). NWF requires the student to apply knowledge of letter-sound correspondence as well as then articulating the correct sound, while PSF requires that the student listen to a word and then separate the phonemes.

This finding is in contrast to findings of previous studies of moderate correlations between measures of Phonological Awareness and other both concurrent and predictive reading and decoding tests. The National Early Literacy Panel in its meta-analyses report reviewing 69 studies measuring the predictive validity of Phonological Awareness tasks in terms of “later” (vaguely defined) reading ability reported an average predictive coefficient of \( r = 0.40 \) (NELP, 2008). Scarborough (2005) reviewed 61 longitudinal research samples and reported an average predictive coefficient for Phonological Awareness of \( r = 0.46 \), also predicting “later” reading ability. Both of these average findings were twice as large as the correlations between PSF and the other reading measures found in the current study, both those administered concurrently and fourth grade MCAS scores. In terms of the longitudinal relationship, the present study identified a very low predictive coefficient between first grade PSF scores and fourth grade MCAS scores (\( r = 0.242 \)) in contrast to the results of previous research just
described.

Several explanations could potentially account for this surprising finding. The most convincing of explanations is the possibility that PSF as a single narrow task does not adequately measuring a child’s broader Phonological Awareness ability, and that most of the previous studies reviewed in the large meta-analyses finding stronger relationships combined more phonological awareness measures, enabling a broader definition of phonological awareness to include more than just phonemic awareness, and more than a single task to measure the construct. PSF is really a measure of one small aspect of phonemic awareness per se, which is only one narrow aspect of phonological awareness, the most advanced manifestation of the construct, at that (Uhry, 2005; McBride-Chang, 2004). The present findings do not appear as surprising when examined in the light of more directly comparable findings of previous studies measuring the predictive validity of the DIBELS PSF measure per se. Goffreda and DiPerna’s (2010) review reported predictive coefficients for PSF predicting performance on standardized reading assessments ranging widely across different studies, from \( r = 0.15 \) to \( r = 0.93 \). This extreme range points to the likelihood of variation based on confounding constructs and calls into question the value of PSF alone as a valid measure of any single construct related to early reading skill. Likely confounds include an instructional variable since phonemic awareness tasks are known to be dependent on direct instruction in phonemic awareness (McBride-Chang, 2004), Students in the present sample actually performed normatively better on the PSF task than on most of the others. Another speculative
possible explanation is that teachers in Holyoke trained students on the specific PSF task. This training could have positive implications for students’ actual phonemic awareness and contribute to improved learning. However, it could also inflate the scores of the entire sample on PSF in particular, making vary less along with the other scores, and lose its predictive sensitivity.

Another explanation, or a concurrent explanation, could have to do with the demographic differences between the present sample and samples used in most previous research. The present sample was comprised of nearly all Hispanic children of low SES, with over half ELLs, while previous research samples were comprised of majority Caucasian students with varying inclusion of racial minorities and low-SES populations.

Yet another explanation has to do with age, both at the predictor variable and the outcome variable. Taken together, predictive studies and modeling studies generally have repeatedly reported widely ranging results and have attributed the lack of consistency to age differences of the samples, with different results reported for preschool, kindergarten, and first-grade measures and their ability to predict distinct outcome variables ranging from simple word-level reading abilities to advanced comprehension. In other words, there has been a problem of comparing apples with oranges which has plagued the field, raising serious doubts as to the validity of comparing results across studies without strictly controlling for age and the specific type of “reading” outcome variable used (Scarborough, 2005; NELP, 2008).

Regardless of the reason, which cannot be fully known without further
investigations isolating these potentially confounding variables, the finding does not support the predictive validity and thus utility of PSF as an indicator of concurrent or future reading ability for the present sample. This finding could be interpreted to add evidence supporting Scarborough's assertions that phonological awareness, while important, is not any more predictive or correlated with early reading skills than several other linguistic measures administered to young children (Scarborough, 2005).

