2008

The Use and Interpretation of the Batería III With U.S. Bilinguals

Julie Esparza Brown
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), and the [Educational Methods Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Brown, Julie Esparza, "The Use and Interpretation of the Batería III With U.S. Bilinguals" (2008).
Dissertations and Theses. Paper 1677.
<https://doi.org/10.15760/etd.1676>

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

THE USE AND INTERPRETATION OF THE *BATERÍA III*
WITH U.S. BILINGUALS

by

JULIE ESPARZA BROWN

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

DOCTOR OF EDUCATION
in
EDUCATIONAL LEADERSHIP:
SPECIAL AND COUNSELOR EDUCATION

Portland State University
©2008

DISSERTATION APPROVAL

The abstract and dissertation of Julie Esparza Brown for the Doctor of Education in Educational Leadership: Special and Counselor Education were presented April 23, 2008, and accepted by the dissertation committee and the doctoral program.

COMMITTEE APPROVALS:

Leslie J. Munson, Chair

Steve Isaacson

Paula Stanovich

Ann Fullerton

Jose Padin
Representative of the Office of Graduate Studies

DOCTORAL PROGRAM APPROVAL:

Randy Hitz, Dean
Graduate School of Education

ABSTRACT

An abstract of the dissertation by Julie Esparza Brown for the Doctor of Education in Educational Leadership: Special and Counselor Education presented April 23, 2008.

Title: The Use and Interpretation of the *Bateria III* with U.S. Bilinguals

Within each classroom, many children excel academically while others struggle. Some students' difficulties are such that they require placement into educational programs different from grade level core and perhaps delivered outside of general education classrooms. For many, special education programs are the lifeline to reach their innate potential. For others misplaced into special education, their opportunities may be truncated. For the past 40 years (Dunn, 1968), disproportionate representation of minority children in some disability categories has been a problem. Educators commonly ask "Is a child's difficulties due to language differences or a learning disability?" One key area of confusion relates to the cognitive assessment of English language learner (ELL) students. While it is understandable that scores generated by English cognitive tests will not reflect an ELL child's true abilities, less is known about the appropriate use of native language (Spanish) cognitive assessments.

This study examined the performance of ELL general education students on the *Batería III: Pruebas de habilidades cognitivas*, a Spanish parallel to the Woodcock-Johnson III. The performance of 34 third- and fourth-grade participants, 16 who have received native language literacy development (NLD), and 18 who have received English language development, was compared to the normative sample's (monolingual Spanish speakers) General Intellectual Achievement (GIA) score, subtest and cluster scores.

ANOVA and *t* test analyses indicated both ELL groups scored significantly lower than the normative sample on GIA, short-term memory, long-term retrieval and crystallized intelligence and higher on auditory processing. Thus, the *Batería's* scores from these factors may underestimate the abilities of ELL students.

Correlation analyses were conducted to examine the relationships between language proficiency, acculturation and performance. No significant relationships were found. The last analysis compared subtest mean scores of the ELD group to Flanagan and Ortiz's (2001) predicted pattern of performance on the Culture-Language Interpretive Matrix (C-LIM) for diverse individuals when tested on the WJ-III. The score patterns of the ELD group did not follow the predicted pattern. A new arrangement of the *Batería's* subtests on the C-LIM is suggested. These findings highlight the need for more research to understand how ELL students perform on Spanish cognitive assessments.

DEDICATION

To my parents

who have always encouraged me,

Frank and Florence Esparza

To my wonderful husband,

Bradley L. Brown

To my precious children,

Nathan, Liana, and Danae

To my sister and her family,

Patrick, Maureen, Caylee and Nicholas Pierce

ACKNOWLEDGEMENTS

Completing a doctoral program is realizing a dream. It would not have been possible without the unconditional support of my husband, Brad, who took over many family duties. My children, Nathan, Liana, and Danae continue to be a great source of encouragement. Their pride and support gave me strength through this long process. My parents and sister and her family also have given their encouragement long distance. Thank you all for being always being there for me.

Dr. Randy Hitz, Portland State University and the Graduate School of Education provided substantial financial support to complete this study. I am humbled by your faith in me. Dr. Leslie Munson, my advisor, spent countless hours reading, commenting and editing. You provided just the perfect balance of support and prodding. Thank you for being my advisor and friend. My committee included an exemplary group of scholars that I wish to acknowledge for assuming this responsibility within their incredibly busy lives. Thank you Drs. Steve Isaacson, Paula Stanovich, Ann Fullerton and Jose Padin.

Working full-time and writing a dissertation, as well as raising a family, is a daunting task but was made possible by my dear colleague and friend, Lynda Pullen, who kept things running in our program when I was preoccupied with completing this dissertation. I appreciate you more than you know.

Finding a district in which to conduct this research was a time-consuming and challenging process. Thank you, Cindy Bauer, for your faith in me and willingness to help out and support this project, even during your busiest times. Thank you also to Claudia Aguilar and Maggie Rosario for your organization and dedication. You made the families and children feel comfortable.

Pati Sluys, you are a marvel. Your eye for detail is exactly what I needed to complete this dissertation and I am most grateful.

Thank you, Dr. Donaldo Macedo who made me start this process nine years ago. Your mentorship and friendship are a great support.

Finally, this research could not have been completed without my "madrinas de la doctorada," Connie Vidotto Ruff, Leslie Sharp, Cynthia Bogert, Martha Villegas-Gutierrez and Jana Swedo. You are the best bilingual school psychologists I know. Each of you is truly committed to equal educational opportunities for all children and are life-long learners. I am so very grateful for your friendship, kindness, brilliance, and being there whenever I needed you.

Conducting a quantitative study worried me at first but this process empowered me. I could not have done this without the help and gentle support of the brilliant, Tyrae Mahan.

I am truly blessed to have such a community of support!

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	x
LIST OF FIGURES	xii
CHAPTER	
I INTRODUCTION	1
Too Many Minority Students in Special Education Programs	4
Disproportionate Representation in Professional Judgment Categories	
What is a Learning Disability	5
Difference or Disability?	
Why Misplacement into Special Education is a Problem	6
Financial Impact on School Systems	
Explaining Disproportionality	9
The Intelligence Debate	
Lack of Resources	
Lack of Knowledge and Role Models	
Emphasis on High-Stakes Testing and Accountability	
Test Bias	
Equal Educational Opportunities	
The Need for New Paradigms	
Introduction to Dissertation Study	14

Glossary	14
II REVIEW OF THE LITERATURE	21
Development of the Bilingual and Special Education Fields.....	21
Bilingualism in the United States.....	22
Bilingual Education: Models and Their Impact on Student Achievement	23
Two-Way Bilingual Immersion Programs	
Late-Exit and Content ESL	
Early-Exit Bilingual and Content ESL	
Early-Exit Bilingual and Traditional ESL	
English-as-a-Second Language Pull-Out	
Major Findings of the Thomas and Collier Study	
Criticisms of Thomas and Collier's Study	
Legislation and Litigation Guiding the Development of Bilingual Education.....	30
Legislation	
Litigation	
The Early Development of the Learning Disabilities Field	32
The U.S. Foundation Period	
The Emergent Period	
The Solidification Period	
The Turbulent Period	
The Period of Reconceptualization	
Litigation and Legislation Guiding the Development of Special Education.....	38
Litigation on Behalf of Students with Disabilities	
Federal Legislation and the Education of Students with Disabilities	
Individuals with Disabilities Education Act	43

Individuals with Disabilities Education Improvement Act	43
The Role of Intelligence Testing in the Diagnosis of SLD	44
Beyond g Factor	
Beyond g	
Two General Factors	
A Modern Intelligence Theory: CHC	
Research on Processing Abilities and their Link to SLD	50
CHC Factors as Abilities – Not Learning Styles	
New Approaches to Eligibility Decision-Making	52
Approach One: Response to Intervention	
Approach Two: Identifying Patterns of Strengths and Weaknesses	
Reconciling the Approaches	55
Standardized Tests Performance and Diverse Populations	56
The Impact of Cultural Experiences	57
Acculturation	
Tests are Culture Loaded	
The Impact of Language Experiences	60
Native Language Standardized Tests	61
Current Practices in Testing ELL and Diverse Learners	62
Nondiscriminatory Assessment	64
Theoretical Framework for this Study	64
The Culture-Language Test Classifications and the Culture-Language Interpretive Matrix	
Research Questions	75

III	METHOD	77
	Participants.....	77
	Study Design	79
	Comparison Groups	
	Data Collections	81
	Assessment Instruments	82
	<i>Batería III</i> Woodcock-Muñoz Pruebas de Habilidades Cognitivas The Language Assessment Scales (LAS) – English and Spanish Acculturation Quick Screen	
	Assessment of Participants.....	87
	Data Analysis	88
	Ethical Considerations	88
IV	PRESENTATION OF DATA	91
	Descriptive Statistics for the Independent Variables	93
	GIA Scores English Proficiency Scores Spanish Proficiency Scores Acculturation Levels Mean Scores for Total Bilingual Group	
	Question One.....	103
	Question 1A Question 1B Question 1C Question 1D Question 1E Question 1F	

