THE RELATIONSHIP OF LANGUAGE PROFICIENCY, GENERAL INTELLIGENCE, AND READING ACHIEVEMENT WITH A SAMPLE OF LOW PERFORMING, LIMITED ENGLISH PROFICIENT STUDENTS

A Dissertation

by

CHARLOTTE KENNEDY JONES

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2006

Major Subject: School Psychology

THE RELATIONSHIP OF LANGUAGE PROFICIENCY, GENERAL INTELLIGENCE, AND READING ACHIEVEMENT WITH A SAMPLE OF LOW PERFORMING, LIMITED ENGLISH PROFICIENT STUDENTS

A Dissertation

by

CHARLOTTE KENNEDY JONES

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Michael J. Ash
Committee Members,	Salvador Hector Ochoa
	Jan N. Hughes
	Luana J. Zellner
Head of Department,	Michael R. Benz

December 2006

Major Subject: School Psychology

ABSTRACT

The Relationship of Language Proficiency, General Intelligence, and Reading Achievement with a Sample of Low Performing, Limited English Proficient Students.

(December 2006)

Charlotte Kennedy Jones, B.S., University of Georgia

Chair of Advisory Committee: Dr. Michael J. Ash

The present study had three purposes. The first was to examine the score reliability of instruments purported to be appropriate in the assessment of students with limited English proficiency (LEP). The second was to investigate the criterion-related validity of the Universal Nonverbal Intelligence Test (UNIT) with a sample of low-performing, Hispanic students. The third purpose was to explore the contribution of language proficiency in the dominant language (L1) and the language proficiency in the subordinate language (L2) in the prediction of reading achievement in L1. Participants included first and third grade students of Hispanic origin who scored below the median for their district on a state-approved, district-administered measure of literacy in first grade.

Satisfactory internal consistency estimates were achieved with a sample of LEP students (n=24) on the UNIT, the Woodcock-Muñoz Language Survey (WMLS) in English and in Spanish, and the Batería Woodcock-Muñoz: Pruebas de Approvechamiento-Revisada (Batería-R APR). For first grade students, scores from the UNIT demonstrated satisfactory concurrent validity with those from the Woodcock-

Johnson III: Tests of Achievement (WJ-III ACH) for a sample of Hispanic, non-LEP students (n=89). However, the concurrent validity of the UNIT was not upheld for a sample of Hispanic, LEP students administered the Batería-R APR (n=56). Regarding predictive validity, results from simple linear regression analyses suggested that performance on the UNIT in first grade accounted for a negligible portion of the variance on the Texas high-stakes reading test in third grade for a group of LEP students (n=51) as well as for a group of non-LEP students (n=77). Language proficiency in L1 emerged as a positive predictor of reading achievement in L1. However, language proficiency in L2 was not shown to be a statistically significant, independent contributor to this relationship with reading achievement on the Batería-R APR (n=79), WJ-III ACH (n=14), TAKS Spanish (n=54), or TAKS English (n=12). Findings are discussed with respect to the restriction of range due to selection criterion and sample size, the use of the Abbreviated battery of the UNIT in the prediction of reading achievement, and the contribution of language proficiency in L2 for low performing, LEP students in the third grade.

DEDICATION

To my mama. You are the wind beneath my wings.

ACKNOWLEDGMENTS

My experiences with the faculty and staff of Texas A&M University have undoubtedly shaped me both professionally and personally. I have grown so much since my arrival at A&M just five years ago. From the very beginning, Dr. Ash, you have been there to lend your support, wisdom, and wit. In your own caring and nurturing way, you have guided me in the discovery of a fresh perspective time after time. No matter the state in which I arrive, I always leave your presence with a smile and thoughts to ponder further. Dr. Ochoa, your passion is contagious. I would not have pursued this topic if I had not had you as a professor and mentor. Further, this study would not have been attainable without the assistance of Dr. Hughes and the Project Achieve team. Dr. Hughes, your drive for precision has led me to further hone my skills as a researcher and a practitioner, and my time with Project Achieve has been especially influential toward my professional and personal development. I also want to sincerely thank you, Dr. Zellner, for your willingness to serve on my committee. Your time and insight is greatly appreciated. A special, heartfelt thank you also goes to "Ms. Carol" and Kathy May. I am forever indebted to your dedication, kindness, support, and friendship. You have seen me through my graduate career.

To my dear friends and colleagues, Rachel and Crystal, your friendship is deeply cherished and the times we have shared will always be fondly remembered. Rachel, you have been my best friend since day one at A&M. We *are* two peas in a pod, and I am so incredibly blessed to have your everlasting friendship. I hope that one of these days we will once again work (and play) side-by-side. Crystal, you have definitely kept the

promise you made during our orientation to always be there for me! Having a fellow South Carolinian in Texas helped me to feel all that much more at home. I would not have made it through without the love and friendship that both of you have bestowed upon me. I look forward to the future memories that we will share.

To my loving family, you have rooted for me as long as I can remember. You have instilled upon me the importance of family, faith, education, and laughter. You have taught me to be myself and to take pride in everything that I do. I am who I am today because of you. I can only hope and pray that I can provide my future family with the same kind of unconditional love and support that you have afforded me.

And, to my devoted husband Jason, you have been there for me like no other. Words cannot begin to express what you mean to me or how blessed I am to be able to spend the rest of my life by your side. With you, I am at home. I love you with all my heart, always and forever.

TABLE OF CONTENTS

Page

Page

CHAPTER

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

The population within the public school systems across the United States is increasingly culturally and linguistically diverse (Kindler, 2002). And, many of these diverse students are disproportionately represented in special education (Artiles et al., 2002, 2005; Rhodes et al., 2005). Although attributed to several factors, the assessment practices and measures employed in determining eligibility are considerable, influential factors in the disproportionate representation of these students and, therefore, are of primary interest in the present study.

To foster appropriate assessment practices, several ethical guidelines and professional standards highlight critical factors that must be considered when working with limited English proficient (LEP) students. Such specifications include that the linguistic demands during testing be kept to a minimal in consideration of the purpose for testing, the examinee's relative language proficiencies should be determined, and, subsequently, the test generally should be administered in the examinee's most proficient language (AERA, APA, NCME, 1999). Furthermore, appropriate assessment and/or treatment procedures, techniques, and strategies should be conducted and include the consideration of the reliability and validity of selected measures (APA, 2002; NASP, 2000).

Several case and federal laws also have been enacted to address appropriate

---

This thesis follows the style of *Journal of School Psychology*.

standards when working with LEP students. For instance, *Diana v. State Board of Education* (1970) affirmed that assessment must be conducted in the child's native language whenever possible or not in English. Public Law 94-142 (Education for All Handicapped Children Act of 1975) upheld and enhanced *Diana* by concluding that assessment of LEP students must be in the child's primary language as well as in a nondiscriminatory manner. All students regardless of disability or culture were also afforded a free and appropriate public education in Public Law 94-142.

Statement of the Problem

Despite the rising likelihood that school psychologists will encounter students from diverse backgrounds, the current ethical guidelines and professional standards, as well as relevant case and federal laws, school psychologists have reported that they feel less than adequately trained in the area of assessment with linguistically diverse students (Ochoa, Rivera, & Ford, 1997), and students from culturally and linguistically diverse backgrounds continue to be disproportionately represented within the special education population (Artiles et al., 2002, 2005; Gersten & Woodward, 1994; Rhodes et al., 2005). Therefore, the assessment practices employed by school psychologists continue to be of concern with regard to the representation of LEP students in special education.

Cognitive Ability

With regard to the assessment of the cognitive functioning of LEP students, some of the most popular measures of intelligence used with LEP students include the Wechsler scales, Bender Visual-Motor, Gestalt test, Draw-A-Person test, and the Leiter (Ochoa et al., 1996b). However, limitations of these measures with diverse populations

include limited psychometric properties, incomprehensive results, and cultural loading (Rhodes et al., 2005). Furthermore, intelligence tests in English are not appropriate for students with LEP because they are not typically included in the standardization sample (Barona & Santos de Barona, 1987).

If the examiner is not bilingual and if tests are not available in the appropriate language, alternate, less ideal avenues (e.g., use of an interpreter, use of a nonverbal test) may be pursued. However, research has identified difficulties with regard to the use of interpreters (Rhodes, Ochoa, Ortiz, 2005). Therefore, a nonverbal test may be the most appropriate method employed in the assessment of students with LEP, especially when students have not achieved CALP in either language. Ochoa (2003) noted that limitations of typical nonverbal instruments include the limited predictive validity, the lack of a comprehensive measure of intelligence, and the underrepresentation of linguistically-diverse children included in the standardization samples.

Of the comprehensive nonverbal measures of general intelligence currently available, the Universal Nonverbal Intelligence Test (UNIT) appears to minimize the limitations associated with other nonverbal tests (Fives & Flanagan, 2002). This nonverbal measure of general intelligence was designed especially for children who may be disadvantaged by language-loaded measures. With regard to test fairness, Bracken and Naglieri (2003) reported that the UNIT is a fair and non-discriminative instrument given its completely nonverbal format, its adequate psychometric properties, the limited influence of culture, and the use of a "comprehensive and inclusive 'sensitivity' panel" to evaluate all items (p. 259). Given the steps taken to ensure fairness as well as the

sound psychometric properties, the UNIT appears to be appropriate with the population of interest in this proposed study. However, its relation to academic achievement with a sample of Hispanic, LEP students remains to be validated by external studies.

Language Proficiency

An individual's proficiency in a particular language should be considered on a continuum and is "the degree of control one has over the language in question" (Hamayan & Damico, 1991, p. 41). To determine language proficiency, students' basic interpersonal communicative skills (BICS) and cognitive academic language proficiency skills (CALP) should be evaluated. While BICS are applied in social, informal language and take typically two to three years to acquire, CALP includes skills needed to succeed in the classroom and require a minimum of five to seven years to obtain in a second language (Cummins, 1984).

The level of CALP in the native language and the second language should be evaluated to appropriately inform the mode of assessment and to accurately interpret results. However, research is limited concerning the relationship among language proficiency in both languages and academic achievement. Consequently, Ochoa (2003) called for studies to "examine the role of CALP level in both the native language and in English" (p. 579). Ochoa further encouraged this research to be completed "across the following methods of assessment used with bilingual students: (1) nonverbal measures, (2) testing in English, (3) testing in a student's native language, and (4) testing in both English and the native language simultaneously" (p. 579). The present study will examine the relation of language proficiency in the dominant language (L1) and the

subordinate language (L2) with academic achievement in reading. Furthermore, given its distinctiveness as a specific measure of CALP, the Woodcock-Muñoz Language Survey (WMLS; Woodcock & Muñoz, 1993/2001) will be employed as a measure of language proficiency in Spanish and in English.

Academic Achievement

In addition to considering the student's previous educational and environmental experiences, the student's measured language proficiency in both languages further informs the examiner of the appropriate language in which to assess academic achievement (Rhodes et al., 2005). The examiner has several comparable options when the child is determined to be fluent in English only; however, only a few measures are available to assess academic achievement when a child is determined to be predominantly fluent in Spanish. One norm-referenced, comprehensive measure of academic achievement available in Spanish is the Batería Woodcock-Muñoz – Revisada (Batería-R APR; Woodcock & Muñoz-Sandoval, 1996). Although one of the more commonly employed measures of achievement in Spanish, little is known about the psychometric properties of the Batería-R APR with LEP students residing in the United States or how performance on the Batería-R APR correlates with language proficiency.

In addition to norm-referenced measures of academic achievement, the passage of acts such as No Child Left Behind Act of 2001 (NCLB; Pub. L. No. 107-110), requires culturally and linguistically diverse students of certain grades to pass high-stakes state assessments of achievement in order to be promoted to the subsequent grade. The Texas Assessment of Knowledge and Skills (TAKS) serves as an example of a

state-administered, criterion-referenced test that attempted to account for the LEP population within the state of Texas. Specifically, the TAKS may be administered in Spanish or in English to LEP students based upon the recommendation of the language proficiency assessment committee (LPAC). Nevertheless, performance on the TAKS has been rarely studied with regard to language proficiency.

Purpose of Study

As discussed, schools within the United States are experiencing an exponential growth in the number of culturally and linguistically diverse students. To adhere to the legal and ethical guidelines, to inform training, and to minimize disproportionate representation of these students in special education, research with regard to appropriate testing instruments and practices is essential. For one, the dearth of reliability studies beyond what is employed with the standardization sample necessitates external studies of the psychometric properties of commonly employed measures of language proficiency (i.e., WMLS), general intelligence (i.e., UNIT), and academic achievement (i.e., Batería-R APR) with linguistically diverse students. Similarly, the criterion-related validity of the UNIT with measures of achievement for LEP versus non-LEP students has yet to be studied. Furthermore, given the recent, yet promising proposition to consider language proficiency in L1 and L2 in the selection of assessment modality with tests of academic achievement must be examined among linguistically diverse students.

Accordingly, the following are the specific objectives of the present research:

1. With the current sample of LEP students, estimate the internal consistency of the scores obtained via the:

a. WMLS-English
b. WMLS-Spanish
c. UNIT: Abbreviated
d. Batería-R APR: Broad Reading composite (Batería-R APR: Reading).

2. Examine the criterion-related validity of the UNIT with a sample of LEP students as well as with a sample of non-LEP students. The following will be specifically investigated:
    a. Estimate the concurrent validity of the UNIT based upon performance on a norm-referenced measure of reading achievement in Spanish (i.e., Batería-R APR) for LEP students and in English (i.e., WJ-III ACH) for non-LEP students.
    b. Estimate the predictive validity of the UNIT based upon performance on a state-administered, large-scale assessment of reading achievement (i.e., TAKS: Reading) administered in the dominant language.
3. Explore the role of language proficiency in L1 and language proficiency in L2 in the prediction of reading achievement in L1. Reading achievement will be assessed via norm-referenced and criterion-referenced measures of achievement, each administered in the child's dominant language.

Research Hypotheses and Implications

Based upon previous research and current recommended practices, the following hypotheses were proposed:

1. Considering that the standardization sample of the UNIT, WMLS, and Batería-R APR purport the inclusion of linguistically diverse students and that these measures are designed for use with linguistically diverse students, the internal consistency of the scores obtained via these instruments is hypothesized to be satisfactory (i.e., $r>.80$).
2. Based upon the premise that the UNIT is a nonverbal measure of general intelligence, the concurrent and predictive validity of the UNIT for students of limited English proficiency and for students who are not considered to be of limited English proficiency is hypothesized to be statistically significant ($p<.05$) and similar for each group.
3. It is further assumed that language proficiency in L1 and L2 will collectively predict reading achievement in the dominant language better than language proficiency in L1 alone.