A similar pattern of results was demonstrated for the GRADE measure Listening Comprehension (LC). Like PSF, this first grade task was only minimally related to any of the other measures. Its highest correlations were with Word Reading (r = 0.376) and with Word Meaning (r = 0.334) but these were still fairly low. Its lowest correlation was with PSF (r = 0.094). Upon first glance this could appear surprising, since both measure some aspect of phonological ability (listening). However, linguistic comprehension of connected speech, particularly in a sample made up of over half students with low English proficiency, is theoretically very different from the ability to discern the phonemes comprising single words, and previous research has shown that comprehension-related linguistic abilities are minimally related to early decoding-related skills such as phonemic awareness (NELP, 2008). In fact, previous research has supported a relatively weak relationship between many aspects of oral language ability and early decoding achievements, while also finding that linguistic abilities demonstrate greater effect on reading comprehension, an increasingly significant effect as students age and the outcome variable directly involves linguistic comprehension per se (Chall, 1990;
Hoover & Gough, 1990; Proctor et Al, 2005; Scarborough, 2005). Given these consistent findings from previous research, it is not at all surprising to find that a measure of listening comprehension would be only minimally related to students' basic phonemic awareness and phonics ability in the first grade.

According to the developmental view of the distinct stages of reading development (Chall, 1990), a linguistic comprehension measure should be more predictive of a student's reading comprehension ability than decoding ability. However, the present data did not support this relationship, either. Listening Comprehension was also only minimally correlated with the students' fourth grade MCAS scores at \( r = 0.291 \). But, considering the specificity of this group and previous research results regarding listening comprehension measures, this finding is not surprising. With over half of this sample identified as ELL upon school entry, a measure of listening comprehension would logically be for most of this sample highly based on their basic English language proficiency, potentially creating a strong floor effect as the ELL students were potentially unable to earn even a minimal score on the measure in the first grade. In fact, the ELL students in the sample did perform much worse than native English speakers on this measure. This effect would have minimized the predictive value of this measure for ELLs. Because their English proficiency continued to grow over the course of the four years leading up to taking the MCAS in the fourth grade, it would have been more valid or potentially interesting to determine whether their concurrent linguistic comprehension in the fourth grade had as much predictive value in estimating their MCAS performance
as had their decoding and reading fluency. Still, it is interesting that this single measure of linguistic comprehension in the first grade was not able to distinguish between students who did well and poorly on a high-stakes measure of reading comprehension in the fourth grade. This would suggest that measures of listening comprehension are not particularly useful in adding information about a student's present reading ability, or his or her potential reading comprehension ability later on.

This finding for LC would seem to contradict previous findings from longitudinal studies demonstrating a growing importance for oral language and many linguistic abilities in determining reading comprehension ability in the older grades (Scarborough, 2005). Scarborough's (2005) review, for example, examining only long-term longitudinal studies, found high average correlations between various linguistic abilities (measured in the early years) and advanced reading comprehension abilities among older students. However, these measured other very different types of linguistic ability including Expressive Vocabulary (average $r = .45$), Sentence/Story Recall (average $r = .45$), and Receptive Language (average $r = 0.37$). Arguably the aspects of linguistic ability which turned out to predict later reading comprehension are more complex language proficiency and processing abilities not well represented by a basic measure of listening comprehension. Tilstra et. al. (2009) found similar results for a measure of Listening Comprehension which was somewhat predictive of other measures of language ability in young children but then poorly-correlated to other measures of English oral language ability and reading comprehension in older children. For older children, they found
strong predictive relationships between vocabulary knowledge (both receptive and expressive) and advanced reading comprehension ability, and hypothesized that measures of listening comprehension measure BICS only, while CALP is better assessed using measures of vocabulary knowledge and is more closely related to reading comprehension in the late elementary and upper grades. They argued that while young children learn new vocabulary items and language skills from listening to oral language, older children learn new vocabulary and language more from written language and are tested more on their CALP ability. BICS becomes irrelevant to the linguistic knowledge required for advanced reading comprehension at some point in late elementary school, and a basic Listening Comprehension measure is measuring BICS, not CALP. This finding certainly supports this theory, proposed decades ago by Chall (1990) and supported by findings of longitudinal research on variables predicting advanced reading comprehension (Scarborough, 2005; Tilstra et. al., 2008; Proctor et. al., 2005).