Question Two	112
Question 2A	
Question 2B	
Question 2C	
Question Three	113
Question 3A	
Question Four	116
Question 4A	
Question 4B	
Question 4C	
Question 4D	
V SUMMARY	118
Discussion	118
Native Language Standardized Test	119
Characteristics of the Research and Normative Sample	119
Discussion of Study Findings	122
Discussion of the Results	123
Analysis of Functioning in General Cognitive Ability	
Analysis of Functioning in Broad Cognitive Abilities	
and Processes	
Comprehension Knowledge (Gc)	
Long-Term Retrieval (Glr)	
Auditory Perception (Ga)	
Short-Term Memory (Gsm)	
Clusters with Significant Differences in Only One Group	135
Fluid Reasoning (Gf)	
Processing Speed (Gs)	
Visual Perception (Gv)	

Relationships Between Variables.....	139
Implications	
Using the C-LIM and the <i>Bateria III</i>	141
The Culture-Language Interpretive Matrix	
Implications	
Participants' Perceptions.....	144
Implications.....	146
Revisiting the Mandates of IDEIA 2004	
Policy Implications	
Implications for Practice	
Limitations	149
Suggestions for Future Research.....	150
Conclusion	151
REFERENCES.....	153
APPENDICES	
A INFORMED CONSENT	170
B TRANSPLANTED PARENT CONSENT	173
C STUDENT ASSENT SCRIPT	176
D STUDENT DATA SHEET	178
E AQS	180

LIST OF TABLES

TABLE	PAGE
1 Program Models for Educating ELL Students.....	28
2 CHC Factors and Achievement Areas.....	51
3 Characteristics of Hispanic Third and Fourth Grade Sample.....	79
4 Independent and Dependent Variables	80
5 Description of 14 Subtests for the <i>Batería III</i> Tests of Cognitive Abilities.....	84
6 Research Questions and Methods of Data Analysis	89
7 CHC Factors, Definition, and Subtests.....	92
8 GIA Scores by English Language Proficiency for NLD Group	93
9 GIA Scores by English Language Proficiency for ELD Group	94
10 Frequency of English Language Proficiency Scores NDL and ELD Groups.....	95
11 GIA Scores by Spanish Language Proficiency for NLD Group.....	96
12 GIA Scores by Spanish Language Proficiency for ELD Group	97
13 Frequency of Spanish Language Proficiency Scores NDL and ELD Groups.....	98
14 GIA Scores by Acculturation Level for NLD Group	99
15 GIA Scores by Acculturation Level for ELD Group.....	100

16	Subtest Scores for Total Bilingual Group (NLD + ELD).....	102
17	Means and Standard Deviations for All Groups.....	103
18	Means and Standard Deviations for Normative Sample and Bilingual Group	105
19	Significant Differences by Group on Gc Cluster and Subtests.....	128
20	Significant Differences by Group on Glr Cluster and Subtests.....	130
21	Significant Differences by Group on Ga Cluster and Subtests.....	132
22	Significant Differences by Group on Gsm Cluster and Subtests.....	134
23	Significant Differences by Group on Gf Cluster and Subtests	136
24	Significant Differences by Group on Gs Cluster and Subtests.....	137
25	Significant Differences by Group on Gv Cluster and Subtests	138

LIST OF FIGURES

FIGURE	PAGE
1 Cattell-Horn-Carroll Integrated Theory of Intelligence (CHC Theory).....	48
2 Culture-Language Interpretive Matrix (C-LIM) for the WISC-IV.....	68
3 Culture-Language Interpretive Matrix (C-LIM) for the WJ-III	69
4 Pattern of Expected Test Performance for Hispanic Children.....	70
5 General Guidelines for Expected Pattern of Test Performance	71
6 Differences in the 14 Subtest Scores for Norm, NLD, and ELD Groups.....	109
7 CHC Cluster Scores for Normative Sample and NLD Group	110
8 CHC Cluster Scores for Normative Sample and ELD Group	111
9 Culture-Language Interpretive Matrix (C-LIM) for the <i>Batería III</i>	114
10 C-LIM Predicted Scores for ELL Students Tested in English and Scores of ELD Participants Tested in Spanish on the <i>Batería III</i>	116
11 Five Characteristics of U.S. Bilingual that Impact Test Performance.....	121
12 Three Characteristics of Normative Sample that Impact Test Performance.....	121

CHAPTER I

INTRODUCTION

Velkommen. Bienvenidos. Merhba. Afio Mai. These are all words of welcome that may resonate in the hallways of any U.S. school today. Rapid diversification across the nation, from coast to coast and into the heartland, has sky rocketed the population of students for whom English is not the home language (Xu & Drame, 2008). Throughout this paper this group of students will be referred to as English Language Learners or ELL students. Today, ELL students represent 3.8 million or 11% of all public school students (Cosentino de Cohen & Clewell, 2007). Yet, observations in classrooms highlight a predominant majority culture teacher workforce. More importantly, data show that of 3,250,600 teachers in the United States in a 2003-2004 survey, only 4% hold an ESL credential (U.S. Department of Education [USDOE], 2004). Thus, there is a shortage of teachers with the experience and training to meet the specific needs of ELL students. When educators are unfamiliar with the challenges ELL students face in schools, the outcome may be unequal educational opportunities. The following vignette is one example.

Felipe was in the second grade and had lived in the United States for only 1 year after moving here with his family. At his school, English was the language of instruction. He, along with the other English Language Learner (ELL) students,

went to his English as a Second Language (ESL) class for 20 minutes each day to learn English. The ESL sessions were taught by an instructional assistant. Felipe's English, however, was not improving substantially nor was he making much academic gain, struggling to read and write simple words. In the middle of second grade Felipe was referred to the school's Pre-referral Team because his teacher was alarmed at his lack of progress and thought he may have a learning disability. This researcher was called in as an itinerant ELL/special education specialist to help the team conduct a psychoeducational evaluation. The monolingual school psychologist did not want to proceed with the evaluation because Felipe "hadn't been in the U.S. for 2 years." Therefore, the tests would not be valid. The psychologist missed the point completely, however, that "waiting 2 years" for an ELL student's English to improve before looking into a potential disability is a myth. All disabilities must be evident in the native language, not only the second language. In other words, a child with a disability will have that disability in their first, second, third language and so on. If a disability is only evident in a non-native language, the struggles may be a function of their developing language proficiency.

After reviewing Felipe's records, school work, and interviewing teachers, parents as well as Felipe, the multidisciplinary team (school psychologist, special education teacher, nurse, and speech/language specialist) decided to proceed with a psychoeducational evaluation. Standardized cognitive and academic assessments

were conducted in English and Spanish. The team, including this researcher serving as an itinerant bilingual special education specialist, determined that Felipe did not exhibit a learning disability. However, his Spanish (native language) seemed atypical as compared to his peers. This was puzzling since he had lived in Mexico until he was six and attended school; thus, his native language should have been age appropriate. There were no bilingual speech and language specialists available locally, language samples were taped and sent to bilingual speech and language specialists in another state. Their consensus was that it appeared that Felipe exhibited language concerns in his native language. Unfortunately, the multidisciplinary team members did not want to proceed because they felt he just needed "time to develop English." Again, they missed the point that Felipe may have had a communication disorder evident in his native language.

Fast forward to the end of the school year – Felipe was retained. The moral of the story is that educators, even assessment specialists, have limited expertise in determining cultural and linguistic differences from disabilities.

To understand the issues, the remainder of this chapter examines how ELL and other minority students have fared in general and special education programs over the last 40 years, identifies barriers to their academic success, and highlights critical areas for investigation.

Too Many Minority Students in Special Education Programs

Although it was first identified by Dunn in 1968, disproportionate representation of minority children in special education continues to plague the field. Dunn's seminal article defined disproportionate representation as the overrepresentation of African American students in special education classes for children with mental retardation. More recently, it was defined as "the representation of a particular group of students at a rate different than that found in the general population" (Gravois & Rosenfield, 2006, p. 42). Zhang and Benz (2006) reported that while racial and ethnic minorities represented 30.9% of the general school population in 2000, they represented 38.3% of the special education population. Today, disproportionality continues to be evident not only in the African American population but also in other ethnic groups and ELL.

Disproportionate Representation in Professional Judgment Categories

A recent report by the National Research Council (2002) stated that the highest incidences of disproportionality were in those categories based on professional judgment (a) Serious Emotional Disturbance, (b) Mental Retardation, and (c) Specific Learning Disability. Those categories are often referred to as a "judgment category" because eligibility teams use their expertise in interpreting data sources rather than a strict numerical formula (e.g., discrepancy formula) to make decisions. Professionals can choose specific test batteries which may be more likely to make some students eligible (Skiba et al., 2006). Nonjudgmental disability

categories are those in which diagnoses do not require inference on the part of professionals (O'Connor & Fernandez, 2006). In the nonjudgmental categories, all groups of students appear to be proportionately represented.