As advised by ethical guidelines, the findings from this study will inform school psychologists of the reliability and validity of recommended instruments for low performing, LEP students. This information will lend to the literature regarding the appropriateness of these instruments with students of similar backgrounds. Furthermore, a better understanding will be gleaned of the relationship between proficiency in two languages in the prediction of reading achievement of low performing LEP students. These results will be particularly useful given that analyses will be conducted via performance on one of the few norm-referenced achievement instruments in Spanish as well as a high-stakes, criterion-referenced instrument available in Spanish and in

English. Furthermore, results from these analyses will inform practitioners of the importance of measuring language proficiency in both languages to aid in test interpretation. Similarly, data will demonstrate the relation of language proficiency in two languages with reading achievement in the dominant language.

CHAPTER II

LITERATURE REVIEW

The assessment of school-aged children referred for academic difficulties has undoubtedly played a central role in the field of school psychology since its infancy in the late 19th and early 20th centuries (Benjamin & Baker, 2004; Braden, Di-Marino-Linnen, & Good, 2001). During this period, psychological testing became extensively used to study individual differences in learning, memory, perception, feeling, and thinking. James McKeen Cattell devoted much of his research to "mental testing" and encouraged the use of such testing in schools to determine academic potential. Although his tests were short-lived given their little to no correlation with academic achievement, his contributions promoted the use of mental testing to identify individual differences of school-aged children (Benjamin & Baker, 2004).

While Cattell was advocating the use of mental tests in America, Alfred Binet and Theophilius Simon of France developed in 1905 the "first technological breakthrough in intelligence testing by developing the first practical intelligence test battery" (Kamphaus, 2001, p. 7). Henry Herbert Goddard subsequently published an English translation of the Binet test in 1908 for its use in America. Unlike Cattell's mental tests, the Binet measures were highly correlated with academic performance and became particularly employed in the differentiation of "morons" (individuals who appear to be of average ability but have borderline intelligence) from "normals" (Benjamin & Baker, 2004).

Lewis Terman, also interested in individual abilities, further revised the Binet scales and introduced the Stanford-Binet in 1916. Given its paramount psychometric properties and norming standards, this measure became the most successful English translation of the original Binet scales (Kamphaus, 2001), and it and its successors would remain the prominent tests of intelligence for the next forty years (Benjamin & Baker, 2004). The impact of the Stanford-Binet tests on the role of assessment in school psychology remains even today (Braden et al., 2001).

In addition to interests in individual difference through child study, the role of the school psychologist has been shaped by the implications of public concerns, regulations, and laws. For example, the child saving movement, instituted to protect children from societal exploitation, propelled child labor laws and compulsory schooling (Benjamin & Baker, 2004). Laws restricting the use of children in industry as well as compulsory school attendance laws enacted and enforced between 1890 and 1930 spawned an exponential increase in the presence of children from an array of backgrounds, including culturally and linguistically diverse, within school systems (Braden et al., 2001; Fagan, 1992). Furthermore, these laws generated the presence of a growing number of diverse children who did not succeed within regular education, which, in turn, instigated the need to evaluate and enroll these "exceptional" children into special classes (Benjamin & Baker, 2004; Braden et al., 2001; Fagan, 1992). As the number of special classes increased, the need for site-based psychological services also increased. Hence, the primary purpose of psychologists became the "sorting" of

children into segregated educational placements based upon measured ability (Benjamin & Baker, 2004; Fagan, 1992).

Despite the growing number of diverse children within schools and the widespread use of mental measures to determine placement of children with special needs, the assessment of bilingual, Spanish-speaking children went relatively unstudied until the 1920s. Of these initial studies, George I. Sánchez (1932) documented 18 manuscripts published between 1922 and 1931 that examined group differences on test results among English-speaking children and Spanish-speaking children. While many of these studies attributed or implied that the low performance of Spanish-speaking children was due to genetic "inferiority," Sánchez was one of the first to affirm that typically executed standardized testing instruments did not provide a valid estimate of bilingual students' abilities because, as he indicated, the inferiority did not reside in the Spanish-speaking children but in the measures employed. He further asserted heredity alone could not account for performance differences but that genetic, environmental (e.g., culture, school experience, socioeconomic status), *and* linguistic factors must be addressed to accurately interpret test results.

Two years later, Sánchez (1934) delved further into issues regarding the use of customary testing practices with bilingual students. He called to question issues of norming and standardization by declaring that the inferiority "determined" from the application of standard intelligence tests with Spanish-speaking children was erroneous because samples of this population were not included in the norming process. He observed, "A test is valid only to the extent that the items of the test are as common to

each child tested as they were to the children upon whom the norms were based" (Sánchez, 1934, p. 766). Sánchez (1934) suggested the need to account for environmental and linguistic differences of the standardization sample throughout the norming process and asserted that an IQ score is only meaningful "to the extent that the past history of the child has been assayed by the test in equal manner, with equal justice, and in equal terms as were the past histories of the children used as the criteria for the test" (p. 767).

Similarly, Sánchez was one of the first to highlight the grave errors of translating a test without establishing proper norms or a proper linguistic context. As stated by Sánchez (1934), "The whole question is that of whether or not the revised test was the same test as the original in terms of difficulty, suitability, validity, reliability, etc." (p. 768). Nevertheless, his battle to develop appropriate standardization and norming practices for children from culturally and linguistically diverse backgrounds continues to exist after more than 70 years. And, although some have credited his work as an impetus for change with respect to the application of testing of bilingual children (Valencia & Suzuki, 2001), Sánchez, indefensibly, is not widely recognized by many in the field of psychological testing.

Laws and Ethics

Although difficulties of assessing children of foreign language backgrounds were referenced in the 1940s to 1960s (McLean, 1995), it was not until legislative decisions were made in the 1970s that more emphasis was placed on the assessment practices employed with culturally and linguistically diverse children. Consequently, legislative

statutes and regulations as well as case laws once again heavily influenced the role of the school psychologist. Two consent decrees that set the stage for the changes in the assessment of culturally and linguistically diverse students were *Diana v. State Board of Education* (1970) and *Guadalupe Organization, Inc. v. Tempe Elementary School District* (1972). In the class action suit *Diana*, nine Mexican-American children were identified as mentally retarded on the basis of their scores on a language-loaded measure of intelligence. One of the plaintiffs, Diana, obtained an IQ score 49 points higher than originally reported when later tested in Spanish and English by a bilingual examiner (Jacob & Hartshorne, 2003). *Guadalupe* was a similar class action filed on behalf of students of Mexican-American or Yaqui Indian origin that pursued the need for bilingual-bicultural education for non-English speaking students (Jacob & Hartshorne, 2003; Rhodes et al., 2005). Of particular importance to school psychology, these decrees ultimately mandated that the assessment of second language learners be conducted in their primary language or via nonverbal measures. The *Guadalupe* consent decree further stipulated that measures of adaptive behavior must be used in conjunction with proper measures of intelligence to identify a child with mild mental retardation (Jacob & Hartshorne, 2003; Rhodes et al., 2005).

Public Law 94-142 (*Education for All Handicapped Children Act,* 1975) espoused and enhanced the aforementioned consent decrees. In addition to requiring that students of limited English proficiency be assessed in their primary language (as stated in *Diana* and *Guadalupe*), this federal legislative law afforded all children with disabilities, regardless of ethnic, cultural, or linguistic background, the right to a

nondiscriminatory evaluation, to a free and appropriate education (FAPE), to an education in the least restrictive environment (LRE), and to the establishment and implementation of an individualized education program (IEP; Rhodes et al., 2005, p. 46). In 1990, the regulations of Public Law 94-142 were adopted into the Individuals with Disabilities Education Act (IDEA) and the term "disability" was substituted for "handicap" (Jacob & Hartshorne, 2003). In 1997, Congress ratified IDEA 1990 and indicated that lack of instruction or limited English proficiency cannot be a determinant factor in qualification for special education (Rhodes et al., 2005). The term "limited English proficiency" (LEP) was previously defined by the Improving America's Schools Act in 1994 as someone

> (A) who--
>
> (i) was not born in the United States or whose native language is a language other than English and comes from an environment where a language other than English is dominant; or
>
> (ii) is a Native American or Alaska Native or who is a native resident of the outlying areas and comes from an environment where a language other than English has had a significant impact on such individual's level of English language proficiency; or
>
> (iii) is migratory and whose native language is other than English and comes from an environment where a language other than English is dominant;

> (B) who has sufficient difficulty speaking, reading, writing, or understanding the English language and whose difficulties may deny such individual the opportunity to learn successfully in classrooms where the language of instruction is English or to participate fully in our society. (sec. 7501)

IDEA 1997 further necessitated that LEP students must be administered tests in their native language or another mode of communication as necessary (Jacob & Hartshorne, 2003). In the assessment of learning disabilities, the exclusionary clause included in IDEA 1997 further required the consideration of environmental, cultural, and economic factors (Rhodes et al., 2005).

The American Psychological Association (APA) and the National Association of School Psychologists (NASP) are two professional organizations to which school psychologists adhere to their established ethical principles and professional standards. Of these principles and standards, several address and complement the aforementioned legislative actions to favorably influence the assessment practices employed with children. With specific regard to the assessment of linguistically diverse students, Standard 9.02 of APA's most recent ethics code advises psychologists to take into account the client's language proficiency as well as the reliability and validity of measures to determine the appropriate assessment modality (APA, 2002). When interpreting assessment results, psychologists are further required to address distinct personal characteristics such as cultural and linguistic differences (APA, 2002).

In addition, the American Educational Research Association (AERA), APA, and the National Council on Measurement in Education (NCME) state in the Standards for Educational and Psychological Testing (1999):

> 7.7. In testing applications where the level of linguistic or reading ability is not part of the construct of interest, the linguistic or reading demands of the test should be kept to the minimum necessary for the valid assessment of the intended construct (p. 82).
>
> 9.3 When testing an examinee proficient in two or more languages for which the test is available, the examinee's relative language proficiencies should be determined. The test generally should be administered in the test taker's most proficient language, unless proficiency in the less proficient language is part of the assessment (p. 98).
>
> 9.10 Inferences about test takers' general language proficiency should be based on tests that measure a range of language features, and not a single linguistic skill (p. 99).

Similar to APA's ethics code, the NASP Professional Conduct Manual for School Psychology (2000) indicates:

> C. 1. School psychologists maintain the highest standard for educational and psychological assessment and direct and indirect interventions.
>
> a. In conducting psychological, educational, or behavioral evaluations…due consideration will be given to individual integrity and individual differences.

b. School psychologists respect differences in age, gender, sexual orientation, and socioeconomic, cultural, and ethnic backgrounds. They select and use appropriate assessment or treatment procedures, techniques, and strategies. Decision-making related to assessment and subsequent interventions is primarily data-based.

C. 2. School psychologists are knowledgeable about the validity and reliability of their instruments and techniques, choosing those that have up-to-date standardization data and are applicable and appropriate for the benefit of the child. (p. 27)

As demonstrated, legal sanctions and ethical principles have attempted to address the assessment practices that may result in the inappropriate placement of culturally and linguistically students in special education. However, given the disproportionate representation of diverse students in special education, the challenge to correctly identify culturally and linguistically diverse students persists.

Disproportionate Representation

Regarding disproportionality in special education, overrepresentation of Hispanics in special education programs has occurred more often in states and districts with a higher minority population (Rhodes et al., 2005). More specifically, the percentage of Hispanic students in the categories of specific learning disability, hearing impairments, and orthopedic impairments was higher than that of the general population (U.S. Department of Education, 2000). However, Hispanics are underrepresented in services such as in early childhood intervention and in gifted education (Ochoa, 2003).

Although evidence for disproportionate representation of minority students is often provided by ethnicity, research is limited regarding the representation of bilingual students in special education programs (Artiles et al., 2005; Ochoa, 2003). Considering the population of limited English proficient (LEP) students as a whole, Macias (1998) reported that, in 1996-1997, approximately 8% of United States student population was considered LEP and 7.6% of LEP students were placed in special education programs. Although this data suggests that students of LEP status appear to be represented fairly in special education programs, variability exists when examining within-group differences of LEP in special education at the state and the district level (Artiles et al., 2002, 2005; Rhodes et al., 2005).

Specifically, Artiles et al. (2002) revealed that in eleven urban school districts in California, the state with the largest LEP population (Macias, 1998), linguistically diverse students were overrepresented in the categories of mental retardation and language and speech impairment. This overrepresentation was amplified in sixth through twelfth grades. More specifically, LEP students at the secondary level were almost twice as likely to be placed in special education classes in comparison to their English-speaking peers. Furthermore, LEP students who were placed in English immersion programs were more likely to be placed in special education than LEP students who received more support in their native language. These same authors revealed in a subsequent study (Artiles et al., 2005) that LEP students with limited proficiency in both languages tended to be overrepresented in two categories that are

typically associated with language (i.e., learning disability, language and speech impairment) at the elementary as well as at the secondary grades.

Demographics

As revealed, disproportionate representation of minority students is more evident in districts with a higher minority population. Therefore, concern for disproportionate representation is exacerbated considering that, across the United States, the public school system is becoming increasingly culturally and linguistically diverse. In the 2000-2001 school year, 9.6% of students enrolled in pre-kindergarten to 12th grade were identified as LEP students; 67% of which were enrolled in the elementary school setting (Kindler, 2002). Kindler (2002) reported, "Since the 1990-91 school year, the LEP population has grown approximately 105%, while the general school population has grown only 12%" (p. 3). Only three states (i.e., Alabama, Kansas, New Mexico) revealed a decrease in the number of LEP students enrolled in the public school system during the 1999-2000 term (Kindler, 2002).

Kindler (2002) specified that more than 460 languages were spoken in the U.S. public schools during the 2000-2001 school year. Spanish comprised the overwhelming majority of the languages spoken other than English in the U.S. From 1997 to 2001, the number of Spanish-speaking students within the LEP population has grown from between 66 to 75% (Ochoa, 2003) to 79.2% (Kindler, 2002). Languages of Asian and Pacific Islanders (e.g., Vietnamese, Hmong, Cambodian, Chinese, Japanese) were ranked second (Kindler, 2002; Ochoa, 2003). Given these statistics, it is not surprising

that 57% of surveyed school psychologists reported they had conducted assessments of bilingual/LEP students (Ochoa, Powell, & Robles-Piña, 1996b).