The individual measures were then grouped in order to identify latent variables. Several patterns of inter-correlation were tested based on hypotheses derived from previous investigations, until one emerged demonstrating comparatively superior model fit. Most previous modeling studies examining relationships among the component skills of early literacy-related abilities have found that measures of linguistic comprehension and measures of early decoding form two separate factors, linguistic comprehension and decoding ability, both contributing to overall reading ability in different proportion depending on age and ability level (Hoover & Gough, 1990; Aaron et. al., 2008; Proctor
et. al., 2005). At first glance at the present results it would appear that they do not support these findings of previous CFA studies. Each time that the measures were divided up with those thought to measure “comprehension” separated from those measuring basic decoding skills, the two resulting latent variables were very highly correlated ($r=.97$), suggesting they were not measuring distinct abilities. However, the best explanation for the strong correlation between the comprehension-related measures and the decoding measures has to do with the construct validity of the measures themselves. Each measure purporting to measure an aspect of reading “comprehension” depended on a minimum level of decoding. Poor decoders, therefore, were not able to demonstrate their actual degree of linguistic comprehension. The measures were in fact more global measures of basic reading ability incorporating both basic decoding and basic literal comprehension simultaneously.

The exception was Listening Comprehension, the only measure which did not involve decoding at all. Students listened to a phrase or sentence and circled the picture depicting its meaning. This measure did not correlate with any of the others enough to load onto a latent variable and thus had to be considered its own separate factor.

Previous studies indicated that phonemic awareness predicts early reading ability but is not highly correlated enough to be considered part of the same basic skill set. The present study unfortunately included only a single measure purporting to measure phonemic awareness, Phoneme-Segmentation Fluency. As discussed, this measure was even less correlated with the other early reading measures than would be predicted by
previous findings. It, too, was therefore separated and considered its own distinct factor. Most of the early measures of literacy ability were highly correlated in individuals and best formed one single factor, which I called “Early Reading”. Listening Comprehension and Phoneme-Segmentation were best described as minimally-to-moderately correlated with Early Reading and thus considered separate factors. These results support the findings of previous models of early reading (Cutting & Denckla, 2001; Burke et. al., 2009; Speece et. al., 2003; Curtis, 1980; Singer & Crouse, 1981; Proctor et. al., 2005).

Once the measurement model with the best empirical fit was identified, various hybrid models were tested for best overall fit. The first attempt placed the three first-grade latent variables (Early Reading, Phonemic Awareness, and Listening Comprehension) with paths predicting fourth grade MCAS performance termed Fourth Grade Reading Comprehension, in order to initially rule in or out whether direct paths would be significant for all three latent variables to account for significant variance in the fourth grade variable. The paths from Phonemic Awareness and Listening Comprehension turned out to not be significant, and were removed. In other words, the first grade measures of Phonemic Awareness and Listening Comprehension had no direct predictive correlation with fourth grade Reading Comprehension. Previous models have also found no direct relationship for similar parameters (Willson & Rupley, 1997; Leppanen et. al., 2008).

Listening Comprehension and Phonemic Awareness were then positioned with paths predicting Early Reading, with a single path estimated from first grade Early
Reading to fourth grade Reading Comprehension. All paths were then significant, and the model demonstrated adequate fit.

These results can be interpreted as supporting the following general findings addressing the second research question: Early Reading ability in the first grade is a strong predictor of fourth grade Reading Comprehension ability for this population, accounting for over half of the variance in the outcome measure; first grade Phonemic Awareness is indirectly (but not directly) predictive of fourth grade Reading Comprehension only through its small but significant direct effect on first grade reading ability overall; and Listening Comprehension is also indirectly (but not directly) predictive of fourth grade Reading Comprehension through its small but significant direct effect on first grade reading ability overall.