What is a Learning Disability?

The construct of learning disability (LD) is characterized as underachievement in an individual with normal or above normal intelligence (MacMillan & Forness, 1998). It is most commonly identified by a discrepancy between measured aptitude (intelligence) and achievement. In 1963, Kirk coined the term Specific Learning Disability (SLD). Subsequently, the category of SLD was established by a national task force in 1966 and was incorporated into federal law in the Learning Disabilities Act of 1969 (Ofiesh, 2006). Since that time, the federal definition of Specific Learning Disability has remained relatively unchanged and is

a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations, including conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. (Hargrove, 1982, p. 366)

Difference or Disability?

With the increase of students across the country who come from a home where English is not the first language, educators often have little training on educating ELL students. Teachers face daily struggles to help these students learn content while at the same time learn English. When students fail to make the gains

anticipated, the question *de jure* is whether the reason for an ELL student's academic difficulties are due to a language and cultural difference or a disability? Unfortunately, the answer is no clearer today than when Dunn (1968) penned his article.

The multitude of factors impacting the schooling of ELL and other minority children is complex and often poorly understood. As a result, it is common for multidisciplinary teams to find ELL students eligible for special education on the basis of invalid standard scores for this population and justify this by contending that at least they will get individual instruction (Salend, Duhaney, & Montgomery, 2002). While the team may have the student's best interest in mind, misplacement may not be an equitable decision.

Why Misplacement into Special Education is a Problem

Misplacement of ELL students into special education programs is particularly problematic in three significant ways. First, the stigmatizing effects of special education labeling are likely to have serious negative consequences on a student's learning career (Artiles & Ortiz, 2002). Stigma is defined as "a discrediting attribute assigned by society to those who differ in some manner from society's expectations, customs, and mores" (McHatton & Correa, 2005, p. 132). Stigma may then be operationalized as discrimination which for minority students add to the discrimination already faced due to their non-White status. Losen and

Orfield (2002) noted that while a label, particularly that of Specific Learning Disability, seems to benefit middle-class, white students by opening the gates to a variety of support services, a label disadvantages black and/or Latino(a) students because they are more likely to be placed in more restrictive and segregated settings.

Second, the segregation of special education service delivery may result in less access to core curriculum, a violation of a student's civil rights when a student does not have a true disability.

Where inability to speak and understand the English language excludes national origin-minority group children from effective participation in the educational program offered by a school district, the district must take affirmative steps to rectify the language deficiency in order to open its instructional program to these students. (Office of Civil Rights, 1970, p. 1)

Special education often excludes minority learners from the general education curriculum that profits Whites and defines standards. To illustrate, studies suggest that K-12 minority students in special education actually receive fewer and more technically oriented services in more segregated settings. (Reid & Knight, 2006, p. 19)

Further, a disability label can improve educational opportunities for some students, but for minority students, the label has a negative impact on postsecondary education (Reid & Knight, 2006).

Third, the quality of special education programs has been scrutinized by some researchers and educators as focusing solely on the remediation of basic skills and narrow curriculum (Artiles & Ortiz, 2002; Hosp & Reschly, 2003; Losen & Orfield, 2002). Criticism of the instructional programs in special education include

structured curriculum whose goals may not match the learner's needs and direct instruction resulting in passive learners (Heward, 2003; Kauffman, 1996; Poplin, 1988). "The teacher cannot act in a professional and intelligent manner, for much is forbidden, much prescribed, and much so rigid that personal initiative is impossible" (Heshusius cited in Heward, 2003, p. 189).

Financial Impact on School Systems

Beyond the student-level impact, there are broader issues that impact the educational system as a whole. In these times of fluctuating school funding and a shortage of resources, the economic impact must be considered. It is more expensive to provide special education services than to provide interventions within general education programs. Chambers, Shkolnik, and Perez (2003) found that the per pupil education expenditure for students receiving services under the category of Specific Learning Disability was \$10,558 or 1.6 times more than for a student in general education. From a financial standpoint alone, it is imperative that decisions regarding the identification of disabilities and placement of ELL students, as well as English-only students, be made as accurately as possible. When students are misdiagnosed, however, due only to linguistic and cultural differences, mislabeling becomes a question of civil rights (Office of Civil Rights, 1970). By placing ELL students in a more restrictive environment, such as a resource or self-contained special education room, students have less access to the core curriculum as well as fewer opportunities to interact with their peers in general education.

Explaining Disproportionality

Theories explaining disproportionality in the SLD category abound. First, hereditarian views claimed minority students have innate cognitive deficits (Artiles, Aguirre-Munoz, & Abedi, 1998; Herrnstein & Murray, 1994). A second theory was that the lack of resources in general education for providing interventions to struggling students led many general educators to refer students for special education assessment so that they could get some help. Third, eligibility teams with limited or no expertise in differentiating language and cultural differences from true disabilities have been implicated as well (Harry & Klingner, 2007). The push toward increased accountability and high stakes testing also created pressure to place some lower achieving students into special education to exempt them from testing and thus attenuating their schools' test scores (Skiba et al., 2006). Finally, the last theory discussed is that of test bias (Larry P. v. Riles, 1979; PASE v. Hannon, 1980). In the next few pages, each theory is discussed in more detail.

The Intelligence Debate

Hereditarian theories have raged for almost two centuries (Herrnstein & Murray, 1994; McDermott, Goldman, & Varenne, 2006). Researchers continue to debate whether intelligence is fixed and measurable or whether some groups have more of it than others.

Since about 1850, first in Europe and then in the United States, classifying human beings by mental ability, accurately or not, has been a politically rewarding activity. Those with power have placed others, usually the downtrodden, into ability and disposition groups that they cannot escape.

The practice has prospered even where the groupings are, as is usually the case, ill defined and, as is always the case in human cultures, arbitrary, in the revealing sense that groupings could be defined differently. (McDermott et al., 2006, p. 12)

Although the legitimacy of intelligence testing and classification can be questioned, the vast majority of professionals conducting special education assessments rely on them (Kranzler, 1997; Reschly, 1997). IQ tests, however, were the focus of important litigation concerning minority students and special education (Larry P. v. Riles, 1979; PASE v. Hannon, 1980). These cases are discussed in more detail in the next chapter.

Lack of Resources

Lack of resources for providing help to an increasingly diverse student population has caused much frustration for educators. Overwhelmed teachers often do not have the time to work individually with struggling ELL students or possess the expertise to meet their specific needs. For example, ELL and other minority students have been viewed as “culturally and linguistically deprived” and deemed ill-prepared for American curriculum (Crawford, 2004). Thus, moving them into another setting where a specialist could be responsible for their progress to provide relief for the general education teacher who must meet the needs of all the “mainstream” students occurs (Gottlieb, Alter, Gottlieb, & Wishner, 1994). Inequities in educational resources have documented that per pupil revenues for high poverty school are 89.4% of revenues for low poverty schools (Donovan & Cross, 2002).

Lack of Knowledge and Role Models

Educators' lack of knowledge about language acquisition and culturally relevant pedagogy may have a negative effect on the academic progress of ELL students. Because most mainstream teachers do not fully understand how culture and language differences impact acquisition of skills and content knowledge, educators may be inclined to consider the norm as middle-class (White) children upon which the other children are measured (O'Connor & Fernandez, 2006). Data also suggest that there will continue to be an insufficient number of minority teachers who have the potential to "radically alter current patterns of school staffing practices" (Donovan & Cross, 2002, p. 176). While it is crucial to have more teachers who hold an ESL credential and have the background to teach ELL students, it is also vital that students have role models who share their world views because of similar life experiences. In summary, the teaching and learning ecology must be sufficiently considered when an ELL or other minority student exhibits academic challenges (Harry, Klingner, Sturges, & Moore, 2002).

Emphasis on High-Stakes Testing and Accountability

The current emphasis on high-stakes testing and accountability also has been implicated as contributing to disproportionality. In a survey Skiba et al. (2006) found that

Both teachers and psychologists felt that accountability standards are expressed both in standardized testing and in local "no social promotion" policies may limit the school's ability or willingness to be sensitive to students' individual developmental needs... (p. 1435)

Test Bias

The final reason implicated for disproportionality is test bias. Much has been written suggesting that standardized tests are biased against minority groups (Harry & Klingner, 2007; MacMillan & Forness, 1998; Skiba, Knesting, & Bush, 2002). In addition, substantial litigation found test bias (Larry P. v Riles, 1979; PASE v. Hannon, 1980). Litigation surrounding both bilingual and special education is more thoroughly discussed in the following chapter. Standardized tests themselves have led to misplacements due to the confusion surrounding the validity of English-only and native language tests. Neither the norm samples of the English nor available Spanish language cognitive tests adequately represent U.S. bilinguals whose acculturation and language proficiency in their home language (L1) and English (L2) vary tremendously (Flanagan, Ortiz, & Alfonso, 2007). Eligibility teams are challenged to find appropriate instruments, determine which language in which to assess, and interpret test scores with little or no training in these areas.