In consideration of these national statistics, important to note is the great variability that exists across states and regions. For example, Blackfoot was listed as the most common language of LEP students in Montana while French was the most common language in Maine. Serbo-Croatian was reported as the language most spoken by LEP students in Vermont, and Hmong was cited as the language most common among LEP students in Minnesota (Kindler, 2002). As shown, students of LEP status are no longer limited to specific languages or to specific geographic regions; therefore, the need for school psychologists across states and districts to be prepared and trained in methods that address the needs of this increasingly diverse population is imperative.

Assessment Practices

As with all children referred for a psychoeducational evaluation, the assessment of LEP students must be multifaceted. For one, factors related to acculturation, second language acquisition, and educational experiences must be duly addressed and understood by the examiner to properly select the mode of assessment as well as to accurately interpret results. Acculturation, as defined by Sattler (2001), is the "process of cultural change that occurs in individuals when two cultures meet; it leads the individuals to adopt elements of another culture, such as values and social behaviors" (p. 639). In psychological evaluations, levels of acculturation may be difficult to reliably assess via standardized measures because, by definition, it is a process of change. Moreover, currently available acculturation questionnaires tend to be unidimensional in

nature and may not address the factors of interest (Rhodes et al., 2005). Therefore, an advantageous choice may be interviews with the parents and, if possible, the individual to address relevant acculturation domains. These domains may include, among others, language use and preference, social affiliation, cultural traditions, cultural identity and values, and generational status (Rhodes et al., 2005).

In addition to acculturation, an understanding of second language acquisition factors is critical in assessing whether the student's academic difficulties should be attributed to an inherent disability or to normal progression in second language development. Cummins' (1984) proposal of two language proficiency thresholds is generally accepted in the fields of education and psychology (Pray, 2005; Rhodes et al., 2005). The first threshold encompasses basic interpersonal communication skills (BICS) while the second, more advanced threshold involves cognitive academic language proficiency (CALP). BICS are the rudimentary conversational skills that are applied to engage in social, informal language and typically take second language learners two to three years to acquire. CALP, on the other hand, involves the more complex, academic communication skills needed to succeed in the classroom and requires a minimum of five to seven years.

In the conceptualization of BICS and CALP, Cummins (1984) applied the iceberg metaphor. The visible surface level of the iceberg (i.e., BICS) embodies simple cognitive (i.e., knowledge, comprehension, application) and language (i.e., pronunciation, vocabulary, grammar) processes needed to be conversationally proficient. The concealed deeper portion of the iceberg (i.e., CALP) involves the sophisticated

cognitive (i.e., analysis, synthesis, evaluation) and language (i.e., semantic meaning, functioning meaning) processes necessary to manipulate language in cognitively demanding, context-reduced settings (Cummins, 1984).

More specifically to bilingual proficiency, Cummins (1984) promoted the theory of a common underlying proficiency (CUP). He expanded upon the aforementioned single iceberg theory to comprise a "dual-iceberg." The two icebergs within this theory connect underneath the surface to create a shared underlying proficiency that facilitates the "transfer of cognitive/academic or literacy-related skills across languages" (Cummins, 1984, p. 143; MacSwan & Rolstad, 2005). This transfer is facilitated by the development of an underlying proficiency (i.e., CALP) in their native language and is more accessible from the minority language to the majority language (e.g., English).

Cummins (1984) also fostered the threshold hypothesis to account for the relation between language proficiency and cognitive and academic functioning. This theory suggests that the level of proficiency in two languages is positively correlated with cognitive and academic development. In other words, bilingual children who achieve a certain threshold of proficiency in two languages may experience greater cognitive and academic success than their monolingual counterparts. On the other end of the continuum, "bilingual" children who demonstrate a low level of proficiency in one or both languages may be more likely to experience negative outcomes.

As proposed by Cummins, subsequent research has revealed that a relationship does exist between a child's native language and English (MacSwan & Rolstad, 2005; Thomas & Collier, 1997). Even more, placement of LEP students in classes that do not

foster the development of cognitive academic language proficiency (CALP) in their native language will likely have detrimental effects on their educational outcomes, especially at the secondary level (Rhodes et al., 2005; Thomas & Collier, 1997). For example, Thomas and Collier (1997) found that LEP students who received less formal schooling in their native language required more time to develop CALP in English. They further cautioned that children might appear to make great gains in the first couple of years of being introduced to a second language but that these gains tend to be only in BICS. As shown, the educational experiences of LEP children can significantly impact academic outcomes. In situations where children have been denied the opportunity to develop CALP in their native language, the examiner must consider whether the experienced academic difficulties are due to an inherent disability or to a lack of appropriate instruction.

Assessment of Language Proficiency

Although the operational definition of language proficiency has long been debated (Del Vecchio & Guerrero, 1995), Hamayan and Damico (1991) offer the following: An individual's proficiency in a particular language is "the degree of control one has over the language in question" (p. 41). Furthermore, language proficiency is not considered a single skill but typically entails listening, speaking, reading, and writing. Despite the lack of a consistent definition, legal mandates, ethical principles, and testing standards require that language proficiency and its relation to BICS and CALP be assessed and evaluated in both languages (Ortiz, 2002; Rhodes et al., 2005). Furthermore, current results (i.e., less than six months old) from dual language

proficiency assessments are essential because they inform the appropriate selection of measures of intelligence and achievement as well as provide key insight into the interpretation of the obtained results (Rhodes et al., 2005).

The assessment of dual language proficiencies should be conducted informally and formally as well as determine receptive and expressive skills (Lopez, 1997; Ochoa, 2003). Informal measures of language proficiency may include observations and language samples conducted across multiple settings (Lopez, 1997) as well as methods such as story-telling, story-retelling, and cloze techniques (Cloud, 1991; Rhodes et al., 2005). Of the available formal measures of language proficiency, the Woodcock-Muñoz Language Survey (WMLS; Woodcock & Muñoz, 1993, 2001) is one of the few that specifically provides information of CALP in English and in Spanish (Rhodes et al., 2005; Woodcock & Muñoz, 1993, 2001).

In addition to measuring CALP in two languages, the WMLS is recommended over other measures of language proficiency given the breadth of skills measured that theoretically comprise language proficiency (i.e., oral language, reading, writing). Results from the WLMS also assist in the eligibility determination of bilingual services, the development of educational goals and objectives, the determination of readiness for English-only instruction, and the evaluation of program efficacy. The authors of the WMLS further endorse the usability of the WMLS in research given its wide range of age norms and accessible results (Woodcock & Muñoz, 1993, 2001). Specific to CALP level classifications, five primary clusters are offered to describe the attained level of proficiency. These clusters range from one to five with one signifying a negligible level

of CALP in either language and five representing advanced language skills. A CALP level of at least a four is indicative of fluency in the observed language. Regarding interpretation, if exposed to English-only curriculum, a Spanish-speaking student with a CALP level of four is expected to find the "language demands of the learning task manageable" (Woodcock & Muñoz, 2001, p. 43).

Despite the aforementioned laws, ethical standards, and recommended practices, ambivalence and noncompliance continues to exist with regard to appropriate assessment methods implemented with bilingual children. For example, Ochoa, Galarza, and Gonzalez (1996) examined school psychologists' assessment practices employed to evaluate the language proficiency of bilingual children. All participants were NASP members and consequently obliged to adhere to the NASP standards and ethics that indicate that school psychologists are to employ best practices in the educational and psychological assessments of children. The majority (62%) of participants in this study who had conducted bilingual assessments reported that they usually administered a measure of language proficiency themselves. Due to inconsistent and lacking responses on key questions, however, the authors surmised that the language proficiency data may have been collected primarily in English and not in both languages. This data suggests that most school psychologists are in fact not adhering to legal requirements, ethical standards, or best practice by not evaluating language proficiency in the first and second languages. A second area of concern was, even though this survey was conducted the same year in which this measure was originally published (i.e., Woodcock-Muñoz Language Survey, 1993), only seven percent of the participating school psychologists

used a measure purported to measure CALP. Furthermore, the majority of utilized language proficiency measures did not assess more than one skill area as required by testing and ethical standards (Ochoa et al., 1996a).

Assessment of Cognitive Ability

The cognitive assessment of linguistically diverse students has been extensively discussed and admonished since the days of Sánchez. But, still, no approach is widely accepted, or even more, practical given the numerous factors previously discussed that must be addressed and the shortage of appropriately trained bilingual examiners. Practice standards suggest that tests should be administered in the child's dominant language when tests in both languages are available (AERA, APA, NCME, 1999). However, current recommended practices advocate for bilingual assessment rather than the mere assessment of bilingual children in a monolingual fashion. Bilingual assessment entails the "evaluation of a bilingual individual, by a bilingual examiner, in a bilingual manner" (Rhodes et al., 2005, p. 161) and is considered more authentic and possibly more valid given the ability of the examiner and/or the examinee to employ either language at any time. However, no currently available measures allow for this naturalistic flow between languages.

Nevertheless, in Rhodes et al. (2005), a unique, promising integrated approach is offered to address some difficulties present in the assessment of culturally and linguistically diverse given the currently available, largely monolingual measures. This approach is the first systematic assessment model devised specifically to address concerns with the selection of the most appropriate assessment modality for culturally

and linguistically diverse children. Salvador Hector Ochoa and Samuel O. Ortiz, the authors of this approach, affirmed

> The most appropriate approach or modality of assessment (i.e., bilingual, nonverbal, English, or native language) depends primarily on knowledge and integration of the individual's current age or grade, the type and nature of formal education he or she has received, and his or current level of proficiency in both languages, not simply relative dominance. (Rhodes, Ochoa, & Ortiz, 2005, p. 168)

Therefore, they stress that language dominance does not examine the relationship between the first language and the second language and that this relationship must be accounted for in the selection of assessment methods and interpretation of assessment results.

In attempt to capture the relationship between the dual levels of proficiency into a manageable design, the authors delineated a "language profile" based upon the level of CALP in both languages. Note, although they caution against the use of one measure of language proficiency, they highlight the benefits of using the WMLS to determine level

of CALP in both languages. They denoted a CALP level of one to two as indicative of a "minimal" level of proficiency in the particular language, a CALP level of three as an "emergent" level of proficiency, and a CALP level of four to five as a "fluent" level of proficiency. The authors further defined the language profile exhibited by a child based on the relationship between the level of proficiency in native language and the level of proficiency in the second language. For example, if "Sarah" demonstrated "minimal" levels of proficiency in L1 and in L2, then her level of proficiency would be categorized as Language Profile 1. If another child, "José" revealed a "fluent" level of proficiency in L1 but a "minimal" level of proficiency in L2, then Language Profile 3 would best characterize his proficiency in both languages. Because three levels of proficiency have been addressed (i.e., minimal, emergent, fluent) for two languages, nine possible language profiles exist. Please refer to Language Profiles of Second Language Learners in Table 1 for further explanation of the nine language profiles formulated by Ortiz and Ochoa (in Rhodes, Ochoa, & Ortiz, 2005).

Table 1
Language profiles of second-language learners

Language Profile	L1 Proficiency Level	L2 Proficiency Level	Description
Profile 1	Minimal	Minimal	CALP levels in native language (L1) and English (L2) are both in the 1-2 range.
Profile 2	Emergent	Minimal	CALP level in L1 is in the 3 range and L2 is in the 1-2 range.
Profile 3	Fluent	Minimal	CALP level in L1 is in the 4-5 range and L2 is in the 1-2 range.
Profile 4	Minimal	Emergent	CALP level in L1 is in the 1-2 range and L2 is in the 3 range.
Profile 5	Emergent	Emergent	CALP levels in L1 and L2 are both in the 3 range.
Profile 6	Fluent	Emergent	CALP level in L1 is in 4-5 range and L2 is in the 3 range.
Profile 7	Minimal	Fluent	CALP level in L1 is in the 1-2 range and L2 is in the 4-5 range.
Profile 8	Emergent	Fluent	CALP level in L1 is in the 3 range and L2 is in the 4-5 range.
Profile 9	Fluent	Fluent	CALP levels in L1 and L2 are both in the 4-5 range.

Adapted from Rhodes, Ochoa, & Ortiz (2005).

Given this information, the level of proficiency for each language should not only be established but the relationship between these two levels of proficiency (e.g., language profile) should also be evaluated.

Although the assessment of cognitive ability in a child's native language is recommended when CALP fluency is achieved, a limited number of measures offered in a child's native language are available. Therefore, the use of nonverbal testing may be "regarded as an acceptable or promising practice" (Ochoa, 2003, p. 576). Bracken and Naglieri (2003) further indicated "nonverbal tests of intelligence are designed to reduce the bias associated with influences of language in an assessment, when language is not the primary construct targeted for assessment" (p. 244). Also, with the increase of ethnically and linguistically diverse students across all regions of the United States, nonverbal measures may be the only option when an appropriately trained bilingual assessor is not available. Nevertheless, limitations of typical nonverbal instruments exist, such as predictive validity, narrow measures of intelligence, and underrepresentation of linguistically-diverse children within in the standardization samples (Ochoa, 2003).

Bracken and Naglieri (2003) specified that the only "true" nonverbal measures of intelligence included the Test of Nonverbal Intelligence (TONI), the Comprehensive Test of Nonverbal Intelligence (CTONI), and Universal Nonverbal Intelligence Test (UNIT). They also recognized the Leiter International Performance Test – Revised (Leiter-R) but noted that some directions are presented verbally. Of these instruments, only two are considered comprehensive nonverbal tests of general intelligence. They

included the Leiter-R and the UNIT (Bracken & Naglieri, 2003; Braden & Athanasiou, 2005). Measures were considered comprehensive if they assess multiple aspects of abilities such as memory, attention, and reasoning.

The Leiter-R (Roid & Miller, 1997) consists of 20 subtests that comprise two cognitive batteries, each with 10 subtests. Battery One examines fluid reasoning and visual-spatial abilities while Battery 2 measures attention, memory, and learning processes (Bracken & Naglieri, 2003). Some significant limitations of the Leiter-R includes the small standardization sample that varies in representation when examining individual age levels, the inclusion of verbal directions with some subtests, and the inconsistent, non-standardized pantomime directions across subtests (Bracken & Naglieri, 2003). The developers of the Leiter-R also did not report obtained stability correlations, only corrected correlations. This practice is not supported by Standards for Educational and Psychological Testing that indicated to report both types of correlations. Furthermore, the test may be more culturally-loaded given the bonus points for speed and accuracy and the allowance for examiners to verbally provide directions (Bracken & Naglieri, 2003; Braden & Athanasiou, 2005). Furthermore, no information regarding internal consistency or factor structure was provided for subgroups within the Leiter-R standardization sample (Braden & Athanasiou, 2005).