These results are interesting primarily because of the uniqueness of the population studied. Previous models have tested similar relationships between measures and latent variables representing a wide variety of skills, abilities, attributes and status variables hypothesized to affect a students' acquisition of literacy skills and ultimately advanced reading comprehension and reached parsimonious conclusions, as previously reviewed. Very few, however, have tested these relationships on a sample of low-SES Hispanic students with over half self-identifying as Spanish-speakers upon school entry. The proportion of variance explained by the model, which in the present study only included observed measures of early literacy skills and did not include any status variables or ability measures, was in line with proportions of variance explaining future predictions of
reading comprehension using early reading measures in previous models with more diverse as well as more affluent samples. These results add evidence to the body of research studies finding differences in literacy performances based on ELL status, racial minority status and SES.

The next set of researched questions had to do with predictive bias between measures based on home language, in the context of the functioning of the entire hybrid model predicting fourth grade reading comprehension for the two groups. Little evidence of predictive bias by home language group membership was found upon examining differences in predictive coefficients for individual measures in isolation, as previously discussed. While all predictive coefficients between measures for the Spanish-speaker group were lower than those for the English-speakers, the differences were not statistically significant. Similarly, multiple group confirmatory factor analysis revealed few statistically-significant differences between the groups in terms of the fit of the hybrid model to account for their variance in performance on the fourth grade outcome measure. Only three paths yielded statistically-significant differences in estimates based on group membership: two involving the GRADE measure Word Reading and one involving the predictive coefficients loading Word Meaning onto the latent variable Early Reading. This means that for the Spanish-speakers, Word Reading and Word Meaning functioned differently than for the English-speakers. Both were less predictive for the Spanish-speakers than for the English-speakers of overall first grade reading ability. This result demonstrates predictive bias in these two measures based on the first language of
the student, and suggests that caution should be used when using these tests with English language learners, since their performance on them could have more to do with their relative language proficiency than their abilities in terms of early decoding and phonological abilities. The results generally support previous findings of no predictive bias based on ethnic group membership for DIBELS measures predicting reading comprehension (Hintze et. al., 2002), and support the use of most of the early measures with ELL students and with low-income Hispanic to accurately and fairly measure their acquisition of early literacy skills, and predict their performance on a high-stakes measure of advanced reading comprehension in the fourth grade without bias based on first language.

The study also provides evidence that low-income Hispanic students as a group struggle on English literacy-related measures in the first grade as well as on high-stakes reading comprehension measures in the fourth, compared with the national averages provided by the tests' normative samples, and that ELL status among them predicts even further deficits in terms of English literacy achievements at both ages. The most important implication for practitioners to take away is that which is least intuitive-- that ELL students continue to be disadvantaged on literacy-related tests even in the fourth grade, long after they appear to be fluent in English.

Unfortunately, the current study was unable to answer the question “why?”. Without an experimental paradigm isolating variables and comparing program types, curricula, teaching methods, home environment factors, or other potentially important
factors it would be impossible to determine a cause for the relatively low achievement
demonstrated across the entire sample in the present study. It is clear that students who
arrived speaking a language other than English at home continued to perform worse on
English literacy tasks in the fourth grade, compared with those who arrived with some
English proficiency. Previous research has suggested that a dual language program,
while not necessarily impacting all of the variables related to low achievement, could
mitigate the achievement gap apparently related specifically to low English CALP that
the fourth grade Spanish-speaking students exhibited as they struggled on the fourth
grade English language arts exam compared with peers who were also Hispanic, also
low-income, but who arrived at school equipped with a base in English on which to build
English CALP through their schooling experiences.
CHAPTER 5
LIMITATIONS

There are many limitations to consider when interpreting the results of the present investigation. Perhaps the greatest limitation involved the use of an extant data set and retrospective longitudinal analysis. Due to the limitations of the measures available, limited to those which the district had chosen to screen for reading difficulties in the first grade, the present study was not able to fully address the question—how much variance in later reading comprehension is explained by linguistic comprehension versus decoding ability? Instead, it was found that the “comprehension” measures were not measuring linguistic comprehension in isolation, but actually required a threshold of decoding skill prior to measuring comprehension of linguistic input. This problem left Listening Comprehension as the sole measure available to describe the first graders' language-related abilities. Previous research indicated that Listening Comprehension is a sorely insufficient measure of overall linguistic ability, but it was all that was available. Because the study was limited by use of extant data, more valid measures of linguistic comprehension which would have enabled this question to be more fully addressed were not included. Ideally these would have included measures of expressive and receptive vocabulary as well as any other measures of language proficiency administered to first graders.