Equal Educational Opportunities

To reiterate, disproportionate representation, both under-representation and over-representation, remains a significant and relevant issue today. Beyond the reasons discussed above, misplacement of ELL students into special education becomes an issue of Civil Rights, as mentioned earlier. Racial and ethnic minorities are protected from discrimination in the Equal Protection Clause of the 14th Amendment to the U.S. Constitution, Title VI of the Civil Rights Act of 1964, and

Section 504 of the Rehabilitation Act of 1973. In other words, disproportionality can serve as a proxy for unequal educational opportunities.

The Need for New Paradigms

Given the rapidly changing demographics currently in the United States, it will become increasingly urgent to identify models to accurately, appropriately, and legally assess ELL students. Assessment personnel must learn how to (a) examine a child's ecology (e.g., home, community, school), and their experiential and linguistic background; (b) administer appropriate and legally defensible assessments to administer; and (c) interpret test scores of available instruments in both English and Spanish and other data collected to make the most appropriate eligibility decisions possible (Flanagan & Ortiz, 2001).

This last area is the focus of this dissertation because little guidance and few frameworks exist. This study investigated the appropriateness of using the *Batería III* (Woodcock, Muñoz, McGrew, Mather, & Schrank, 2004; referred to throughout document as *Batería III*), a Spanish-language cognitive test, that was normed on monolingual Spanish speakers, with U.S. bilingual students. Then, a framework that accounts for the cultural loading and linguistic demand on each of the *Batería's* subtests was examined. The results suggest a guide for assessors in interpreting the results of the *Batería III* Cognitive Test when the examinee is a bilingual student in the United States.

Introduction to Dissertation Study

The remainder of this dissertation is presented in the subsequent chapters. Chapter 2 traces the roots of the fields of special and bilingual education, highlights seminal litigation that has guided their development, and summarizes the history and current status of intellectual testing and modern intelligence theories. Thus, the theoretical framework guiding this research is established. In addition, the research questions are proposed and their significance discussed. Chapter 3 describes the research methodology while chapter 4 presents the results of the study. To conclude, chapter 5 discusses the implications of the findings from this research as well as limitations and future research needs and directions. First, a glossary of terminology is presented.

Glossary

Bilingual Education: Bilingual education is the practice of teaching non-native English speaking students in either their native language or in structured English language development programs. Developed in the 1960s as an outgrowth of the Civil Rights Movement, such programs were intended to allow children to progress in subjects such as math, science and social studies while they learned English in a separate class.

Bilingual Group: The total research sample (NLD + ELD groups).

Circumstantial Bilingualism: refers to children that because of their circumstances such as entering an English-only school must learn a second

language to survive. In these cases, a student's home language generally does not have prestige.

Basic Interpersonal Communication Skills (BICS): Language proficiency needed to function in everyday interpersonal contexts and carry on a conversation in familiar face-to-face situations. Also called social language.

Bilingual Education Act (BEA): Enacted in 1968, this was the first piece of federal legislation related to language minority students. The purpose of the BEA was to provide federal funds to school districts to establish educational programs to serve students with limited English speaking ability. In 1968 the BEA merged into Title VII of the Elementary and Secondary Education Act (ESEA). This act allowed school district to provide bilingual education programs without violating segregation laws.

Cognitive Academic Language Proficiency (CALP): CALP refers to formal academic learning. This includes listening, speaking, reading, and writing about subject area content material. This level of language learning is essential for students to succeed in school. Students need time and support to become proficient in academic areas. This usually takes from 5 to 7 years. Recent research (Thomas & Collier, 1997) has shown that when a child has no prior schooling or has no support in native language development, it may take 7 to 10 years for ELLs to catch up to their peers.

Content English-as-a-Second Language (ESL) (Sheltered Immersion):

Content ESL models rely on the classroom teacher of record, who has been ESL trained, to teach the ELL child the entire curriculum while keeping the ELL child together with the rest of the students at all times. The ESL trained teacher employs ESL techniques to ensure that ELL students learn the academic curriculum while they are in the process of becoming English proficient. This model is used where the ESL student group represents many different home languages. Content ESL models are effective from K-12th grades and for all subject areas (Thomas & Collier, 1997).

Culture: Culture “is experienced in local, face-to-face interactions that are locally constrained and heterogeneous with respect to both ‘culture as a whole’ and the parts of the entire toolkit experienced by any given individual” (Cole & Engestrom, 1993, p. 15).

Dominant Language: The language that the student speaks most fluently. It is the language the child prefers to speak when given the choice. The dominant language can be situational in nature. For example, a child schooled only in English will ultimately become dominant in English academic language. However, the primary language may remain dominant in other social situations (e.g., church, community events).

Dual-Language Enrichment Programs (Two-Way Bilingual Enrichment Instruction): This model uses two languages to teach students (commonly at the

elementary level) the core curriculum. Participating students are equally divided between native English speakers and native speakers of the program's other language. The regular curriculum is divided into language groups with half of the content taught in the student's stronger language and the other half in the student's weaker language. Seminal research by Thomas and Collier (1997, 2002) found this model to be the most effective bilingual model.

Early-Exit Bilingual Models: This model is identical to the late-exit model. However, this model is designed to move ELL children from their native language to English in the first 3-years of primary grades. The language of instruction in the native language is directly proportional to the degree to which the student has acquired English language proficiency.

Elective Bilingualism: refers to when individuals choose to learn a second language.

English Language Development (ELD) Group: The portion of the research sample that received structured English language development as a specific content area and no native language instruction.

English-as-a-Second-Language (ESL) Pull-Out (Elementary): ESL "pull-out" programs continue to be the commonly used model but also the least effective (Thomas & Collier, 1997, 2002). ESL students are "pulled" out of their mainstream classrooms each day for approximately 30-45 minutes each day. In this model, a

teacher or paraprofessional provides students with focused assistance either in English language development or academic assistance.

English-as-a-Second-Language (ESL) (Secondary Level): ESL programs are designed to provide ELL students focused English language development while concurrently taking the regular curriculum in English. This can be accomplished by scheduling ELL children into an ESL classroom in which an ESL trained teacher develops a student's English language proficiency as quickly as possible utilizing a communicative-based approach.

English Language Learners (ELL): The terms Culturally and Linguistically Diverse Students (CLD), Limited English Proficient Students (LEP), and English Learners (EL) are used interchangeably. These terms are used as descriptors rather than categorical terms for a group of students whose home language is other than English. They may be U.S.- or foreign-born. Some may have immigrated prior to school age and others may have received formal education in their first language prior to immigrating. The level of English language proficiency will vary from Non-English proficient to Fully English proficient as measured by a language proficiency test. Thus, these students comprise a heterogeneous group. The term "limited English proficient" has the meaning given the term in Section 9101 of the Elementary and Secondary Education Act of 1965.

GIA: General Intellectual Ability (synonymous with Intelligence Quotient scores). On the *Batería III* the GIA scores are computer-generated scores and

represent common ability underlying intellectual performance. They are not simply an average of subtest scores but are obtained from first seven subtests.

Late-Exit Bilingual Education (Transitional Bilingual Education): This model is for English Language Learners (ELL) only. This is a transitional model designed to move ELL children from their native language to English over the 5 or 6 year period of their school's primary grades. This model relies on the teacher's instruction in the students' native language.

Native Language Development (NLD) Group: The portion of the research sample that received native language development in literacy since kindergarten.

Oral Language Proficiency: The level of an individual's ability to comprehend and speak a language.

People First Language: People First Language puts the person before the disability and describes what a person *has*, not what a person *is*.

Primary Language: The language that the student learns first and uses most frequently in the early stages of language development. It is the language of the home and the one children use to make and establish meaningful communicative relationships with their family members. The determination of primary language is best made through home language surveys and carefully conducted parent interviews.

Sequential Bilingualism: refers to when a second language is introduced after the first language is learned. In this case, research shows that there is usually a

silent period when the child is first exposed to a second language. Continued input in the first language should occur in order to decrease *language loss*. Language loss occurs when the first language is abruptly replaced by a second language and can have negative consequences for cognition and language.

Simultaneous Bilingualism: refers to when two languages are taught from infancy. Research shows that in the Simultaneous Bilingual model, the child initially has one undifferentiated language system (Valdes & Figueroa, 1994). As the child gets older (preK), the two languages begin to become recognizable. However, some mixing continues (i.e., mixing up word order). Finally, by school age, the two languages are mastered and are completely unique. The child is then able to speak both languages fluently.

CHAPTER II

REVIEW OF THE LITERATURE

In the context of chapter 1, chapter 2 expands on the key concepts of educational access and equity by tracing the roots of the fields of bilingual and special education. Key litigation and legislation guiding the development of both fields is reviewed. Next, the history and current status of intellectual testing and modern intelligence theories are summarized. At the end of the chapter, the epistemological framework and research questions are presented and their significance discussed.