On the other hand, one nonverbal test, the UNIT, appears to minimize the limitations associated with other nonverbal tests (Fives & Flanagan, 2002). The UNIT is especially applicable to children from diverse linguistic and cultural backgrounds. For one, directions are completely nonverbal and do not require overt expressive or receptive

linguistic abilities (via the use of pantomime and gestures) (Bracken & McCallum, 1998; Fives & Flanagan, 2002). Checkpoint, sample, and demonstration items are also provided to promote understanding. The standardization sample included children of Hispanic origin as well as students who were receiving services in English as a second language (ESL) classes and bilingual education (Bracken & McCallum, 1998). Fives and Flanagan (2002) also noted that the UNIT reduces bias because only two of the six subtests are timed (one of which comprises only the extended battery). With regard to test fairness, Bracken & Naglieri (2003) reported that the UNIT is a fair and non-discriminative instrument because this test is completely nonverbal, has adequate psychometric properties, minimized the influence of culture, and all items were evaluated by a "comprehensive and inclusive 'sensitivity' panel" (p. 259).

Following their review, Fives and Flanagan (2002) advocated the use of the UNIT as a nonverbal measure of general intelligence. According to Fives and Flanagan (2002), "The UNIT appears superior to other nonverbal measures in that administration is completely nonverbal, it measures multiple abilities and it is available in Abbreviated, Standard and Extended Battery forms" (p. 443). They further applauded the steps taken to ensure fairness of the UNIT for different racial and ethnic groups as well as the sound psychometric properties.

To determine practices typically employed by school psychologists to assess the cognitive functioning of bilingual children, Bainter and Tollefson (2003) recruited NASP members from states with a high percentage of bilingual students. Participants from this study indicated that Spanish was spoken by 94% of second language learners,

while 17% of the participating school psychologists spoke Spanish. The majority of respondents (85%) reported that the employment of a bilingual school psychologist to administer tests in the child's native language and in English was a "usually to an always" acceptable practice. Of concern, however, is that 87% of respondents reported that the use of traditional intelligence tests in English when the child is dominant or "prefers using" English is acceptable practice. No mention was made to how English dominance or "preference" was determined, to whether the tests' norming procedures were considered, or to the types of "traditional intelligence tests" being administered. Fortunately, most respondents (i.e., 56 to 74%) concluded that it is rarely or never acceptable to administer tests in English when the child is dominant in another language, to administer nonverbal measures without an interpreter when oral instructions are presented, or to translate traditional English tests into Spanish during the testing session. Although this study reveals that school psychologists are relatively familiar with the very basics of best practices (or what is suggested not to be done), the authors did not take the opportunity to query what practices are actually employed by school psychologists in the intellectual assessment of bilingual children.

Based on findings from earlier studies, however, the Wechsler scales appear to be the most commonly employed measures of intellectual ability in the assessment of bilingual children despite its well-documented limitations with this population (McCloskey & Athanasiou, 2000; Ochoa et al., 1996b). The Test of Nonverbal Intelligence-Third Edition (TONI-3) also appears to be a favorite among school psychologists (McCloskey & Athanasiou, 2000; Ochoa et al., 1996b). The Bender

Visual-Motor, the Gestalt test, the Draw-A-Person test, and the Leiter were ranked as instruments typically used in the intellectual assessment of bilingual children (Ochoa et al., 1996b). Nevertheless, limitations of these measures with diverse populations (Rhodes et al., 2005) include limited psychometric properties, incomprehensive results, and suspected test bias. Furthermore, norms of intelligence tests in the English language typically do not include bilingual students (Barona & Santos de Barona, 1987).

Assessment of Academic Achievement

Achievement testing is often performed to assess the academic skills and abilities such as reading, mathematics, writing that children learn through instruction (Stetson, Stetson, & Sattler, 2001). Similar to the assessment of cognitive ability, the selection of appropriate measures of academic achievement with LEP students continues to perplex psychologists. When a child is considered bilingual, the assessment of academic performance in both languages is suggested to account for the influence of previously described psychosocial variables (i.e., acculturation, educational experience, second language acquisition). However, limitations exist with the academic assessment in a child's native language. For one, a shortage of appropriately trained bilingual examiners endures. Secondly, of the few norm-referenced academic measures available in languages other than English, most are available only in Spanish. Third, of these measures in Spanish, the standardization sample are often comprised of monolingual, Spanish-speaking students from several different countries. Furthermore, studies have not examined the psychometric properties of these instruments with a sample of LEP students.

Overall, the Woodcock tests of achievement in English and in Spanish were reported as the most commonly used instruments in the academic assessment of bilingual children (Ochoa et al., 1996b). Sound psychometric properties as well as satisfactory norming properties for Spanish-speaking students were cited as unique strengths of the Woodcock instruments. However, Ochoa et al. (1996b) urge the examination of the psychometric properties and the differential item functioning of these instruments with Spanish-speaking students.

Another area of concern with the academic assessment of LEP children includes the large-scale assessments associated with high-stakes testing. The recent adoption of No Child Left Behind Act of 2001 (NCLB; Pub. L. No. 107-110) was designed to "close the achievement gap with accountability, flexibility, and choice, so that no child is left behind" (Jacob & Hartshorne, 2003, p. 37). As federal law, NCLB currently requires annual statewide assessments in the major areas of reading and math, beginning in the third grade. Results from this testing must also include information on specific subgroups such as economically disadvantaged, racial and ethnic populations, children with disabilities, and students with limited English proficiency (LEP; Abedi, 2004).

The Texas Assessment of Knowledge and Skills (TAKS) is one example of a state-administered assessment that has attempted to appropriately evaluate the annual yearly progress of LEP students. For one, the TAKS is available in English and in Spanish. The language proficiency assessment committee (LPAC) determines whether a student of limited English proficiency should be administered the TAKS in English or in Spanish in grades three through six. Note, however, the Spanish version of the TAKS

may only be administered for a maximum of three years. Secondly, the Texas English Language Proficiency Assessment System (TELPAS) was developed to adhere to NCLB's requirement that LEP students must be assessed annually in the areas of listening, speaking, reading, and writing. Beginning in the third grade, LEP students in Texas are required to participate in the Reading Proficiency Tests in English (RPTE) and the Texas Observation Protocols (TOP). Both components of the TELPAS attempt to measure the annual progress made by LEP students in English reading proficiency. However, it is unclear if the TELPAS is designed to account for CALP. As such, the relation among CALP and academic achievement, as measured by this state-administered, criterion-reference test, has yet to be studied.

The Relationship of Language Proficiency, Intelligence, and Achievement in the Assessment of LEP Students

Although widely accepted as a measure of CALP, the psychometric properties of the scores obtained via the WMLS in English and in Spanish and their relation to academic achievement have not been adequately researched. For example, DiCerbo (2003) examined the relation of English language proficiency (as measured by the WMLS) with an English measure of intelligence (i.e., Wechsler Intelligence Scale for Children-Third Edition [WISC-III]) and of academic achievement in English (i.e., Wechsler Individual Achievement Test-Second Edition [WIAT-II]). Participants of this study included 172 Hispanic children referred for a psychoeducational evaluation due to reading difficulties. Results of this study revealed that English proficiency was positively correlated with IQ and accounted for a significant portion of the variance on

the WISC-III Verbal and Performance domains (after controlling for demographic variables). Similarly, with regard to achievement, DiCerbo found that, even after controlling for demographic variables and IQ, English proficiency was significantly, positively correlated with scores on the WIAT-II Reading composite scores. Based upon her results, DiCerbo surmised that the WMLS might not be a pure measure of language proficiency given the relationship among scores of language proficiency, intelligence, and academic achievement.

However, two primary limitations emerge from this summation. For one, a statistically significant and positive correlation among scores on the WMLS, WISC-III, and WIAT-II should be expected given the WMLS authors' aim that CALP "should correlate well with important aspects of school achievement" (Woodcock & Muñoz, 2001, p. 68). Secondly, DiCerbo employed the Language Assessment Scale (LAS) to establish English as the language modality for assessment, not the WMLS. The LAS is not a formal measure of CALP and, therefore, may not accurately classify language proficiency in terms of what is needed to be successful in school. To elaborate, DiCerbo (2003) reported that for her sample the mean for the WMLS Oral Language scores was 76.77 (SD=8.37) while the mean for the WMLS Reading-Writing scores was 68.63 (SD=7.02). Therefore, it appears that the majority of these participants did not actually demonstrate fluency in English oral language or reading-writing skills on a purported measure of CALP. Consequently, the employed measures of intelligence and achievement in English, may have served more as a measure of English language skills than of their intended purposes.

García-Vázquez, Vázquez, López, and Ward (1997) examined the relation among the individual subtest scores of the WMLS English and Spanish versions with the performance on the Iowa Tests of Basic Skills (ITBS) in English. As would be expected, a statistically significant, positive correlation was determined for overall language proficiency in English (i.e., Broad English Ability) and achievement scores on the ITBS in English. Written language performance on the WMLS-English was shown to be the most highly correlated variable with overall academic achievement as measured by the ITBS (r = .84, p<.01). Furthermore, second language (i.e., Spanish) proficiency in reading and written language also demonstrated a positive, statistically significant relationship with general academic success in English (r = .21, .30, respectively). These results suggest that reading and writing proficiency in Spanish are associated with overall academic success in English and that language proficiency should be ascertained by measures that tap reading and writing skills, not simply oral language proficiency, to predict academic success.

Laija-Rodríguez, Ochoa, and Parker (2006) also examined the crosslinguistic role of language proficiency in the Spanish and in English on academic achievement. Using the WMLS as measures of CALP and curriculum-based oral reading probes as measures of academic achievement, these researchers assessed whether accounting for Spanish and English concurrently would serve as a better predictor of reading growth than the frequently employed method of separately examining Spanish and English. Participants were 77 Hispanic students identified as LEP enrolled in either the second grade (71.4%) or the third grade (28.6%). Although adequate reliability coefficients for

the curriculum-based measures were reported, no information for the score reliability of the WMLS-English or the WMLS-Spanish was provided. Contrary to expectations, results from this study revealed that the crosslinguistic role of Spanish and English did not serve as a better predictor of reading growth in English or in Spanish. They, nevertheless, hypothesized that these results may actually demonstrate a crosslinguistic relationship between the two languages based on the shared variance that may be anticipated between the two measures of language proficiency. This shared variance may serve to regulate the ability to differentially predict achievement based on language proficiency. Another noteworthy consideration that may have further minimized the variance was the small number of children (n=11) who obtained English CALP levels above the moderate range (i.e., CALP $\geq$ 3). Ultimately, Laija-Rodríguez et al. (2006) addressed the need for additional studies to further address the crosslinguistic role of CALP on academic achievement.

Regarding cognitive functioning and academic achievement, Laija (2001) also examined the role of nonverbal cognitive reasoning to predict reading growth in English and in Spanish. In her analyses, the TONI-3 was used to assess nonverbal cognitive reasoning while reading growth was measured by curriculum-based oral reading fluency probes administered in English and Spanish. Results from this analysis revealed that nonverbal cognitive reasoning contributed to the predictive relationship for reading growth in English or in Spanish for this sample of second and third grade Hispanic, LEP students. However, as indicated by Laija, the use of a measure of nonverbal intelligence to predict academic achievement has revealed limited results for monolingual, English-

speaking students as well as for LEP students (Figueroa, 1990). Therefore, a nonverbal measure of general intelligence, such as the UNIT, may be more appropriate to examine the relationship of intelligence and achievement with LEP students.

Summary

The primary purposes of intelligence tests are to determine eligibility for special education services and to predict future school achievement (Holtzman & Wilkinson, 1991). As previously insinuated in the fall of Cattell's mental measures and the rise of the Binet scales, results from tests of intelligence should correlate with academic achievement in order for the intelligence test to be useful and subsequently successful. Although some evidence suggests that nonverbal measures of general intelligence are able to successfully predict academic achievement (Williams & McCallum, 1995), the reliability and validity of such measures with low-performing LEP students is limited. This is of particular concern because professional standards and ethical guidelines emphasize the need to use instruments that demonstrate adequate reliability and validity with the population under study.

Research has further demonstrated a relationship between language proficiency and academic achievement (DiCerbo, 2003; García-Vázquez, Vázquez, López, & Ward, 1997). Although initially believed that bilingualism had adverse effects on children's intellectual and academic outcomes (Garcia, 1992), research has now revealed that language proficiency in two languages enhances the academic progress of children (Thomas & Collier, 1997), and this progress seems particularly apparent once a threshold of in the second language is reached (Cummins, 1984). However, the

relationship of proficiency in two languages in the prediction of reading achievement has yet to be adequately studied, especially with regard to achievement via commonly employed standardized instruments in Spanish.

CHAPTER III

METHOD

The present study used a descriptive, correlational design to examine the internal consistency of scores of linguistically diverse students on theoretically appropriate, commonly employed measures of language proficiency, general intelligence, and academic achievement. The criterion-related validity of scores on a nonverbal measure of general intelligence was also investigated. Multiple regression analyses were conducted to explore the influence of language proficiency on the strength of the relationship between language proficiency in the subordinate language and reading achievement.

Participants

The participants of interest in the present study were elementary-age students participating in a larger longitudinal study of school achievement. At the time of recruitment, participants attended one of three school districts (one urban, two small city) in Southeastern Texas. Participants of the two small city districts were recruited during their first grade year via two sequential cohorts; cohort 1 was recruited during fall 2001 and cohort 2 was recruited during fall 2002. Participants of the one urban district were recruited from one cohort during fall 2001. Students were eligible to participate in the larger study if they scored below the median score for their district on a state-approved, district-administered measure of literacy.

Of the 1,374 eligible participants from cohort 1 and cohort 2 of the larger longitudinal study, 784 (57.1%) parents provided written consent for their child to

participate in the study. Of these 784 participants, 57.3% were recruited in cohort 1 and 47.4% were female. Approximately 76% attended a small city district while approximately 24% attended an urban district. The ethnic composition of the total consented sample was: Native American/Alaskan Native (n=2), Asian/Pacific Islander (n=28), African-American (n=182), Hispanic (n=293), Caucasian (n=267), Other (n=12). Furthermore, 128 of these 784 students were identified as limited English proficient in first grade. Participants for the current study were purposely selected based upon two criteria: Hispanic origin and limited English proficiency status. These two variables were provided by the respective school district, and the demographics will be described separately for the two time periods of interest.