The above limitation points to the drawbacks of using a school district's extant student data to analyze interrelationships among ability constructs, but it does not
necessarily mean that the measures themselves were inherently flawed in terms of their practical utility. While researchers would prefer to isolate each component skill or ability to the maximum extent possible in order to minimize confounds when attempting to quantify and describe the interrelationships among them, practitioners often have the opposite goal— to use tests which can be administered quickly and economically which encapsulate perhaps many related abilities into one task and thus to increase the potential predictive power of each measure. For example, in the present study the first-grade GRADE measure Word Meaning was fairly highly predictive of fourth grade MCAS scores for the sample as a whole, even though the measure was problematic from a research perspective in that it combined both decoding and linguistic (receptive vocabulary) skills into a single task, making it difficult to answer which was more important to the child's performance.

On the other hand, the use of the available test scores that the district had administered to its students to inform instruction uncovered several real shortcomings inherent in the district's assessment practices in terms of which tests were chosen, which were missing, and how much variance in later performance the district's chosen screening battery was able to account for. Missing were measures of both oral language ability and CALP for all students, which based on previous research findings are important predictors of future reading comprehension ability as well as useful for intervention purposes. According to the Component Model of Reading (Aaron et. al., 2008) which itself is based on a synthesis of modeling research, as well as the original more simplified
Simple View of Reading (SVR; Hoover & Gough, 1990), knowing whether a student's reading problems are related to a phonological deficit, a letter-sound correspondence knowledge deficit, a fluency deficit, or a linguistic deficit would lend itself to targeted remediation in the needed area. On the other hand, if screening measures are limited to only proxy measures such as oral reading fluency, the result is a high degree of predictive validity without any information regarding where to remediate. In other words, if a student is only administered ORF probes because these enjoy the most robust ability to predict their reading comprehension ability, we still do not know if his or her fluency deficit is a deficit in fluency per se, or in decoding which would affect their fluency rate in turn, or in vocabulary knowledge which would also affect their fluency rate. The screening battery administered to the general population of first graders in Holyoke included several measures of basic letter-sound correspondence, but only one fairly irrelevant linguistic measure, and only one very narrow measure of phonemic awareness. There were no measures of receptive or expressive vocabulary, CALP knowledge, or measures of other aspects of phonological awareness apart from phonemic awareness administered. Also, since the phonics measures included a fluency component, confounding phonics knowledge with general fluency and perhaps RAN ability and/or processing speed, they are be more useful for making normative comparisons between children, but less able to pinpoint the student's area of weakness even in terms of phonics ability. For this purpose, a more criterion-referenced phonics inventory would need to be additionally administered.
HPS does administer language proficiency measures to its ELL students and this information was not included in the present study because it was not consistently available for all of the ELL students in the sample, such that its inclusion would have reduced the size of the sample to an unacceptable level. Students' scores from the various subtests of the MELA-O and the MEPA would have been valuable additional data regarding the students' language abilities; however, even if it had been available for all of the ELL students, it is additionally problematic to not have any similar information for the native English speaking students. This again points to not only a psychometric limitation of the current study, but an implicit and incorrect assumption made by the district that only ELL students would benefit from assessments of language proficiency or linguistic abilities. Further, results from language proficiency testing is then only used to judge a student's progress towards attaining English language proficiency, a theoretically slippery goal which is itself linked to normative performance on reading comprehension and writing tasks since advanced reading comprehension is required to pass the MEPA. The implicit assumption is that students whose parents identify them as native speakers of English upon school entry do not have needs in the areas of linguistic abilities which will later affect their ability to handle more complex reading tasks, or that CALP ability is not relevant to this group. And further, the assumption is implied that ELL students' language proficiency and their reading ability are two separate skill areas. In fact, previous modeling research has consistently demonstrated that advanced reading comprehension in the older grades is increasingly based on linguistic comprehension for
all students, not just ELLs. By not measuring language ability, Holyoke Public Schools is ignoring half of the information needed to truly predict a student's future reading achievement as well as arm teachers to intervene in actual areas of deficit to prevent future reading problems.