Development of the Bilingual and Special Education Fields

Each year many professionals including doctors, teachers, administrators, lawyers, advocates and counselors are involved in the field of special education (DeValenzuela, Connery, & Musanti, 2000). One of the basic activities of these professionals is to identify children who have instructional needs that may warrant removal from the general education classroom for some percentage of the day. This process is designed to provide specific instruction to students to maximize their potential. However, sometimes educational institutions sort children into groups that include race, class, gender, and abilities (both high and low). These educational hierarchies can serve to limit educational opportunities for some if the focus is on

what students *cannot* do rather than what they *can* (Gartner & Lipsky, 1987). This occurs in general, special, and bilingual education.

In the field of bilingual education, myriad obstacles have been identified to explain why ELL students are not more successful in school. These factors include cultural deprivation (the view that children from diverse backgrounds lacked culture), limited English proficiency, and lack of prior knowledge, to name but a few (Crawford, 2004). An unfortunate parallel between special and bilingual education is that in finding reasons for lack of achievement in some groups of children, little attention has been given to finding solutions. One approach is for educators in both fields to have knowledge of the other field because of these parallels. Therefore, the following sections provide overview of the bilingual education and special education field including relevant legislation and litigation that have impacted their development.

Bilingualism in the United States

Defining bilingualism is not an easy task. Broadly, bilingualism refers to the ability to communicate in two languages without a distinction as to competence. To understand how varied bilingual individuals are Valdes and Figueroa (1994) described different types of bilingualism. First, *sequential* bilinguals were exposed to their native language (L1) since birth and later exposed to a second language (L2), usually when they entered public school. Variability is the hallmark of sequential bilingualism.

Second, *simultaneous* bilinguals were exposed to and learned both languages *at the same time*. The additional distinctions are based on an individual's choice in becoming bilingual. *Elective* bilinguals are those who choose to acquire a second language. *Circumstantial bilinguals* must learn their second language to survive in the environment. Their first language is generally not valued by society. Besides these distinctions, bilingual individuals vary in the proficiency levels of their languages. These factors add to the complexity of knowing how to best educate bilingual students.

Bilingual Education: Models and Their Impact on Student Achievement

The USDOE (2004) defined bilingual education as an:

...educational program for limited English proficient students that:
(a) makes instructional use of both English and a student's native language;
(b) enables limited English proficiency students to achieve English proficiency and academic mastery of subject matter content and higher order skills, including critical thinking, so as to meet age-appropriate grade-promotion and graduation standards in concert with the National Education Goals. (p. 1)

It is crucial to note that the goal of all bilingual program models is English proficiency and literacy. The major difference between the models is whether initial literacy is taught in the native language or in English.

Bilingual education was not a recent invention but a natural outgrowth of earlier waves of immigrants (Rethinking Schools, 2003). Ohio adopted the first bilingual education law in 1839 authorizing instruction in German. "By the end of the 19th century, about a dozen states had passed similar laws" (Rethinking

Schools, 2003, p. 1). By the turn of the 20th century, about 4% of American children were receiving bilingual instruction in languages such as Norwegian, Italian, Polish, Czech, and Cherokee (Rethinking Schools, 2003, p. 1). After World War I, however, bilingual education programs were banned until 1968 when the civil rights movement spawned the Bilingual Education Act of 1968. Since that time, Bilingual Education has remained controversial and instruction in a language other than English is currently outlawed in some states (Crawford, 2004). However, the decision to mandate English-only instruction so that ELL students can academically succeed in English is not supported by the research (Crawford, 2004; Cummins, 1982, 1984; Thomas & Collier, 1997, 2002; Tse, 2001).

Research over the last 20 years demonstrates that positive cognitive gains are associated with learning a second language in childhood (Baker, 2000; Bialystok, 1991; Krashen, 1999; Tse, 2001). Bilingualism has been shown to foster classification skills, concept formation, analogical reasoning, visual-spatial skills, creativity, and other cognitive gains. Between 1996 and 2001, Thomas and Collier (1997, 2002) conducted a longitudinal study that built on 14 years of related research to investigate the effectiveness of bilingual programs. They documented the academic achievement of ELLs over the long-term (4-12 years) and across content areas. Data were collected from five school districts throughout the U.S. A description of each model studied by Thomas and Collier and its impact on students' academic success is found in the following sections.

Two-Way Bilingual Immersion Programs

The goal of two-way bilingual immersion programs is to promote academic achievement, bilingualism, and biliteracy for ELLs and native English speakers. Such programs are usually provided for at least 5-6 years. ELLs and English speakers, in equal numbers, work together to learn both languages. Thomas and Collier (2002) further distinguished these programs as either 90/10 (i.e., students receive 90% of their instruction in a language other than English and 10% of instruction in English in the early years) programs or 50/50 (i.e., instructional time in English is equal to instructional time in the non-English language) programs throughout the program.

Late-Exit and Content ESL

Late-exit and content ESL programs are also known as maintenance bilingual programs and share the goals and duration of Two-Way Bilingual Immersion programs. Development of the students' first language is desirable. In this model, language minority students of one language background are grouped together in a class and receive instruction in their native language. ESL instruction is generally tied to the content instruction in the classroom. The aim is to promote high levels of academic achievement in all curricular areas and full academic language proficiency in the students' first and second languages.

Early-Exit Bilingual and Content ESL

Early-exit bilingual and content ESL programs offer classes presented in the ELLs' native language for 2 or 3 years after which time all-English instruction is provided. ESL is also provided and the instruction generally ties into classroom units. While early-exit programs are the most common forms of bilingual education models, they do not aim for full bilingualism. Most of these programs begin in kindergarten or grade one.

Early-Exit Bilingual and Traditional ESL

Early-exit and traditional ESL program models offer classes presented in the ELLs' native language for at least 2 or 3 years after which time ELLs receive all-English instruction. ESL is a separate curriculum and unrelated to classroom units or content.

English-as-a-Second Language Pull-Out

English-as-a-second language pull-out programs offer instruction in an environment outside of the general education classroom. Content is frequently unrelated to classroom content curriculum. The goal of this model is the rapid acquisition of English with no regard for the development of L1.

Major Findings of the Thomas and Collier Study

Three models (e.g., the 90/10 and 50/50 Two-Way Bilingual Immersion and One-Way Developmental Bilingual Education) were the only programs in which students reached the 50th percentile in both their native language and in English in

all subject areas. Plus, students were found to maintain the level of high achievement through the end of their schooling. These models showed the lowest dropout rates.

In some cases, ELL students attended English-only mainstream programs because their parents refused language support services. This ESL-only group showed large decreases in reading and math achievement by Grade 5 when compared to students who participated in language support programs. The largest number of dropouts came from this group.

When ELL students exited a language support program into the English mainstream, those schooled in all-English programs initially outperformed those schooled in the bilingual programs when tested in English. The students who received bilingual programs, however, reached the same levels of achievement as those schooled all in English by the middle school years. During the high school years, this group *outperformed* the students schooled in all English.

To summarize, the amount of formal schooling in L1 that a student received was the strongest predictor of student achievement in English. The greater the number of years in L1 during grade-level schooling, the higher the student's academic achievement in English. Table 1 summarizes the program models and their goals.

Thomas and Collier's (1997) research shows substantial benefits when children are allowed to build on their L1 while gaining fluency and ultimately

Table 1

Program Models for Educating ELL Students

Program	Language	Components	Duration	Goals
ESL Pullout	90-100% English; may include some L1 support	Mainstream instruction; ELL students are pulled out for 30-45 minutes of daily instruction and taught by ESL certified teachers.	2-4 years; ELLs exit on becoming English proficient	Assimilating ELL students; L2 acquisition
Early-exit Transitional Bilingual Education and Traditional ESL	10-50% L1; 50-90% English	L1 and sheltered subject-matter instruction; daily grammar-based ESL, initial literacy usually in L1, teachers certified in Bil. Ed., and proficient in L1 and L2.	2-4 years; ELLs exit on becoming English proficient	Assimilating ELL students; L2 acquisition
Program	Language	Components	Duration	Goals
Early-exit Transitional Bilingual Education and Content ESL	10-50% L1; 50-90% English	L1 sheltered subject-matter instruction; daily ESL with curriculum tied to general education classroom's unit, initial literacy usually in L1, teachers certified in Bil. Ed. and proficient in L1 and L2.	2-4 years; ELLs exit on becoming English prof.	Assim. ELL stu. L2 acquisition without falling behind aca.
Late-exit Developmental Bilingual Education with Content ESL	90% L1 in K-1; gradually decreased to 50% or less by Grade 4	L1 and sheltered subject-matter instruction; daily ESL, initially literacy in L1; teachers certified in Bil. Ed. And proficient in L1 and L2.	5-6 years; former ELLs continue some classes in L1	Bil. and biliteracy; long-term academic achievement in English.
Program	Language	Components	Duration	Goals
Two-Way Bilingual	90/10 -90% L1, 10% English in early grades; 50/50 model parity of both L1/L2.	Ells and native English speakers taught literacy and subjects in both languages; peer tutoring; teachers certified in Bil.Ed., proficient in L1 and L2.	5-6 years	Bilingualism and biliteracy, long-term academic achievement in English.

literacy in L2. Beyond the academic benefits, bilingual programs can also provide the catalyst for valuing the diverse cultures and languages represented in the school. Bilingualism also aides later employment opportunities, and perhaps most importantly, more ELL students remain in school and graduate from bilingual programs (Greene, 1998).