The first time period (Time 1) consists of first grade students while the second time period of interest (Time 3) includes third grade students. As a preliminary step in the evaluation of the predictive validity of the UNIT, the sample of Time 1 participants were employed to estimate the concurrent validity of the UNIT. The sample of Time 3 participants is of primary interest to the goal of this study.

Of the 784 children with consent to participate in the larger study, 114 (13.0%) of the Time 1 participants met criteria for the present study. Of these 114 Time 1 participants, 43.9% (n=50) children were enlisted in cohort 1. Approximately 98% (n=112) of the Time 1 participants attended a small city district while approximately 2% (n=2) attended an urban district. This Time 1 sample consisted of 51 male participants and 63 female participants.

Of the original 784 children with consent from Time 1, 144 (18.4%) participants in Time 3 met criteria for the current study. The increase in students of LEP status may have been due to the time at which they were identified. Members of cohort 1 comprised 83 (57.6%) of the selected 144 participants. Approximately 74% (n=106) of the Time 3 participants attended a small city district while approximately 26% (n=38) attended an urban district. The gender composition of this sample was 78 male and 66 female.

Internal Consistency of Measures

The sample used to estimate the internal consistency of measures typically employed with LEP students was based upon a random selection of approximately 35% of students. This random selection was completed via SPSS 11.5 for Windows, and the this percentage of participants was utilized to provide adequate representation of the general sample. Of the 24 students randomly selected, 13 (54.2%) were cohort 1 participants and 13 (54.2%) were female. Seventeen (70.8%) attended a small city school district while the remaining seven (29.2%) attended an urban district. As specified by the original selection criteria, all participants in this sample were Hispanic and identified as LEP.

Criterion-Related Validity of UNIT

The preliminary step employed to estimate the predictive validity of the UNIT was to estimate the concurrent validity of the UNIT with two measures of achievement, each administered in Time 1. The concurrent validity of the UNIT with a sample of LEP students was estimated with the Batería-R APR: Reading. This sample was comprised

of 56 Hispanic, LEP participants. Twenty (35.7%) were cohort 1 participants and 26 (46.4%) were female. All participants of this sample were enrolled in a small city district.

The concurrent validity of the UNIT with a sample of non-LEP participants was estimated via the WJ-III ACH: Reading. This sample was comprised of 89 Hispanic, non-LEP participants. Of these 89 participants, 58 (65.2%) were enlisted in cohort 1 and 48 (53.9%) were female. Fifty-seven (64.0%) were enrolled in a small city school district while 32 (36.0%) were enrolled in an urban school district.

The predictive validity of the UNIT was estimated for a group of LEP students and a group of non-LEP students based upon their performance, respectively, on the Spanish and English versions of the TAKS. The sample of LEP students administered the TAKS-Spanish consisted of 51 participants. Twenty-nine (56.9%) were recruited in cohort 1 and 23 (45.1%) were female. Of these 51 participants, 36 were enrolled in a small city district and 15 were enrolled in an urban district.

Within the non-LEP comparison group, all participants also were of Hispanic origin (n=77). Of these non-LEP participants, 45 (58.4%) were enlisted in cohort 1 and 42 (54.5%) were female. Fifty-nine (76.6%) attended one small city school district while the remaining 18 (23.4%) attended an urban school district. Table 2 displays the demographic characteristics of each subsample used to explore the validity of the UNIT for the present study.

Table 2
Sample demographic characteristics for criterion-related validity analyses

	Concurrent validity		Predictive validity	
	Bateria-R	WJ–III	TAKS-Spanish	TAKS-English
N	56	89	51	77
Cohort				
1	20	58	29	45
2	36	31	22	32
Gender				
Female	26	48	23	42
Male	30	41	28	35
District				
Small City	56	57	36	59
Urban	0	32	15	18

The Relation of Language Proficiency in the Prediction of Reading Achievement

Four subsamples were used to explore the relationship of language proficiency in the prediction of reading achievement. Performance on two norm-referenced measures of reading achievement (i.e., Batería-R APR, WJ-III ACH) as well as on two criterion-referenced measures of reading achievement (i.e., TAKS-Spanish, TAKS-English) was examined. Participants were administered the achievement measure in their dominant language. If equivalent, then the English version of the measure was administered. As previously indicated, all participants in each of the four subsamples were Hispanic and identified as LEP. The first sample who achieved Spanish dominance were administered the Batería-R APR and consisted of 79 participants. Of these participants, 46 (58.2%) were recruited in cohort 1 and 41 (51.9%) were female. Sixty (75.9%) participants attended a small city school district and 19 (24.1%) attended an urban district. The sample administered the WJ-III ACH as a measure of reading achievement was

comprised of 14 participants. Twelve (85.7%) were cohort 1 participants while four (28.6%) were female. Ten attended a small city district.

Of the 54 participants administered the Spanish version of the TAKS, approximately 61% (n=33) were recruited in cohort 1. Thirty-eight participants (70.4%) were from a small city district and 16 (29.6%) were from an urban district. The number of males to females was equivalent. Twelve participants comprised the subsample administered the English version of the TAKS. Ten were enlisted in cohort 1 and, as with the previous sample, males and females were equally represented. Eight were enrolled in a small city district while four were enrolled in an urban district. Table 3 displays the demographic characteristics of the four subsamples utilized to explore the association of language proficiency in the prediction of reading achievement.

Table 3
Sample demographic characteristics of hierarchical multiple linear analyses

	Norm-referenced		Criterion-referenced	
	Bateria–R	WJ–III	TAKS-Spanish	TAKS-English
N	79	14	54	12
Cohort				
1	46	12	33	10
2	33	2	21	2
Gender				
Female	41	4	27	6
Male	38	10	27	6
District				
Small City	60	10	38	8
Urban	19	4	16	4

Measures

Language Proficiency

To assess the participants' language proficiency in English and in Spanish, the respective forms of the Woodcock-Muñoz Language Survey (WMLS; Woodcock & Muñoz-Sandoval, 2001) were individually administered to the participants. Four subtests (i.e., Picture Vocabulary, Verbal Analogies, Letter-Word Identification, Dictation) of each form were administered to assess participants' proficiency in oral language, reading, and writing in both languages. These four subtests yield an overall measure of language competence (i.e., Broad English Ability, Broad Spanish Ability) as well as a level of cognitive-academic language proficiency (CALP) (Cummins, 1984). The Broad Ability scores are represented by standard scores with a mean of 100 and standard deviation of 15. CALP scores range from one (negligible proficiency) to five (advanced proficiency). The WMLS is unique among language proficiency measures in that it is one of the few that provides specific data on the language skills required to be successful in the academic setting (Rhodes et al., 2005).

The standardization sample of the WMLS in English consisted of 8,818 participants representative of the projected United States population for the year 2000. The sample was stratified based upon significant demographic data such as census region (i.e., northeast, midwest, south, west), community size (i.e., central city and urban fringe, larger community and associated rural area, smaller community and associated Rural area), sex (i.e., male, female), race (i.e., White, Black, American Indian, Asian and Pacific Islander), Hispanic (i.e., Hispanic, non-Hispanic), type of school (i.e., public,

private, home). Performance on the Spanish version of the WMLS was equated to similar levels of performance on the English version via its administration to over 2,000 participants from Argentina, Costa Rica, Mexico, Peru, Puerto Rico, Spain, and the United States.

Although only reported for the WMLS English Form, the internal consistencies of the Broad English Ability cluster scores for the norm group was high for all age groups (r_{11}>.94). Concurrent validity of the WMLS with school-aged children was assessed via correlational studies with the Wechsler Intelligence Scale for Children – Third Edition (WISC-III) as well as with three other common measures of language proficiency in English and in Spanish. Correlations for the scores on the WMLS Broad English Ability and the WISC-III clusters were .80 for the Verbal IQ, .76 for the Verbal Comprehension Index, and .55 for the Performance IQ. As reported in the manual, the "higher correlation of all WMLS Normative Update tests and clusters with the WISC-III Verbal IQ and Verbal Comprehension Index is evidence that the WMLS Normative Update measures the construct of CALP" (Woodcock & Muñoz-Sandoval, 2001, p. 60). Studies of the relationship between the WMLS and three other tests of language proficiency (i.e., Language Assessment Scales-Oral Short Form, preLAS, and IDEA Oral Language Proficiency Tests-Oral) with bilingual children in grades kindergarten through third grade suggested that scores from the WMLS-English generally demonstrated correlations between .70 and .80 with the English versions of other language proficiency measures, and scores from the WMLS-Spanish generated

correlations between .60 and .70 with the Spanish versions of the three other language proficiency measures.

To further illustrate the utility of the WMLS as a measure of CALP in English and in Spanish, performance on both versions of the WMLS was compared to performances on the Woodcock-Johnson – III Tests of Achievement (English) and the Batería Woodcock Psico-Educativa en Español (Spanish). For norm sample participants in grade 1 and in grade 3, the reported correlations among WMLS-English and the WJ-III ACH in reading as well as the WMLS-Spanish and the Batería in reading were reported as similar and satisfactory (i.e., correlation coefficients between .79 and .85). Achievement in mathematics tended to maintain the lowest correlation with the respective forms of the WMLS for grade 1 and grade 3 with correlation coefficients between .62 and .66 (Woodcock & Muñoz-Sandoval, 2001).

Cognitive Ability

The Abbreviated Battery of the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998), comprised of the subtests Symbolic Memory and Cube Design, was individually administered to the participants of the present study at their respective home school. The UNIT was selected due to its distinctiveness as an entirely non-verbal, yet comprehensive measure of general intelligence for children and adolescents (ages 5 through 17) "who may be disadvantaged by traditional verbal and language-loaded measures" (Bracken & McCallum, 1998, p. 1). The subtest Symbolic Memory, as insinuated by its title, is reported to assess complex cognitive memory abilities as well as verbal mediation via a nonverbal format. Cube Design, on the other

hand, provides a measure of reasoning competencies as well as of the ability to process nonsymbolic stimuli. The overall general ability scores obtained via the UNIT (i.e., FSIQ) is represented by standard scores with a mean of 100 and standard deviation of 15.

The UNIT standardization sample was comprised of 2,100 children and adolescents aged 5 years, 0 months to 17 years, 30 days. The standardization sample was randomly stratified to represent the United States population based upon the 1995 U.S. Census Data with regard to sex, race (i.e., White, African American, Asian/Pacific Islander, Native American, Other), Hispanic origin (i.e., Hispanic, non-Hispanic), geographic region (midwest, northeast, south, west), and community setting (i.e., urban/suburban, rural). Additional factors addressed in the UNIT standardization sample were classroom placement (i.e., full-time regular classroom, full-time self-contained classroom, part-time special education resource, other), education services (i.e., learning disability, speech and language impairments, mental retardation, giftedness, serious emotional disturbance, English as a second language and bilingual education, regular education), and parental educational attainment (i.e., less than high school degree, high school graduate or equivalent, some college or technical school, four or more years of college).

With regard to score reliability based upon the performance of participants within standardization sample, the UNIT Abbreviated Battery was reported to demonstrate satisfactory internal consistency with a FSIQ reliability coefficient average of .91. Although the internal consistency of the Abbreviated Battery FSIQ scores exceeded .90

for four clinical/exceptional samples, the score internal consistency with a culturally and linguistically diverse sample was not examined. The corrected test-retest stability coefficient of the Abbreviated Battery FSIQ scores from the UNIT standardization sample over a three week time interval was .80 for children five to seven years old (n=46) and .74 for children eight to ten years old (n=42).

In addition to acceptable score reliability, the UNIT demonstrated satisfactory evidence for validity with the standardization sample. The Abbreviated Battery FSIQ evidenced a high intercorrelation average with the Standard Battery FSIQ (.91) and the Extended Battery FSIQ (.86). As anticipated, exploratory factor analyses of the Standard Battery upheld one primary factor (i.e., the UNIT FSIQ) as well as two secondary factors (i.e., Memory, Reasoning). Confirmatory factor analyses tested three models: a one-factor model, a two-factor memory-reasoning model, and a two-factor symbolic-nonsymbolic model. Although they interpreted the results to suggest that the two factor memory-reasoning model is somewhat of a better fit than the other two models, Bracken and McCallum (1998) noted that each of the three models appeared to be a good fit with the data.

The validity of the UNIT was externally examined via correlational studies with other measures of intelligence. The corrected correlations between the Abbreviated Battery FSIQ scores and the Wechsler Intelligence Scales for Children – Third Edition (WISC-III) FSIQ scores were .78 for a sample with learning disabilities, .86 for a sample with mental retardation, .75 for a sample with intellectually gifted, and .87 for a sample with Native American children. The corrected correlation between the Abbreviated

Battery FSIQ scores and the Woodcock-Johnson – Revised Tests of Cognitive Ability (WJ-R Cognitive) Broad Cognitive Ability scores were .80 for a White, Non-Hispanic sample in regular education classes. Interestingly, the Abbreviated Battery FSIQ scores were found to have little to no correlation with the parallel Spanish version of the WJ-R Cognitive (i.e., Batería-R Cognitiva) for a sample of students receiving English as a Second Language (ESL) services or a sample of students receiving bilingual education services. Bracken and McCallum attributed the limited relationship to the variability of the scores exhibited on the Batería-R Cognitiva.

With regard to the predictive validity of the UNIT with academic achievement, scores from the UNIT standardization sample were compared to scores obtained via the Woodcock-Johnson – Revised Tests of Achievement (WJ-R ACH), the Spanish Form of the Woodcock Language Proficiency Battery – Revised (WLPB-R), the Wechsler Individual Achievement Test (WIAT), and the Peabody Individual Achievement Test – Revised (PIAT-R). The Abbreviated Battery FSIQ scores demonstrated low correlations with WJ-R ACH Broad Reading scores for the sample classified as Intellectually Gifted and for the sample identified as Learning Disabled. The corrected correlation between the Abbreviated Battery FSIQ scores and the WLPB-R scores were .55 for the bilingual education sample and .01 for the ESL sample. Validity studies also were conducted for specific clinical and exceptional samples; however, these studies did not include linguistically and culturally diverse students.