Another major limitation of the current study was that the LISREL software would not properly run the configuration model syntax comparing the functioning of the hybrid model between the two groups. The reason for the crash remained unknown to the investigators, but points to potential problems with the distribution of the scores in each category between the groups. In order to compare the correlation parameters estimated within the hybrid model, a different statistical technique will need to be used such as multiple regression analysis. Further analysis will utilize alternative techniques in order to discover and account for any problems in the data and offer a more statistically valid assessment of group differences in terms of the final hybrid model.

The score distributions within each measure were extremely non-normal, representing another major weakness in terms of the validity of the predictive correlation coefficients produced.

Floor effects were noted for several first grade measures, which potentially negatively affected the coefficients (Catts et. al., 2009). This again points to the need for more sensitive early literacy measures.

Low power also was a major limitation of the current study. The statistical power of an analysis is affected by the size of the sample as well as the normality of the
distribution of the observed variables, or test scores. 232 students is not terribly small, but once divided between two groups, the small size of each group limited the ability of the analysis to find potential group differences beyond differences expected based on test error and the large variance of the scores in each group. So, there was a large potential and likelihood for type 2 errors-- or for differences based on group membership which existed in the population to be missed by the analysis due to the relatively small size of the sample taken. In other words, the mean predictive coefficients were different between groups with

The potential generalizability of results is also limited by the way that the sample was selected. To minimize potential selection bias, ideally a sample would be selected at random from the entire population of persons meeting the criteria of inclusion into the study. However, students were selected for inclusion based firstly upon their matriculation into Holyoke Public Schools, and secondly upon the availability of three sets of data for each: first grade GRADE scores, first grade DIBELS scores, and fourth grade MCAS scores. Score availability was limited; many students were not included because one of these three types of scores was missing in their district profile data. This again points to the limitations inherent in using extant data for analysis. On the other hand, it is unlikely that any specific identifiable characteristic determined which students had missing data, and even less likely that whatever caused the missing data was also relevant to reading ability, thus hopefully avoiding construct-irrelevant confounds based on a third factor, as is often the main limitation of correlational research as well as quasi-
experimental research which has been unable to achieve the ideal of random selection. It might be speculated that students for whom no fourth grade data was available were also those students whose families were more mobile, moving them out of the district in between first and fourth grades, and that that characteristic could be correlated with lower achievement and serve as a construct-irrelevant third factor explaining greater predictive validity of early measures for those students who remained in the district and thus were included in the study. However, this seems a bit of a stretch for an explanatory variable for predictive validity, and it is thought that although random selection was not achieved, inclusion into the study was fairly random and not based upon factors particularly relevant to the constructs being studied.
## APPENDIX A

### MEAN RAW SCORES AND GROUP DIFFERENCES

<table>
<thead>
<tr>
<th>Measure Significant difference?</th>
<th>Mean raw score</th>
<th>SD</th>
<th>Mean-English N= 107</th>
<th>Mean-Spanish N= 125</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsense-Word Fluency (NWF)</td>
<td>48.92</td>
<td>24.95</td>
<td>52.64</td>
<td>46.3</td>
<td>No P=0.535</td>
</tr>
<tr>
<td>Oral Reading Fluency (ORF)</td>
<td>24.24</td>
<td>21.76</td>
<td>27.58</td>
<td>21.77</td>
<td>Yes P=0.0429</td>
</tr>
<tr>
<td>Word Reading (WR)</td>
<td>15.37</td>
<td>3.64</td>
<td>15.97</td>
<td>14.94</td>
<td>Yes P=0.0273</td>
</tr>
<tr>
<td>Word Meaning (WM)</td>
<td>21.65</td>
<td>5.14</td>
<td>22.81</td>
<td>20.75</td>
<td>Yes P=0.0020</td>
</tr>
<tr>
<td>Sentence Comprehension (SC)</td>
<td>8.89</td>
<td>5.33</td>
<td>10.08</td>
<td>7.99</td>
<td>Yes P=0.0029</td>
</tr>
<tr>
<td>Paragraph Comprehension (PC)</td>
<td>9.89</td>
<td>5.03</td>
<td>11.26</td>
<td>8.74</td>
<td>Yes P=0.0001</td>
</tr>
<tr>
<td>MCAS English Language Arts Grade 4</td>
<td>34.38</td>
<td>11.98</td>
<td>37.44</td>
<td>32.12</td>
<td>Yes P=0.0006</td>
</tr>
<tr>
<td>Phoneme-segmentation fluency (PSF)</td>
<td>42.42</td>
<td>14.63</td>
<td>44.1</td>
<td>41.3</td>
<td>No P=0.1462</td>
</tr>
<tr>
<td>Listening Comprehension (LC)</td>
<td>14.93</td>
<td>1.88</td>
<td>15.42</td>
<td>14.49</td>
<td>Yes P=0.0001</td>
</tr>
</tbody>
</table>
**APPENDIX B**