Criticisms of Thomas and Collier's Study

Thomas and Collier's work was criticized by some because they chose to avoid "laboratory-style research methods" (Crawford, 2004, p. 235; Rossell, 1999). They claimed random assignment of students to experimental and control groups were not only impractical but unethical when working with students who needed rigorous instruction in the best program models. Instead, they pursued a "noninterventionist philosophy" (Crawford, 2004, p. 235) and collaborated with district staff to evaluate the programs that were already in place and student performance in those programs. In 1997, when Thomas and Collier released their report, it did not include the usual statistical analysis, data tables, or clear descriptions of the research conditions. Their findings were thus criticized by many in the field and the results questioned (Rossell, 1999). In 2002, Thomas and Collier released a second report on the same data that included more of the usual research features. Regardless of the controversy, the results of their study are the most commonly cited data when comparing the various models of bilingual programs

and their achievement outcomes. Other research (August & Hakuta, 1997; Hakuta, 1990; Ramirez, 1992) supports the positive impact on student achievement of specialized educational programs using native language instruction. Given that bilingual education was present at the beginning of the 20th century and then virtually nonexistent until the late 1960s, it is helpful to review the key legal findings guiding the reoccurrence of the field.

Legislation and Litigation Guiding the Development of Bilingual Education

Legislation

The 1960s civil rights era moved the nation toward equality in the law as well as education. Seminal legislation for bilingual education includes the Civil Rights Act of 1964 prohibiting discrimination against students on the basis of their language minority status. The Bilingual Education Act of 1968 was the first law to provide federal funding to school districts to provide appropriate programs for students with limited English proficiency. This law, however, did not require L1 instruction. A third directive, known as the OCR May 25th Memorandum (Office of Civil Rights, 1970), outlined specific guidelines for districts to (a) take affirmative steps to rectify the language deficiencies of LEP students, (b) refrain from assigning these students to classes for the mentally handicapped on the basis of criteria which reflected their English language skills, (c) insure that any ability grouping designed to meet these students' language skills did not result in

permanent tracks, and (d) required that parents be notified of school activities in a language that they understood. In 1984, the BEA, reauthorized every 4 years, specifically addressed language minority students with disabilities by funding the Special Populations Program, the first time bilingual special education was funded at the federal level (Crawford, 2004). Throughout this period, the courts were also making impacting the field.

Litigation

There are several significant court cases addressing equality in education for language minority students. *Lau v. Nichols* (1974) ruled that the San Francisco Unified School District violated Title VI of the Civil Rights Act when it did not provide services to help Chinese-speaking students learn English claiming that basic English skill are at the core of what schools teach. In the same year, *Diana v. State Board of Education* (1970) addressed assessment issues of Mexican-American children. The ruling included that (a) all Mexican-American children who had been placed in special education on the basis of IQ tests administered in English be reassessed in their first language and in English with a nonverbal IQ test, (b) the mandate that IQ tests appropriate for Mexican-American students be developed, and (c) the requirement that school districts be monitored to identify racial and ethnic disparities in special education programs. *Castaneda v. Pickard* (1981) set a three-pronged standard for courts in examining programs for LEP

students. It directed districts to have (a) a pedagogically sound plan for LEP students, (b) sufficient qualified staff to implement the plan, and (c) a system established to evaluate the programs. The Court, however, fell short of requiring bilingual programs to meet these standards and required only that appropriate action to overcome language barriers be taken through well implemented programs (Crawford, 2004).

In summary, over the past several decades legislation and litigation have guided educational equity for students who are not native English speakers. The same occurred for the special education field. It is evident that many parallels between the fields exist including their roots and focus on equity for those deemed different in some way.

The Early Development of the Learning Disabilities Field

Hallahan and Mercer (2001) divided the history of learning disabilities into several periods. The first period, the European Foundation Period (1800-1920), focused on the exploration between brain injury and behaviors (Swanson, Harris, & Graham, 2002). Research in language disorders led to the discovery of “word blindness” that was later termed “dyslexia” with an etiology attributed to faulty memory for words and letters (Swanson et al., 2002).

The U.S. Foundation Period

During the second period, United States Foundation Period (1920-1960), compulsory education for children was established and there was a focus on intellectual testing. Thus, researchers began working with children in educational settings and focused much of their work on reading disabilities (Swanson et al., 2002). During this time, Fernald advocated for an integrated sensory approach for teaching reading and writing, with teaching strategies developing for struggling students. Because of the success of new teaching approaches, previously held beliefs about IQ were challenged. Orton (cited in Swanson et al., 2002) hypothesized that IQ was “not always reflective of true intellectual capacity, especially in students with reading deficits...” (p. 19) – a view that continues today (Stanovich, 1986). Monroe (1932) foreshadowed the modern discrepancy approach by beginning a practice of calculating a reading index, the discrepancy between actual and expected level of reading achievement for a student. Establishing the need in the field for assessments that guided instruction, Kirk (cited in Swanson et al., 2002) “found a need for assessments that could isolate and identify abilities and disabilities” (p. 20).

Work in the field of mental retardation saw a focus on improving learning environments as well as the use of standardized tests. As far back as 1937, however, Werner and Strauss (cited in Swanson et al., 2002) cautioned the field to

interpret standardized test scores with caution saying that to “understand normal child psychology, as well as mental deficiency, one must go beyond mere standardized achievement test scores” (p. 21). During this period, Cruickshank, Bentzen, Ratzeburg, and Tannhauser (1961) bridged the research on students with mental retardation with students who today would be considered learning disabled.

The Emergent Period

From 1960 to 1975, the Emergent Period saw the birth of the field of learning disabilities. The term “learning disabilities” was introduced to the educational field in the early 1960s by Kirk (cited in Hallahan & Mercer, 2002). Kirk used this term to describe a group of children who had disorders in the development of language, speech, reading, and associated communication skills needed for social interaction. Children with sensory (e.g., blindness or deafness) disabilities or those with generalized mental retardation were not included. In 1965, Batemen (one of Kirk’s students), offered the following definition:

Children who have learning disorders are those who manifest an educationally significant discrepancy between their estimated potential and actual level of performance related to basic disorders in the learning process, which may or may not be accompanied by demonstrable central nervous system dysfunction, and which are not secondary to generalized mental retardation, educational or cultural deprivation, severe emotional disturbance, or sensory loss. (p. 220)

Bateman’s definition is significant since it reintroduced an earlier notion of using a discrepancy between achievement and potential or ability as a way of identifying

students with a learning disability. The use of a “discrepancy formula” is discussed in more detail in a later section.

After this group of children with learning disorders was identified, the federal government began to include the topic on its agenda. From the 1960s to mid 1970s, many organizations for learning disabilities were founded. In addition, schools focused on the educational programming for students with learning disability. It should be noted that the original 1966 version of Education of the Handicapped Act did not include the category of learning disabilities (Hallahan & Mercer, 2001).

The Solidification Period

The Solidification Period (1975-1985) saw a focus on definition and federal regulations. Education for All Handicapped Children Act (P.L. 94-142), passed in 1975, became the first national law to include learning disabilities as a category eligible for funding direct services (Polloway, 2002). The final regulations, established in 1977, specified an ability-achievement discrepancy model for special education eligibility under the category of Specific Learning Disability (SLD). The regulations required a multidisciplinary team to determine that a student had a severe discrepancy between achievement and ability in at least one of seven areas: (a) listening, (b) thinking, (c) speaking, (d) reading, (e) writing, (f) spelling, or (g) doing mathematical calculations. With the new regulations, the field of special

education moved toward a consensus definition of learning disabilities and methods to identify students.

The Turbulent Period

The period between 1985 and 2000 is considered the Turbulent Period (Swanson et al., 2002). During this time, the number of students identified as SLD doubled, creating the need for a firm definition of SLD. In 1988, the definition was revised to essentially the current definition. The federal definition of SLD is “a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written...” (Hargrove, 1982, p. 366). While the regulations specified that a child exhibit a severe discrepancy between ability and achievement and a processing disorder, the operationalization of the regulations was left to the states. Some states left this to the individual local education agencies. Thus, systems interpreted these regulations in many different ways. While some systems required that a processing disorder be identified in addition to a discrepancy, others did not. Sometimes a single measure, the “discrepancy formula” was used.

Because of the historic reliance on the discrepancy formula in making eligibility decisions, it is defined here but discussed in more depth later in this chapter. The discrepancy formula compares a student’s performance on a cognitive or IQ test with performance on a standardized achievement test and looks for a

discrepancy of a pre-determined magnitude. There are different ways a discrepancy is quantified. Some districts used a straight point difference to operationalize a “significant discrepancy” while others used more sophisticated formulas (e.g., regression formula) (Kavale, Holdnack, & Mostert, 2005). A regression formula takes into consideration the tendency for extreme scores on one variable, such as an IQ score, to predict less extreme scores on the other variable (i.e., achievement scores) or, in other words, closer to the mean. Therefore, without correction for regression, discrepancies may be identified for some children when none actually exist. Currently, the discrepancy formula is under controversy.