To address this limitation, Jiménez (2002) investigated the reliability and validity of the UNIT with a sample of Puerto Rican children in second through fourth grades

with a matched group of non-Hispanic children of the UNIT standardization sample. Results from the analysis of internal consistency suggested that scores from the UNIT Standard and Extended batteries failed to demonstrate adequate item homogeneity with a sample of Puerto Rican children. Specifically, the reliability coefficient for the standard battery was .68 while the reliability coefficient for the extended battery was .62. However, a statistically significant relationship (r=.36, p<.01) was observed between scores on the Batería-R APR Broad Reading composite and the Extended battery Full Scale IQ scores. Furthermore, examination of mean score differences revealed that the ascertained statistically significant differences were positive for the non-Hispanic comparison sample.

Williams and McCallum (1995) further assessed the predictive validity of the UNIT with a state-mandated, norm-referenced measure of achievement (i.e., Comprehensive Tests of Basic Skills [CTBS/4]). The reading portion of the CTBS/4 is composed of Reading Vocabulary and Reading Comprehension. Internal consistency estimates for subtests of the Abbreviated battery of UNIT were greater than .85. The correlation between the Full Scale IQ and the CTBS/4 Reading test was positive and statistically significant (r=.48, p<.01). These researchers determined that the subtest Analogic Reasoning was the best predictor of achievement in reading, math, and written language. Specific to reading achievement, Analogic Reasoning accounted for approximately 21% of the variance. Cube Design also was statistically significant in the prediction of reading achievement, although to a less degree (i.e., change in R^2=.06).

Note the generalizability of these results to other samples is limited given that key demographic characteristics such as race/ethnicity and language status were not reported.

Nevertheless, the UNIT appears appropriate for the current sample given the great strides that the authors took to ensure test fairness and the limited research conducted with a sample of Hispanic, LEP students within the United States. In addition to its completely nonverbal administration, tasks on the UNIT are considered less reliant on acquired knowledge and previous experiences and have less demands of speed. A committee of culturally, ethnically, and racially diverse psychologists also were consulted to examine content and procedures that may lend to potential bias. The mean Abbreviated Battery FSIQ score differences between Hispanic and Non-Hispanic groups of the standardization sample were negligible (difference score = 2.00, effect size = .13) and were described as "smaller than the performance differences between Hispanic and non-Hispanic examinees reported in the literature" (Bracken & McCallum, 1998, p. 188-189). Furthermore, bilingual and ESL examinees performed similar to a demographically matched English-speaking comparison sample on the Abbreviated Battery FSIQ with a difference score of 2.82 and an effect size of .19.

Academic Achievement in Reading

Batería Woodcock-Muñoz: Pruebas de Approvechamiento-Revisada

As one of the few achievement batteries in Spanish, the Batería Woodcock-Muñoz: Pruebas de Approvechamiento-Revisada (Batería-R APR; Woodcock & Muñoz-Sandoval, 1996) was individually administered to participants who obtained a higher CALP score on the WMLS-Spanish. The Batería-R APR is the parallel Spanish version

of the Woodcock-Johnson Tests of Achievement-Revised (WJ-R ACH, Woodcock & Johnson, 1989), which is the direct predecessor to the WJ-III ACH. The Batería-R APR subtests administered to achieve the reading cluster score (i.e., Amplia Lectura) included Identificación de Letras y Palabras (Letter-Word Identification) and Comprensión de Textos (Passage Comprehension). This reading cluster score is represented by standard scores with a mean of 100 and standard deviation of 15.

The standardization sample for the cognitive and achievement versions of the Batería-R APR included 3,911 participants. At least one of the Batería-R APR batteries was administered to 1,325 participants in five U.S. states. The remaining 2,586 participants were administered at least one of the calibration tests in Costa Rica, Mexico, Peru, Puerto Rico, and Spain. Only monolingual Spanish speaking students were included in the standardization of the Batería-R APR because the authors theorized that "subjects that are truly bilingual would present data that would diverge from normal Spanish language development" (Woodcock & Muñoz-Sandoval, 1996, p. 21).

Although internal consistency reliability coefficients (r_{11}) are not reported for the reading cluster score at any age, the manual provided the internal consistency reliability coefficients, respectively, for Identificación de Letras y Palabras (Letter-Word Identification) and Comprensión de Textos (Passage Comprehension) at age 6 (r_{11}=.95, .89) and at age 9 (r_{11}=.93, .92). The manual further noted that the validity results are generalizable from the WJ-R (English) to the Batería-R APR given that their corresponding content and structure (Woodcock & Muñoz-Sandoval, 1996, p. 3).

Woodcock-Johnson III Tests of Achievement

To obtain a broad sampling of the present participants' overall reading level in English, the three subtests that comprise the Broad Reading Cluster (i.e., Letter-Word Identification, Reading Fluency, Passage Comprehension) of Woodcock-Johnson III Tests of Achievement (WJ-III ACH; Woodcock, McGrew, & Mather, 2001) were administered to participants who attained equal or greater proficiency via the WMLS-English CALP score. With its theoretically driven design and recognized psychometric properties, the WJ-III ACH is a commonly used and widely respected comprehensive measure of achievement (Rhodes, Ochoa, Ortiz, 2005). As with the previous measures, the Broad Reading cluster score is represented by standard scores with a mean of 100 and standard deviation of 15.

The standardization sample of the WJ-III ACH is similar to the standardization sample of the WMLS with regarding to number of participants and demographics; therefore, this information will not be repeated here. For the Broad Reading Cluster scores, the internal consistency score reliability estimates, as reported in the WJ-III ACH manual, are .92 for age 6 and .95 for ages 7 through 10. The one-year test-retest correlations for the Broad Reading Cluster scores is .92 (ages 4 to 7) and .93 (ages 8 to 10). Regarding validity, the WJ-III ACH Broad Reading Cluster scores were found to have a positive relationship with the Weschler Individual Achievement Test (WIAT) Reading Composite scores (r=.67).

Texas Assessment of Knowledge and Skills

The Texas Assessment of Knowledge and Skills (TAKS; Texas Education Agency, 2004) is a state-mandated, criterion-referenced standardized measure used to evaluate the performance of all children in specified grades based upon the state-mandated curriculum entitled the Texas Essential Knowledge and Skills (TEKS). The TAKS may be administered in Spanish or in English to LEP students based upon the recommendation of the language proficiency assessment committee (LPAC). The criterion used to make a recommendation for the administration of the TAKS in English or Spanish appears unclear. The scores from the third grade version of the TAKS – Reading were of interest in the present study. However, data regarding the reliability and validity of this measure is unknown.

Procedures

Initially, permission was obtained from each of the three school districts. Subsequently, at the beginning of first grade, for each cohort, letters requesting parental permission were sent to all parents of first grade children who had not been retained previously in first grade and who scored below the 50th percentile on the year's most recent screening measure. Regardless of parents' consent decision, incentives (e.g., McDonald's gift certificates, name entered into a lottery for larger prizes) for children were conferred to increase the rate of return of permission slips. Parents were told that the purpose of the larger longitudinal study is to learn more about children's adjustment to school over time.

Undergraduate psychology students with a junior or higher standing administered the measures of interest in the current study at each participant's home school, with the exception of the TAKS. These research assistants received a minimum of 12 hours of training on the administered measures and were required to pass an administration proficiency check-out on each measure. The measure of cognitive ability (i.e., the UNIT) was administered in the late fall/early spring of the participants' first year in the larger study. Measures of language proficiency were considered current as they were administered in less than six months of measures of academic achievement. Specifically, the WMLS-English, WMLS-Spanish, Batería-R APR, and WJ-III ACH were administered in the late fall/early spring of each participating year. Academic performance from the participants' third grade year on these measures was of particular interest in the present study given that the second measure of academic achievement of interest in this study (i.e., TAKS: Reading) was initially administered during the spring of participants' third grade year. The TAKS: Reading is administered by the presiding district, and results were provided by the respective districts and represented by raw scores and the percent correct.

Measures of language proficiency were administered to participants if the school identified them as Spanish-speaking, limited English proficient (LEP), if they were enrolled in a bilingual education classroom, if their parent and/or teacher indicated Spanish as a language spoken at home, or at examiner request. The language in which the measure of achievement was administered was dependent upon relative CALP dominance obtained via the respective forms of the WMLS. Participants who attained

equivalent CALP levels were administered measures of achievement in English. Although the LPAC team determined the language in which the TAKS was administered, for purposes of the current study, only participants who were administered the TAKS in their dominant language based upon the results of the WMLS were selected for the present study.

CHAPTER IV

RESULTS

Three primary research questions were investigated in this study. First, the score reliability of each of the following four measures was examined: Woodcock-Muñoz Language Survey, Spanish and English versions (WMLS-Spanish/English), Universal Nonverbal Intelligence Test: Abbreviated Battery (UNIT), and Batería Woodcock-Muñoz: Pruebas de Approvechamiento-Revisada (Batería-R APR: Reading). Second, the criterion-related validity of the UNIT was assessed for participants of limited English proficiency (LEP) status as well as for participants not of LEP status. Third, analyses were conducted to examine the predictive relationship of language proficiency in Spanish and in English on reading achievement in the dominant language. Due to sample variability, descriptive statistics and bivariate correlations of the measures of interest will be discussed in conjunction with the associated results.

Internal Consistency of Measures

With a random sample of LEP participants, the score reliability of the WMLS Spanish and English versions, the UNIT, and the Batería-R APR: Reading was explored using Cronbach's coefficient alpha as an index of internal consistency. The goal of these analyses was to establish item homogeneity (i.e., performance consistency across items) as well as to assess the effects of content sampling error and sources of measurement error such as administration errors, scoring errors, and guessing (Crocker & Algina, 1986). Because of its applicability to various types of tests, Cronbach's coefficient alpha is the most commonly used formula for establishing internal consistency (Pedhazur &

Schmelkin, 1991; Thompson, 2003). Descriptive statistics of the measures applied in this analysis are reported in Table 4.

Table 4
Descriptive statistics of measures examined for internal consistency (n=24)

	Range					
	Minimum	Maximum	Mean	S.D.	Skewness	Kurtosis
WMLS-Spanish	86	140	102.71	12.17	1.13	2.38
WMLS-English	38	127	76.63	17.79	0.38	1.45
UNIT FSIQ	60	117	92.08	15.81	- 0.31	- 0.78
Batería-R APR: Reading	92	145	112.42	11.97	0.73	1.07

As revealed in Table 5, the internal consistency estimates using Cronbach's coefficient alpha were .92 for the WMLS-Spanish, .97 for the WMLS-English, .84 for the UNIT, and .87 for the Batería-R APR: Reading.

Table 5
Cronbach's coefficient alpha reliability coefficients

	N of items	Cronbach's coefficient alpha
WMLS-Spanish	207	.92
WMLS-English	206	.97
UNIT FSIQ	45	.84
Batería-R APR: Reading	101	.87

Criterion-Related Validity of UNIT

Analyses were conducted to estimate the concurrent validity of the UNIT with two measures of achievement (i.e., Batería-R APR, WJ-III ACH), each administered in Time 1. Based upon previous reviews, scores from the UNIT were hypothesized to demonstrate a statistically significant relationship with reading scores from the Batería-R APR as well as the WJ-III ACH. Table 6 displays the descriptive statistics of scores on the UNIT and on norm-referenced measures of reading achievement.

Table 6
Descriptive statistics to estimate concurrent validity of UNIT

Group	Minimum	Maximum	Mean	S.D.	Skewness	Kurtosis
LEP (n=56)						
UNIT FSIQ	60	132	94.30	14.53	0.09	0.28
Batería-R APR: Reading	61	156	117.25	22.46	- 0.59	- 0.29
non-LEP (n=89)						
UNIT FSIQ	63	126	94.54	13.01	0.15	- 0.18
WJ-III: Reading	69	125	96.79	12.43	- 0.20	- 0.41

Pearson-product moment correlations were calculated to examine the concurrent validity of the UNIT. Results did not reveal a statistically significant relationship between scores from the UNIT and the Batería-R APR (r=-.06) for the present sample of Time 1 LEP participants. However, for the sample of non-LEP Time 1 participants, scores from the UNIT were found to demonstrate a statistically significant relationship with scores from the WJ-III ACH (r=.30, p<.01).

To evaluate the predictive validity of the UNIT, two separate simple linear regression analyses were conducted for a group of LEP students administered the TAKS in Spanish as well as for a group of non-LEP students administered the TAKS in

English. Scores from the UNIT were posited to be a statistically significant predictor for performance on the TAKS in both languages. Descriptive statistics of the UNIT and the TAKS: Reading for both groups are depicted in Table 7.

Table 7
Descriptive statistics to estimate predictive validity of UNIT

Group	Minimum	Maximum	Mean	S.D.	Skewness	Kurtosis
LEP (n= 51)						
UNIT FSIQ	69	132	93.59	14.95	0.18	- 0.19
TAKS: Reading in Spanish	6	36	24.71	7.01	- 0.69	- 0.18
non-LEP (n=77)						
UNIT FSIQ	75	126	96.92	11.99	0.47	- 0.43
TAKS: Reading in English	16	36	29.68	4.81	- 1.13	0.66

Results of the simple regression analysis examining the predictive validity of the UNIT for a group of LEP students demonstrated that UNIT accounted for less than one percent of the variance (R^2= <.00, t=-0.13; *p*=.90) on the reading portion of the TAKS in Spanish. Similarly, the UNIT accounted for approximately one percent of the variance in the English version of the TAKS: Reading (R^2=.01, t=0.78; *p*=.44). The UNIT did not predict a statistically significant portion of the variance in either group administered the TAKS: Reading. Results of this analysis may be found in Table 8.

Table 8
Predictive validity of the UNIT

Group	R^2	Adj. R^2	B	β	t	Sig.
LEP (n=51)						
UNIT FSIQ	< .00	- .02	- 0.01	- 0.02	- 0.13	.90
non-LEP (n=77)						
UNIT FSIQ	.01	- .01	0.04	0.09	0.78	.44

Language Proficiency in the Prediction of Reading Achievement

Hierarchical multiple regression analyses were employed to explore the role of language proficiency in the dominant language (L1) and in the subordinate language (L2) in the prediction of reading achievement in L1. The hypothesis for this analysis suggested that the inclusion of proficiency in L2 with proficiency in L1 would contribute to the prediction of reading achievement in the L1. Although CALP scores were appropriately used to determine dominance and, thus, the language in which the achievement measure was administered, the WMLS Broad Ability standard scores in Spanish and in English were applied in the following analyses. CALP scores provide "cutoff points" for five levels of proficiency that are more accessible to practitioners. The WMLS Broad Ability standard scores provide a greater range of possible scores that produce more reliable, statistically sound results. Nevertheless, as indicated in the WMLS manual, the Broad Ability scores obtained in English and in Spanish are still a "broad-based measure of language ability" (Woodcock & Muñoz-Sandoval, 2001, p. 8). Similar to WMLS CALP scores, the Broad Ability scores provide an estimate of a child's expressive vocabulary, verbal reasoning, reading identification, and writing skills that theoretically comprise cognitive-academic language proficiency. Therefore, for the purposes of this study, the WMLS Broad Ability standard scores were deemed more appropriate for inclusion in the following statistical analyses.