**CORRELATION MATRIX**

<table>
<thead>
<tr>
<th></th>
<th>NWF</th>
<th>ORF</th>
<th>WR</th>
<th>WM</th>
<th>SC</th>
<th>PC</th>
<th>MCAS</th>
<th>PSF</th>
<th>LC</th>
</tr>
</thead>
<tbody>
<tr>
<td>NWF</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF</td>
<td>0.704</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WR</td>
<td>0.553</td>
<td>0.635</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WM</td>
<td>0.548</td>
<td>0.596</td>
<td>0.786</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>0.590</td>
<td>0.755</td>
<td>0.634</td>
<td>0.614</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>0.445</td>
<td>0.654</td>
<td>0.527</td>
<td>0.542</td>
<td>0.714</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCAS</td>
<td>0.475</td>
<td>0.581</td>
<td>0.551</td>
<td>0.536</td>
<td>0.644</td>
<td>0.599</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSF</td>
<td>0.447</td>
<td>0.272</td>
<td>0.259</td>
<td>0.263</td>
<td>0.195</td>
<td>0.069</td>
<td>0.242</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>LC</td>
<td>0.261</td>
<td>0.252</td>
<td>0.376</td>
<td>0.334</td>
<td>0.296</td>
<td>0.283</td>
<td>0.291</td>
<td>0.094</td>
<td>1.000</td>
</tr>
</tbody>
</table>
APPENDIX C