The Period of Reconceptualization

The current period might be considered the period of Reconceptualization of Specific Learning Disability. By the late 1990s, some professionals were calling for a reconceptualization of how SLD should be assessed and identified (Kavale et al., 2005). The discrepancy formula, as discussed earlier, was deemed problematic for several reasons. The Matthew Effect (Stanovich, 1986) contended that better readers were likely to learn more about the world. Therefore, IQ tests probably underestimated the IQs of poor readers. The use of a discrepancy model also made it difficult to identify children in early elementary grades when scores may not yet be significantly discrepant. Thus, it was difficult to provide early intervention for reading difficulties. The next criticism stated that researchers had not been able to

discriminate between students with low reading achievement and *no* discrepancy from students *with* a discrepancy on measures related to reading achievement.

Given the long-held concerns by many groups about the discrepancy formula, the U.S. Department of Education Office of Special Education Programs committed to “carefully review research findings, expert opinions, and practical knowledge over several years to determine whether changes should be proposed to the procedures for evaluating children suspected of having a specific learning disability” [USDOE, 1999, p. 12541]. The outcome was a change in the eligibility procedures for SLD, discussed later, as well as a stronger focus on disproportionality in the 2004 reauthorization of Individuals with Disabilities Act. At both the state and national level, however, policy was not the only influence on the field. Legislation and litigation also propelled special education forward.

Litigation and Legislation Guiding the Development of Special Education

Prior to the 1970s, during the U.S. Foundation Period, equal access to a free and appropriate education for students with disabilities was seen as a privilege rather than a right with only 20% of students with disabilities served (Yell & Rogers, 1998). In the 1950s and 1960s, the civil rights movements focused the nation on racial equity and equal educational opportunities for people of color. The landmark case, *Brown v. Board of Education* (1954), set the stage by challenging the constitutionality of school racial segregation (Laosa, 2001; Losen & Orfield,

2002). The Supreme Court unanimously held that “in the field of public education, the doctrine of ‘separate but equal’ has no place” (Laosa, 2001, p. 36). Further, the Court found that the plaintiffs had been deprived of the equal protection of the laws guaranteed by the 14th Amendment of the U.S. Constitution (Laosa, 2001).

The 1960s, however, saw unequal implementation of this decision while, in general, public awareness of inequity and discrimination increased (Laosa, 2001). In 1964, The Civil Rights Act required a cessation of federal funds to school districts and other institutions that discriminated against *any* group. The Supreme Court continued to provide strong leadership on civil rights during 1963 and 1964, and 1969 to 1973 as did Presidents Kennedy and Johnson (Laosa, 2001). The civil rights movement, however, also broadened the scope of equity in education. It provided a “catalyst to parents and advocacy groups to begin using the courts in an attempt to force states to provide a public education that was appropriate for their children’s unique needs” (Katsiyannis, Yell, & Bradley, 2001, p. 325).

Litigation on Behalf of Students with Disabilities

Parents of children with disabilities saw changes in segregation laws as their opportunity to advocate for expanded rights for the disabled. They claimed that if segregation by race was a denial of equal educational opportunities, parent groups excluding students with disabilities from any aspect of education was also a denial of equal opportunity. Using *Brown v. Board of Education* (1954) where equal

protection to a “class” of people (i.e., racial minorities) was mandated, advocates for students with disabilities claimed that these students were entitled to the same rights as those without disabilities (Yell & Rogers, 1998). This line of logic led to two other landmark decisions, *Pennsylvania Association for Retarded Citizens (PARC) v. Commonwealth of Pennsylvania* (1972) and *Mills v. D.C. Board of Education* (1972).

In *PARC v. Commonwealth of Pennsylvania* (1972), the Court ordered that all children with mental retardation between the ages of 6 and 21 be provided a free and appropriate public education in programs like those provided for their peers without disabilities (Losen & Orfield, 2002; Yell & Rogers, 1998). The *Mills v. D.C. Board of Education* (1972) case, a class action suit, claimed that seven students with a variety of disabilities were excluded from public education (Yell & Rogers, 1998). The court held in favor of the plaintiffs and established due process safeguards that were to become the framework for the first federal special education laws.

These two landmark decisions created a litigious climate and during the next few years, lawsuits were filed in 28 states (Zettel & Ballard, 1982). Although the outcomes of these cases paralleled those of *PARC v. Commonwealth of Pennsylvania* (1972) and *Mills v. D.C. Board of Education* (1972), in practice, many students with disabilities continued to be denied appropriate educational

opportunities (Zettel & Ballard, 1982). The basis of this denial at the state level was the claim of insufficient funds. By the early 1970s, it became apparent that federal involvement was necessary (Yell & Rogers, 1998).

Federal Legislation and the Education of Students with Disabilities

The initial special education law was the brainchild of President Johnson. The Elementary and Secondary Education Act of 1965 (ESEA) provided federal funding for the education of students with disabilities. ESEA required that the federal government provide direct financial support to states to improve the education of students with disabilities in state schools for children who were blind, deaf, or had mental retardation. In 1966, this Act was amended and the Bureau of Education for the Handicapped in the Department of Health, Education, and Welfare, later to become the Office of Special Education Programs, was created.

In 1970, P.L. 91-230 Education of the Handicapped Act (EHA) amended the ESEA. This became the first law that exclusively addressed students with disabilities. At the same time, there were concerns regarding the generalizability of court cases across all states which propelled a number of organizations seeking federal intervention to form an alliance. The alliance included the Council for Exceptional Children, the American Federation of Teachers, The Association of Retarded Children, Council of Chief State School Officers, National Association of State Directors of Special Education, National Education Association, National

School Boards Association, and United Cerebral Palsy Association of America (Weintraub, 2005). For the first time such diverse organizations came together to support children with special needs.

In 1975, the EHA was amended and was undoubtedly the most significant law concerning students with disabilities. First, it called for a free appropriate education (FAPE) that would meet the unique needs of all children with a wide range of disabilities including physical handicaps, mental retardation, speech, vision and language problems, emotional and behavioral disorders and other learning disorders in kindergarten through grade 12. Public Law 94-142 or the Education for All Handicapped Children Act (EAHCA) was signed into law by President Gerald Ford on November 29, 1975. Second, it required that all states receiving federal funds submit state plans detailing how to educate children with special needs (Weintraub, 2005).

New Mexico was the only state to decline funds to avoid complying with the legal mandates of EAHCA. In response, the New Mexico Association for Retarded Citizens (NMARC) sued the state under Section 504 of the Rehabilitation Act of 1973 for failing to provide FAPE for students with disabilities (Yell & Rogers, 1998). The plaintiffs prevailed and the state was mandated to comply. After this decision, all 50 states, in essence, became partners with the federal government in educating students with disabilities (Heufner, 2000).

Individuals with Disabilities Education Act

The law required reauthorization of EAHCA (P.L. 94-142) every 4 years. Thus, the law was first amended in 1983 with the latest reauthorization occurring in 2004. Each reauthorization expanded the rights of students with disabilities (Katsiyannnis et al., 2001). The 1990 amendment (P.L. 101-476) changed the name of the law to the Individuals with Disabilities Education Act (IDEA). IDEA was amended in 1997 with important new safeguards. IDEA 1997 was the first to require the administration of tests in the student's native language, *if feasible*, increasing the focus on inappropriate identification of children who are racially, ethnically and linguistically diverse. Other mandates of this reauthorization included strengthening the role of parents and ensuring equal access to the general education curriculum. P.L. 101-476 was significant in two other ways as well. The categories of autism and traumatic brain injury were added and for the first time people first language (defined in the Glossary) was included (Hallahan & Mock, 2003).

Individuals with Disabilities Education Improvement Act

In December 2004, IDEA was reauthorized and officially titled Individuals with Disabilities Education Improvement Act (IDEIA). The IDEIA strengthened the language regarding equitable assessment and disproportionality. It states in part:

Procedures to ensure that testing and evaluation materials and procedures utilized for the purposes of evaluation and placement of children with disabilities for services under this title will be selected and administered so as not to be racially or culturally discriminatory. Such materials or procedures shall be provided and administered in the child's native language or mode of communication, unless it is clearly not feasible to do so, and no single procedure shall be the sole criterion for determining an appropriate educational program for a child. (20 U.S.C. §1412 (6)(B))

OVERIDENTIFICATION AND DISPROPORTIONALITY. – The State has in effect, consistent with the purposes of this title and with section 618(d), policies and procedures designed to prevent the inappropriate overidentification or disproportionate representation by race and ethnicity of children as children with disabilities, including children with disabilities with a particular impairment described in section 602. (20 U.S.C. §1412 (24))

While the above mandates are critical directives, there is little consensus as to how to operationalize nondiscriminatory assessments that can potentially rectify the problem of disproportionality. These changes in IDEA in eligibility procedures will force the field to revisit current processes and procedures. Many in the field (Lyon & Chhabra, 2002; Pasternack, 2002; Stanovich & Siegel, 1994) already felt IQ/cognitive assessment unnecessary. To examine this debate, the history of IQ testing in the SLD field is briefly examined and then IDEA's broadened eligibility procedures outlined to underscore their significant implications.