The hierarchical multiple regression analyses were conducted separately for each language in which achievement measure was administered. For each analysis, the reading achievement standard score in the L1 was entered as the dependent variable.

The L1 Broad Ability standard score was entered the independent variable in Model 1 and the L2 Broad Ability standard score was entered in Model 2. For example, scores from the Batería-R APR were applied as the dependent variable for Spanish-dominant participants. Scores from the WMLS-Spanish (L1) were entered into the independent variable block of Model 1. Scores from the WMLS-English (L2) were entered into the independent variable block of Model 2.

Norm-Referenced Measures of Reading Achievement

Descriptive statistics for the analyses involving the norm-referenced measures of reading achievement are displayed in Table 9. In addition, bivariate correlations for these measures administered in Spanish or in English are respectively revealed in Tables 10 and 11.

Table 9
Descriptive statistics of norm-referenced measures of reading achievement

	Range					
Measure	Minimum	Maximum	Mean	S.D.	Skewness	Kurtosis
Spanish (n=79)						
WMLS-Spanish	76	122	100.75	9.73	- 0.07	- 0.53
WMLS-English	32	110	73.38	14.53	0.03	0.43
Batería-R APR	87	131	109.62	10.91	- 0.04	- 0.58
English (n=14)						
WMLS-Spanish	32	106	70.36	23.39	- 0.15	- 0.91
WMLS-English	42	107	80.71	19.66	- 0.65	- 0.39
WJ-III ACH	39	114	81.00	22.86	- 0.64	- 0.51

Table 10
Bivariate correlations of scores from WMLS-Spanish, WMLS-English, and Batería-R APR,

	WMLS-Spanish	WMLS-English	Batería-R APR
WMLS-Spanish	1.00		
WMLS-English	.51**	1.00	
Batería-R APR	.69**	.36**	1.00

** $p<.01$.

Table 11
Bivariate correlations of scores from WMLS-Spanish, WMLS-English, and WJ-III ACH

	WMLS-Spanish	WMLS-English	WJ-III ACH
WMLS-Spanish	1.00		
WMLS-English	.79**	1.00	
WJ-III ACH	.76**	.98**	1.00

** p<.01.

Results from the hierarchical multiple regression analyses that employed norm-referenced measures of reading achievement in the dominant language are located in Table 12. For reading achievement in Spanish, as assessed via the Batería-R APR, language proficiency in Spanish and in English accounted for approximately 48% of the variance. However, when examining the specific role of the second language, the addition of English proficiency did not account for a statistically significant portion of the explained variance (ΔR^2 = <.00, β=0.01, t = 0.12; p = .907). For this regression analysis, the squared semipartial correlation for English proficiency was .0001, thereby suggesting that English proficiency uniquely contributed less than 1% to reading achievement in Spanish.

Somewhat similar results emerged with the WJ-III ACH in English. While a large majority of the variance in reading achievement in English was accounted for by language proficiency in English and in Spanish in Model 2 (i.e., R^2 = .97, Adjusted R^2 = .96), language proficiency in Spanish was not a statistically significant contributor (ΔR^2 = <.00, t= -0.37; p=.720). As with the prior analysis, proficiency in L2 independently contributed less than 1% of the explained variance in reading achievement in L1 (i.e., squared semipartial correlation = .0004).

Table 12
Hierarchical multiple regression analyses predicting norm-referenced reading achievement

	R^2	Adj. R^2	B	β	t	Sig.
Batería-R APR: Reading in Spanish (L1) (n=79)						
Model 1	.48	.47				
Spanish language proficiency			0.77	0.69	8.36	.000**
Model 2	.48	.46				
Spanish language proficiency			0.77	0.68	7.07	.000**
English language proficiency			0.01	0.01	0.12	.907
WJ-III ACH: Reading in English (L1) (n=14)						
Model 1	.97	.97				
English language proficiency			1.14	0.98	18.89	.000**
Model 2	.97	.96				
English language proficiency			1.17	1.01	11.41	.000**
Spanish language proficiency			- 0.03	- 0.03	- 0.37	.720

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

Criterion-Referenced Measures of Reading Achievement

The contribution of language proficiency in L1 and L2 was further explored in the prediction of reading achievement as measured by the TAKS. Descriptive statistics for the variables of interest within this analysis are reported in Table 13. Bivariate correlations with the TAKS: Reading in Spanish appear in Table 14. Bivariate correlations with the TAKS: Reading in English are shown in Table 15.

Table 13
Descriptive statistics of criterion-referenced measures of reading achievement

	Range					
Version of TAKS	Minimum	Maximum	Mean	S.D.	Skewness	Kurtosis
Spanish (n=54)						
WMLS-Spanish	83	114	100.69	8.43	- 0.24	- 0.76
WMLS-English	38	99	71.72	11.88	- 0.16	0.24
TAKS: Reading	6	36	25.24	6.51	- 0.76	0.18
English (n=12)						
WMLS-English	77	124	92.17	13.20	1.26	2.22
WMLS-Spanish	34	140	82.00	28.46	0.38	0.40
TAKS: Reading	11	36	28.08	7.34	- 1.44	1.83

Table 14
Bivariate correlations of WMLS-Spanish, WMLS-English, and TAKS: Reading in Spanish

	WMLS-Spanish	WMLS-English	TAKS: Reading in Spanish
WMLS-Spanish	1.00		
WMLS-English	.44**	1.00	
TAKS: Reading in Spanish	.44**	.09	1.00

** $p<.01$.

Table 15
Bivariate correlations of WMLS-English, WMLS-Spanish and TAKS: Reading in English

	WMLS-English	WMLS-Spanish	TAKS: Reading in English
WMLS-English	1.00		
WMLS-Spanish	.83**	1.00	
TAKS: Reading in English	.69*	.74**	1.00

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

Results of the hierarchical multiple regression analyses with scores from the TAKS are depicted in Table 16. Although language proficiency in Spanish was found to be a statistically significant contributor to TAKS reading achievement in Spanish (β=0.50, t = 3.56; p = .001), the addition of English proficiency was not found to be beneficial in predicting reading achievement in Spanish (ΔR^2 = .013, β=-0.13, p=.366).

While L1 proficiency (Spanish) uniquely accounted for approximately 20% of the variance in this analysis, L2 proficiency (English) alone contributed approximately 1% of the variance.

Regarding reading performance on the TAKS in English, language proficiency in L1 and L2 together explained approximately 57% of the variance ($\Delta R^2 = .095$). However, when examined within the same model, neither L1 nor L2 emerged as statistically significant contributors to reading achievement in English. Although not statistically significant, interesting to note is that Spanish proficiency explained more unique variance (squared semi-partial correlation = .09) than English proficiency (squared semi-partial correlation = .02) in this model.

Table 16
Hierarchical multiple regression analyses predicting criterion-referenced reading achievement

	R^2	Adj. R^2	B	β	t	Sig.
TAKS: Reading in Spanish (L1) (n=54)						
Model 1	.19	.18				
Spanish language proficiency			0.05	0.44	3.53	.001**
Model 2	.21	.18				
Spanish language proficiency			0.06	0.50	3.56	.001**
English language proficiency			- 0.01	- 0.13	- 0.91	.366
TAKS: Reading in English (L1) (n=12)						
Model 1	.47	.42				
English language proficiency			0.05	0.69	3.00	.013*
Model 2	.57	.47				
English language proficiency			0.02	0.23	0.59	.570
Spanish language proficiency			0.02	0.55	1.41	.193

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

CHAPTER V

SUMMARY AND CONCLUSIONS

With a sample of Hispanic, limited English proficient (LEP) students, the purpose of the present study was to explore the relationship of general intelligence and reading achievement as well as the relationship of language proficiency in two languages and reading achievement. Preliminary analyses of this study suggest that the internal consistency of scores obtained on each of the examined measures by the present sample of LEP students was satisfactory. More specifically, the score reliability estimate of .92 of the WMLS-Spanish is considered high and thus within acceptable parameters. As the first known study to evaluate the internal consistency of the WMLS-Spanish, this finding is particularly notable. The present estimate of internal consistency on the WMLS-English is also high (r=.97) and is comparable to the estimates of internal consistency reported in the manual.

Although the developers of the UNIT took numerous steps to ensure fairness for clinical and exceptional samples, the internal consistency estimates of the UNIT with a sample of linguistically diverse participants were not reported within the manual. Nevertheless, the internal consistency of the Abbreviated version of UNIT with the present sample of Hispanic, LEP participants is considered adequate (r=.84). The present sample of LEP participants also performed consistently across items of the Batería-R APR: Reading (r=.87). Similar to the WMLS-Spanish, this investigation was one of the first known analyses to evaluate the score reliability of the Batería-R APR.

Regarding concurrent validity of the UNIT, Pearson-product moment correlations demonstrated a statistically significant relationship between scores on the UNIT and the WJ-III ACH with a sample of Hispanic, non-LEP students. These results suggest that the concurrent validity of the UNIT with the WJ-III ACH with a sample of Hispanic, non-LEP students is satisfactory. However, a statistically significant relationship between scores on the UNIT and the Batería-R APR for the present sample of Hispanic, LEP students did not emerge. These results did not corroborate previous reviews that advocate the use of the UNIT with students of diverse linguistic backgrounds (Bracken & McCallum, 1998; Fives & Flanagan, 2002). One plausible explanation is the great variability in scores on the Batería-R APR for this sample. The mean Batería-R APR standard score was 117.25 while the standard deviation was 22.46. Given this wide, atypical variability in scores, results from this testing may not accurately reflect performance within the general population and therefore may not be expected to correlate with scores from the UNIT.

The UNIT was not shown to be a useful tool in the prediction of reading achievement on a criterion-referenced, high-stakes test for a sample of LEP students or for a sample of non-LEP students. More specifically, a statistically significant relationship was not found among scores on the UNIT for the group of LEP third grade students administered the TAKS in Spanish or for the group of non-LEP third grade students administered the TAKS in English. In fact, the UNIT accounted for less than one percent of the variance for each version of the TAKS: Reading. Results of this examination diverge from previous research that concluded that the UNIT demonstrated

adequate predictive validity with scores from a state-administered measure of reading achievement (Williams & McCallum, 1995).

Although scores approached normality with regard to the bell curve, the lack of correlation among scores from the UNIT with reading achievement for this sample of Hispanic LEP students is of particular concern. The UNIT is often recommended and employed in the assessment of general intelligence of low performing LEP students referred for a psychoeducational evaluation. Results of the study determined that this instrument might not be appropriate for children with similar characteristics. Two hypotheses emerge concerning the negligible relationship. One, this limited correlation may be due to the nonverbal format in which the UNIT is administered. Results of this study are similar to results of a previous study that examined the predictive validity of the TONI-3, a purported measure of nonverbal cognitive reasoning (Laija, 2001). Therefore, the UNIT may in fact be a measure of nonverbal intelligence, not a nonverbal measure of general intelligence as purported. If this inference is true for this sample, then the utility of the UNIT to predict academic achievement is limited.

Another reason for these findings may be due to UNIT subtests administered. The present study employed the Abbreviated version of the UNIT, which is comprised of the Symbolic Memory and Cube Design subtests. A previous study found that although the contribution of Cube Design with reading achievement was statistically significant, Analogic reasoning was best predictor of reading achievement on a state-administered, norm-referenced measure of academic achievement. Therefore, despite other reviews and research, the employed subtests of the Abbreviated version of the

UNIT may not be the best predictor of reading achievement as measured by the TAKS, in either language.

A third explanation may be the restriction of range due to the selection of participants who scored below the median on a test of literacy. This selection reduced the variability of observed scores, which, in turn, may have limited the ability to detect statistically significant differences.

With regard to the relationship of language proficiency and reading achievement, language proficiency in L1 was highly correlated with language proficiency in L2 for all examined samples ($r>.51$, $p<.01$). This relationship provides further support for the common underlying proficiency (CUP) model, as prescribed by Cummins. Moreover, language proficiency in L1 was found to demonstrate a statistically significant relationship with reading achievement in L1, as assessed by both norm-referenced and criterion-referenced measures. Language proficiency in L2 also revealed a statistically significant correlation with reading achievement in L1, as measured by the Batería-R APR in Spanish, the WJ-III ACH in English, and the TAKS in English. Therefore, language proficiency in either L1 or L2 appears to be related to reading achievement in L1.

Nevertheless, contrary to initial speculation, language proficiency in L2 did not produce a statistically significant contribution to the explained variance of reading achievement in L1 for this sample of LEP students when also considering proficiency in L1. As a result, with this sample, language proficiency in L2 did not emerge as a beneficial, independent contributor to the relationship between language proficiency in

L1 and reading achievement in L1. These results may be expected, however, considering that most children in this sample had unfortunately not reached fluency in the second language. Specifically, approximately 9% of participants administered the Batería-R APR in Spanish achieved a CALP level of four or greater on the WMLS-English. Of the participants administered the WJ-III ACH in English, approximately 14% obtained a CALP level of four or greater on the WMLS-Spanish. Less than 2% of participants administered the TAKS-Spanish demonstrated a CALP level of four or higher on the WMLS-English. And, three out of twelve participants given the TAKS-English attained a CALP level equal to or above four in Spanish.

Based upon these findings, these participants may not be considered truly "bilingual" given that fluency in L2 was not obtained. Consequently, these results may provide further support for the threshold hypothesis that suggests a minimum threshold must be attained in the second language in order for positive outcomes to be realized (Cummins, 1984). The majority of participants in this study did not reach a minimum threshold or fluency in the second language and thus proficiency in L2 was not shown to contribute to predictive relationship between language proficiency in L1 and reading achievement in L1.

Another plausible explanation is the positive, statistically significant correlation or collinearity demonstrated between the two predictor variables (i.e., L1 proficiency, L2 proficiency). Therefore, although in different languages, these variables may have been measuring the same characteristic and may have shared variance with the reading

achievement. The variance that L2 contributed to the relationship may have already been accounted for by L1.