BETWEEN-GROUP COMPARISON OF CORRELATION COEFFICIENTS

<table>
<thead>
<tr>
<th></th>
<th>English group N = 107</th>
<th>Spanish group N = 125</th>
<th>Statistically significant difference?</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF vs. NWF</td>
<td>0.75</td>
<td>0.623</td>
<td>no</td>
</tr>
<tr>
<td>WR vs. NWF</td>
<td>0.63</td>
<td>0.491</td>
<td>no</td>
</tr>
<tr>
<td>WR vs. ORF</td>
<td>0.67</td>
<td>0.620</td>
<td>no</td>
</tr>
<tr>
<td>WM vs. ORF</td>
<td>0.54</td>
<td>0.647</td>
<td>no</td>
</tr>
<tr>
<td>WM vs. NWF</td>
<td>0.56</td>
<td>0.527</td>
<td>no</td>
</tr>
<tr>
<td>WR vs. WM</td>
<td>0.77</td>
<td>0.78</td>
<td>no</td>
</tr>
<tr>
<td>SC vs. NWF</td>
<td>0.62</td>
<td>0.522</td>
<td>no</td>
</tr>
<tr>
<td>SC vs. ORF</td>
<td>0.77</td>
<td>0.725</td>
<td>no</td>
</tr>
<tr>
<td>WR vs. SC</td>
<td>0.72</td>
<td>0.58</td>
<td>no</td>
</tr>
<tr>
<td>SC vs. WM</td>
<td>0.58</td>
<td>0.628</td>
<td>no</td>
</tr>
<tr>
<td>PC vs. NWF</td>
<td>0.514</td>
<td>0.305</td>
<td>no</td>
</tr>
<tr>
<td>PC vs. ORF</td>
<td>0.672</td>
<td>0.611</td>
<td>no</td>
</tr>
<tr>
<td><strong>WR vs. PC</strong></td>
<td>0.64</td>
<td><strong>0.432</strong></td>
<td>yes</td>
</tr>
<tr>
<td>PC vs. WM</td>
<td>0.544</td>
<td>0.519</td>
<td>no</td>
</tr>
<tr>
<td>PC vs. SC</td>
<td>0.732</td>
<td>0.664</td>
<td>no</td>
</tr>
<tr>
<td>MC vs. NWF</td>
<td>0.472</td>
<td>0.443</td>
<td>no</td>
</tr>
<tr>
<td>MC vs. ORF</td>
<td>0.605</td>
<td>0.537</td>
<td>no</td>
</tr>
<tr>
<td>WR vs. MC</td>
<td>0.65</td>
<td>0.475</td>
<td>no</td>
</tr>
<tr>
<td>MC vs. WM</td>
<td>0.524</td>
<td>0.515</td>
<td>no</td>
</tr>
<tr>
<td>MC vs. SC</td>
<td>0.685</td>
<td>0.568</td>
<td>no</td>
</tr>
<tr>
<td>MC vs. PC</td>
<td>0.603</td>
<td>0.561</td>
<td>no</td>
</tr>
<tr>
<td>Comparison</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Result</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>PSF vs. NWF</td>
<td>0.484</td>
<td>0.401</td>
<td>no</td>
</tr>
<tr>
<td>PSF vs. ORF</td>
<td>0.302</td>
<td>0.217</td>
<td>no</td>
</tr>
<tr>
<td>WR vs. PSF</td>
<td>0.34</td>
<td>0.167</td>
<td>no</td>
</tr>
<tr>
<td>PSF vs. WM</td>
<td>0.344</td>
<td>0.166</td>
<td>no</td>
</tr>
<tr>
<td>PSF vs. SC</td>
<td>0.248</td>
<td>0.109</td>
<td>no</td>
</tr>
<tr>
<td>PSF vs. PC</td>
<td>0.084</td>
<td>0.0001</td>
<td>no</td>
</tr>
<tr>
<td>PSF vs. MC</td>
<td>0.169</td>
<td>0.257</td>
<td>no</td>
</tr>
<tr>
<td>LC vs. NWF</td>
<td>0.36</td>
<td>0.23</td>
<td>no</td>
</tr>
<tr>
<td>LC vs. ORF</td>
<td>0.14</td>
<td>0.21</td>
<td>no</td>
</tr>
<tr>
<td>LC vs. WR</td>
<td>0.44</td>
<td>0.35</td>
<td>no</td>
</tr>
<tr>
<td>LC vs. WM</td>
<td>0.37</td>
<td>0.29</td>
<td>no</td>
</tr>
<tr>
<td>LC vs. SC</td>
<td>0.35</td>
<td>0.25</td>
<td>no</td>
</tr>
<tr>
<td>LC vs. PC</td>
<td>0.31</td>
<td>0.23</td>
<td>no</td>
</tr>
<tr>
<td>LC vs. MC</td>
<td>0.36</td>
<td>0.23</td>
<td>no</td>
</tr>
<tr>
<td>LC vs. PSF</td>
<td>0.14</td>
<td>0.05</td>
<td>no</td>
</tr>
</tbody>
</table>
APPENDIX D

MEASUREMENT MODELS

Measurement Model 1:

Chi-Square=141.39, df=23, P-value=0.00000, RMSEA=0.148
Measurement Model 2:

Chi-Square=175.34, df=23, P-value=0.00000, RMSEA=0.168
Measurement model 3:

Chi-Square=129.46, df=20, P-value=0.00000, RMSEA=0.153
Measurement model 4 (Final):

Chi-Square=172.13, df=24, P-value=0.00000, RMSEA=0.162
BIBLIOGRAPHY


Abella, R., Urrutia, J. & Shneyderman, A. An Examination of the Validity of English Language Achievement Test Scores in a LEP Student Population. Paper presented at the annual meeting of the National Association for Bilingual Education, February 1, 2003, New Orleans, LA.


Garcia & Cuellar (2006) Who are these linguistically and culturally diverse students? Teachers College Record, 108(11), p.2220-2246


154