In summary, results from the preliminary analyses suggest that the observed measures demonstrated satisfactory content homogeneity and item quality with the employed sample. Hence, examinees of similar backgrounds may perform consistently across items within each test. This finding is especially advantageous given that little to no previous studies have examined the internal consistency of measures administered in Spanish.

Based upon the results of the present study, however, practitioners should use caution when using the UNIT: Abbreviated battery to predict future reading achievement on high-stakes testing for Hispanic English-speaking children and Spanish-speaking children. The UNIT: Abbreviated was not found to demonstrate a statistically significant relationship with reading achievement on either measure. As a result, the UNIT: Abbreviated may not be as comparable to the UNIT: Standard when used to predict reading achievement with a sample of low-performing Hispanic LEP and non-LEP students. Secondly, this battery may not be a comprehensive measure of general intelligence as originally purported. These inferences, however, are directed toward future samples of low performing Hispanic students and may not generalize to the larger population.

Results from the analyses conducted to explore the contribution of proficiency in L1 and L2 to reading achievement in L1 demonstrated a relationship between L1 proficiency and L2 proficiency. Moreover, a relationship was demonstrated between L2

proficiency and reading achievement in L1 as assessed by three measures of achievement. However, L2 proficiency did not emerge as an independent contributor to the relationship between L1 and reading achievement in L1. Although not supportive, these results do not refute the original hypothesis. As discussed, higher proficiency in both languages is believed to be related to higher levels of achievement. Therefore, as shown by the present results, when participants do not achieve CALP in L2 then proficiency in L2 is not predictive of reading achievement in L1.

Limitations and Future Directions

As shown, an unexpected limitation that emerged following the analysis of these results was the use of the UNIT: Abbreviated as a measure of general intelligence for this sample of low performing Hispanic students. The UNIT: Abbreviated was administered for this study given its high correlation with the standard battery of the UNIT and its purported likeness with the UNIT: Standard. However, a previous study suggested that Analogic Reasoning, which is not included in the administration of the UNIT: Abbreviated, was the best predictor of overall academic achievement, and especially in reading achievement (Williams & McCallum, 1995). Therefore, the UNIT: Abbreviated may have underestimated the relationship between the Full Scale IQ scores on the standard battery and academic achievement. This suggestion, however, should be considered within the scope of the selection criteria for the presently examined sample and should be ascertained by additional research. Furthermore, the psychometric properties of the UNIT were of interest in the present study and provide favorable information regarding the internal consistency of the UNIT: Abbreviated. Nevertheless,

future research should explore the standard battery of the UNIT in addition to other measures of general intelligence when examining the predictive relationship of intelligence and reading achievement for culturally and linguistically diverse students.

With regard to language proficiency and reading achievement, another limitation that was detected was the small percentage of participants who achieved CALP in the second language. Therefore, future research should consider examining the independent contribution of L2 proficiency to L1 reading achievement with a sample who demonstrates greater variability in L2 CALP. This variability may be evident in populations who have participated in academic programs that promote proficiency in two languages over a longer period of time (i.e., five to seven years). Additionally, a larger sample size will increase the likelihood of finding statistically significant results.

The goal of the present study was to examine the relationship of language proficiency and intelligence with reading achievement of a sample of low performing, Hispanic students who were identified as LEP. However, the generalizability of these results to students of other ethnic and linguistic backgrounds may be limited. Similarly, findings may vary when similar statistical methods are applied to students who demonstrate higher achievement. Future studies, therefore, may wish to include students of other ethnic and linguistic backgrounds who demonstrate a broader range of academic achievement.

REFERENCES

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*, 4-14.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Psychological Association (APA). (2002). *Ethical principles of psychologists and code of conduct*. Washington, DC: Author.

Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2002). English-language learner representation in special education in California urban school districts. In D. J. Losen & G. Orfield (Eds.), *Racial inequity in special education* (pp. 117-136). Cambridge, MA: Harvard Education Press.

Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Council for Exceptional Children, 71*, 283-300.

Bainter, T. R., & Tollefson, N. (2003). Intellectual assessment of language minority students: What do school psychologists believe are acceptable practices? *Psychology in the Schools, 40*, 599-603.

Barona, A., & Santos de Barona, M. (1987). A model for the assessment of limited

English proficient students referred for special education services. In S. H. Fradd & W. J. Tikunoff (Eds.), *Bilingual education and bilingual special education* (pp. 183-210). Boston: College-Hill Press.

Benjamin, L. T., & Baker, D. B. (2004). *From séance to science: A history of the profession psychology in America*. Belmont, CA: Wadsworth/Thomson Learning.

Bracken, B., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test: Examiner's manual*. Itasca, IL: Riverside.

Bracken, B. A., & Naglieri, J. A. (2003). Assessing diverse populations with nonverbal tests of general intelligence. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (2nd ed., pp. 243-274). New York: Guilford.

Braden, J. P., & Athanasiou, M. S. (2005). A comparative review of nonverbal measures of intelligence. In D. P. Flanagan & P. L. Harrison (Eds), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 557-577). New York: Guilford.

Braden, J. S., DiMarino-Linnen, E., & Good, T. L. (2001). Schools, society, and school psychologists: History and future directions. *Journal of School Psychology, 39,* 203-219.

Cloud, N. (1991). Educational assessment. In E. V. Hamayan & J. S. Damico (Eds.), *Limiting bias in the assessment of bilingual students* (pp. 219-246). Austin, TX: Pro-Ed.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, & Winston.

Cummins, J. (1984). *Bilingualism and Special Education: Issues in assessment and pedagogy*. San Diego, CA: College-Hill.

Del Vecchio, A., Guerrero, M. (1995) Handbook of English language proficiency tests. Retrieved July 14, 2006 from http://www.ncela.gwu.edu/pubs/eacwest/elptests.htm.

Diana v. State Board of Education, Civ. Act. No. C-70-37 (N. D. Cal., 1970, *further order*, 1973).

DiCerbo, K. E. (2003). English language proficiency and tests of intelligence and academic achievement. *Dissertation Abstracts International, 64* (3-A), pp. 793.

Education for All Handicapped Children Act of 1975 (Public Law No. 94-142), renamed the Individuals with Disabilities Education Act in 1990, 20 U.S.C. Chapter 33.

Fagan, T. K. (1992). Compulsory schooling, child study, clinical psychology, and special education: Origins of school psychology. *American Psychologist, 47*, 236-243.

Figuero, R. A. (1990). Best practices in the assessment of bilingual children. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 2, pp. 93-106). Washington, DC: National Association of School Psychologists.

Fives, C. J., & Flanagan, R. (2002). A review of the Universal Nonverbal Intelligence Test (UNIT): An advance for evaluating youngsters with diverse needs. *School Psychology International, 23*(4), 425-448.

García, E. E. (1992). "Hispanic" children: Theoretical, empirical, and related policy issues. *Educational Psychology Review, 4*, 69-93.

García-Vázquez, E., Vázquez, L. A., López, I. C., & Ward, W. (1997). Language proficiency and academic success: Relationships between proficiency in two languages and achievement among Mexican American students. *Bilingual Research Journal, 21*, 334-347.

Gersten, R., & Woodward, J. (1994). The language-minority student and special education: Issues, trends, and paradoxes. *Exceptional Children, 60*, 310-322.

Guadalupe Organization v. Tempe Elementary School District No. 3, Civ. No. 71-435 (D. Ariz., 1972).

Hamayan, E. V., & Damico, J. S. (1991). Developing and using a second language. In E. V. Hamayan & J. S. Damico (Eds.), *Limiting bias in the assessment of bilingual children* (pp. 39-75). Austin, TX: Pro-Ed.

Holtzman, W. H., & Wilkinson, C. Y. (1991). Cognitive ability. In E. V. Hamayan & J. S. Damico (Eds.), *Limiting bias in the assessment of bilingual children* (pp. 247-280). Austin, TX: Pro-Ed.

Improving America's Schools Act (Pub. L. No. 103-382), 108 Stat. 3518 (1994).

Individuals with Disabilities Education Act (Pub. L. No. 101-476), 20 U.S.C Chapter 33 (1990). Amended by Pub. L. No. 105-17 in June, 1997. Regulations appear at 34 C.F. R. Part 300.

Jacob, S., & Hartshorne, T. S. (2003). *Ethics and law for school psychologists* (4th ed.). Hoboken, NJ: John Wiley & Sons.

Jiménez, S. (2002). An analysis of the reliability and validity of the Universal Nonverbal Intelligence Test (UNIT) with Puerto Rican children (Doctoral dissertation, Texas A&M University, 2001). *Dissertation Abstracts International, 62*, 5424.

Kamphaus, R. W. (2001). *Clinical assessment of child and adolescent intelligence* (2nd ed.). Boston: Allyn & Bacon.

Kindler, A. L. (2002). *Survey of the states' limited English proficient students and available educational programs and services 1999-2000 summary report.* Washington, DC: National Clearinghouse for English Acquisition and Language Instruction Educational Programs.

Laija, W. (2001). The role of cognitive academic language proficiency and nonverbal cognitive reasoning on reading growth with limited English proficient students (Doctoral dissertation, Texas A&M University, 2000). *Dissertation Abstracts International, 61*, 2594.

Laija-Rodríguez, W., Ochoa, S. H., & Parker, R. (2006). The crosslinguistic role of cognitive academic language proficiency on reading growth in Spanish and English. *Bilingual Research Journal, 30*, 87-106.

Lopez, E. C. (1997). Cognitive assessment of LEP and bilingual children. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (p. 503-516). New York: Guilford.

Macias, R. F. (1998). *Summary report of the survey of the states' limited English proficient students and available educational programs and services 1996-1997.* Washington, DC: National Clearinghouse for Bilingual Education.

MacSwan, J., & Rolstad, K. (2005). Modularity and the facilitation effect: Psychological mechanisms of transfer in bilingual students. *Hispanic Journal of Behavioral Sciences, 27*, 224-243.

McCloskey, D., & Athanasiou, M. S. (2000). Assessment and intervention practices with second-language learners among school psychologists. *Psychology in the Schools, 37*, 209-225.

McLean, Z. Y. (1995). History of bilingual assessment and its impact on best practices used today. *New York State Association for Bilingual Education Journal, 10*, 6-12.

National Association of School Psychologists (NASP). (2000). *Professional conduct manual*. Bethesda, MD: Author.

Ochoa, S. H. (2003). Assessment of culturally and linguistically diverse children. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (2nd ed., pp. 563-583). New York: Guilford.

Ochoa, S. H., Galarza, A., & Gonzalez, D. (1996a). An investigation of school psychologists' assessment practices of language proficiency with bilingual and limited-English-proficient students. *Diagnostique, 21*(4), 17-36.

Ochoa, S. H., Powell, M. P. & Robles-Piña, R. (1996b). School psychologists' assessment practices with bilingual and limited-English-proficient students. *Journal of Psychoeducational Assessment, 14*, 250-275.

Ochoa, S. H., Rivera, B. D., & Ford, L. (1997). An investigation of school psychology

training pertaining to bilingual psycho-educational assessment of primarily Hispanic students: Twenty-five years after *Diana v. California*. *Journal of School Psychology, 35*(4), 329-349.

Ortiz, S. O. (2002). Best practices in nondiscriminatory assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 1321-1336). Washington, DC: National Association of School Psychologists.

Pedhazur, E. J. & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum.

Pray, L. (2005). How well do commonly used language instruments measure English oral-language proficiency? *Bilingual Research Journal, 29*, 387-409.

Rhodes, R. L, Ochoa, S. H., & Ortiz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide.* New York: Guilford.

Roid, G. H., & Miller, L. J. (1997). *Leiter International Performance Scale – Revised*. Wood Dale, IL: Stoelting.

Sánchez, G. I. (1932). Group differences and Spanish-speaking children: A critical review. *Journal of Applied Psychology, 16*, 549-558.

Sánchez, G. I. (1934). Bilingualism and mental measures: A word of caution. *Journal of Applied Psychology, 18*, 765-772.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). La Mesa, CA: Sattler.

Stetson, R., Stetson, E. G., & Sattler, J. M. (2001). Assessment of academic

achievement. In Sattler, J. M. (Ed.), *Assessment of children: Cognitive applications* (4th ed., pp. 576-609). La Mesa, CA: Sattler.

Thomas, W. P., & Collier, V. P. (1997). *School effectiveness for language minority students.* Washington, DC: National Clearinghouse for Bilingual Education.

Thompson, B. (2003). Understanding reliability and coefficient alpha, really. In B. Thompson (Ed.), *Score reliability: Contemporary thinking on reliability issues* (pp. 3-30). Newbury Park, CA: Sage.

U.S. Department of Education (2000). Twenty-Second Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act (chap. 2). Retrieved August 6, 2006 from http://www.ed.gov/about/reports/annual/osep/2000/chapter-2.doc

Valencia, R. R., & Suzuki, L. A. (2001). *Intelligence testing and minority students: Foundations, performance factors, and assessment issues*. Thousand Oaks, CA: Sage.

Williams, P. C., & McCallum, R. S. (1995). Predictive validity of the Universal Nonverbal Intelligence Test with achievement. *Assessment in rehabilitation and exceptionality, 2*, 221-232.

Woodcock, R. W., & Muñoz-Sandoval, A. F. (1993). *Woodcock- Muñoz Language Survey*. Chicago: Riverside.

Woodcock, R. W., & Muñoz-Sandoval, A. F. (1996). *Batería Woodcock-Muñoz-Revisada: Pruebas de Aprovechamiento.* Itasca, IL: Riverside.

Survey. Chicago: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *The Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside.

Woodcock, R. W., & Muñoz-Sandoval, A. F. (2001). *Woodcock- Muñoz Language* Survey Normative Update. Chicago: Riverside.

VITA

Charlotte Kennedy Jones
Lampasas Independent School District
Department of Special Education
207 West 9th Street
Lampasas, TX 76550

Education:

2006 Texas A&M University
College Station, Texas
Doctor of Philosophy, School Psychology
American Psychological Association - Accredited Program

2001 The University of Georgia
Athens, Georgia
Bachelor of Science, Psychology
Magna Cum Laude, with Honors

Pre-doctoral Internship:

2005-2006 Virginia Beach City Public Schools
American Psychological Association - Accredited Program

Publication:

Lease, M.A., **Kennedy, C.A.**, & Axelrod, J.L. (2002). Children's social constructions of popularity. *Social Development, 11,* 87-109.