The Role of Intelligence Testing in the Diagnosis of SLD

Since the beginning of the 20th century, the construct of intelligence has been a subject of debate. At that time, Wechsler (cited in Wasserman & Tulskey,

2005), whose work in the field greatly influenced intelligence testing, noted that there were 14 different definitions of intelligence. Wechsler's definition, "Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment" (p. 15) is still widely accepted.

The g Factor

In 1904, Spearman (cited in Wasserman & Tulsky, 2005) introduced the concept of the *g* factor or global ability factor. He found that children's grades across unrelated school subjects were positively correlated, and discovered that these correlations reflected the influence of a dominant factor. He termed this dominant factor *g* for "general" intelligence and claimed that this common factor is manifested in individual differences on all mental tests. Further, all branches of intellectual activity have one fundamental function (or group of functions) in common which he described as the amount of general mental energy or *g* (Wasserman & Tulsky, 2005). "The *g* factor is a mathematically derived general factor, stemming from the shared variance that saturates batteries of cognitive/intelligence tests" (p. 16). Further discoveries lead Spearman to the development of a two-factor theory of intelligence where performance of any intellectual act required some combination of *g* and specific factors or *s* which vary in strength from one act to another. Spearman remained convinced that the most

important information to have about a person's intellectual ability was an estimate of *g* or general ability.

During much of the 20th century, researchers in intelligence were unable to dismiss the concept of *g* or global capacity. Even today, "most contemporary modelers of intelligence retain a 'g' factor" (Wasserman & Tulsky, 2005, p. 17). Eventually, variance in intelligence/cognitive test performance began to be explained by a variety of theories and the debate over the existence of *g* continued.

Beyond g

In the 1930s, Thurstone (cited in Wasserman & Tulsky, 2005) posited that intelligence was made up of several primary mental abilities rather than a general (*g*) ability and several specific factors. Thurstone showed multiple ways in which someone could be intelligent. He identified several primary mental abilities: (a) verbal comprehension, (b) word fluency, (c) number facility, (d) spatial visualization, (e) associative memory, (f) perceptual speed, and (g) reasoning. He recommended describing individuals by a profile of mental abilities rather than by a single index of intelligence (i.e., IQ).

Two General Factors

During the next decade, Cattell (cited in Wasserman & Tulsky, 2005) presented a new theory that described two separate general factors, (a) fluid ability/intelligence and (b) crystallized ability/intelligence. These two general factors gave rise to the modern-day constructs of fluid ability/intelligence (*Gf*) and

crystallized ability/intelligence (G_c). Later, Horn and Cattell (1966) increased the number of ability factors (i.e., visualization, retrieval capacity, cognitive speed) adding them to the original two factors of fluid and crystallized abilities. In the 1990s, Horn revised the theory to include nine ability factors (Horn & Noll, 1997).

A Modern Intelligence Theory: CHC

In the late 1980s and early 1990s a new empirically based psychometric taxonomy of human cognitive abilities, the Cattell-Horn-Carroll (CHC) theory of cognitive abilities, appeared (McGrew, 2005). McGrew (2005) described this as a “consensus model” between the Cattell-Horn theory and one developed by Carroll (p. 149). “CHC theory of intelligence is the tent that houses the two most prominent psychometric theoretical models of human cognitive abilities” (McGrew, 2005, p. 137). The model consists of three strata: (1) the concept of g or general ability in stratum I, (2) nine broad abilities at stratum II, and (3) approximately 70 narrow abilities in stratum III. An overview of the broad abilities as well as narrow abilities is presented in Figure 1. This model serves as the theoretical foundation for some of the latest and most sophisticated cognitive assessments and is gaining acceptance by special education professionals (Fiorello & Primerano, 2005). As with any theory, research must validate its premise. Current research on the use of the CHC model in special education eligibility assessments appears to be validating the framework as useful in guiding test selection and making differential diagnoses (Fiorello & Primerano, 2005).

Figure 1. Cattell-Horn-Carroll Integrated Theory of Intelligence (CHC Theory).

The Cattell-Horn-Carroll (CHC) theory of cognitive abilities is supported by a large network of validity evidence, which includes more than half a century of factor analytic, developmental, heritability external outcome validity, and neurocognitive research evidence. (Floyd, Evans, & McGrew, 2003, p. 156)

CHC theory is the theoretical framework underpinning modern cognitive tests including the Woodcock-Johnson Tests, the first battery to do so.

The Woodcock Johnson Psycho-Educational Battery-Revised (Woodcock & Johnson, 1989; referred to throughout document as WJ-R) and its Spanish counterpart, the *Batería* (Woodcock, 1982), were the first to use the predecessor to CHC theory, Gf-Gc theory, as its framework. The concept of combining theories by Cattell, Horn, and Carroll was first described in the Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2001, referred to throughout document as WJ-III) Technical Manual in 2001. By then the model had evolved to a model of eight broad abilities (McGrew, 2005). The WJ-III assesses the broad CHC abilities; and thus, provides a profile of an individual's specific cognitive strengths and weaknesses.

From this profile, targeted, research-based interventions can be provided that are individualized specifically to that student's strengths and weaknesses, thus narrowing the gap between theory and practice. Thus, the only way to measure cognitive strengths and weaknesses is through a standardized cognitive/intelligence test. A significant body of research related to this is beginning to emerge and is presented in the next section.

Research on Processing Abilities and their Link to SLD

Emerging research in the school psychology field supports the importance of establishing a picture of an individual's processing strengths and weaknesses to help guide intervention.

With modern theories about the importance of processing skills replacing the outdated psychological processing views (e.g., perceptual-motor deficits) that were associated with the SLD concept when first proposed, it becomes critical to reemphasize process deficits in an operational definition of SLD. (Kavale et al., 2005, p. 4)

Thus, the IQ score, or Global Intellectual Ability (GIA) score on the WJ-III, is becoming less important than the scores for the broad processing abilities in which strengths and weaknesses can be identified. Table 2 shows the broad processing abilities, or CHC factors, their definitions and link to achievement.

CHC Factors as Abilities – Not Learning Styles

There has been a movement to deemphasize general ability scores and focus more on specific intellectual constructs “because of the *belief* that subtest scores yield useful diagnostic and treatment validity” (Fiorello & Primerano, 2005, p. 525). For example, the recommendations by the National Research Council's Committee on the Prevention of Reading Difficulties (Snow, Burns, & Griffin, 1998) identified several areas that are necessary to acquire literacy. They are: (a) linguistic proficiency, (b) verbal memory, (c) lexical and syntactic skills, (d) general language abilities, and (e) phonological awareness. These abilities can be categorized within several CHC factors – Gc, Gsm, Glr and Ga and are the

Table 2

CHC Factors and Achievement Areas

CHC Factor	Definition	Achievement Area
Comprehension-Knowledge (<i>Gc</i>)	The breadth and depth of knowledge including verbal communication and information. Reasoning, when using previously learned procedures, is also included.	The strongest predictor of Basic Reading Skills and Reading Comprehension. Impacts math reasoning.
Long-Term Retrieval (<i>Glr</i>)	The ability to store information efficiently and retrieve it later through association.	Moderate relation with Basic Reading Skills and Comprehension during early school years.
Visual Processing (<i>Gv</i>)	Spatial orientation, the ability to analyze and synthesize visual stimuli, and the ability to hold and manipulate mental images.	Weak relationship to reading between ages 6 – 19. May impact math reasoning.
Auditory Processing (<i>Ga</i>)	The ability to discriminate, analyze, and synthesize auditory stimuli. Also related to phonological awareness.	Relationship to early reading skills (basic reading and comprehension) only through formative years.
CHC Factor	Definition	Achievement Area
Fluid Reasoning (<i>Gf</i>)	The ability to reason and solve problems that often involve unfamiliar information or procedures. Manifested in the reorganization, transformation, and extrapolation of information.	Does not appear to be a significant predictor of reading achievement. Impacts math reasoning.
Short-Term Memory (<i>Gsm</i>)	The ability to hold information in immediate awareness and then use it within a few seconds, also related to working memory.	Moderate relation to Basic Reading Skills and Reading Comprehension during ages 6 through adolescence. As reading process becomes more automatized, less use of working memory.
Processing Speed (<i>Gs</i>)	Speed and efficiency in performing automatic or very simple cognitive tasks.	Moderate relation to Basic Reading Skills and Comprehension between ages 6- 10. Impacts math reasoning.

Sources: Evans, Floyd, McGrew, and Leforgee (2001); Proctor, Floyd, and Shaver (2005).

underlying constructs of early literacy development (Fiorello & Primerano, 2005). However, current research supports the predictive validity of cognitive processes to learning (Evans, Floyd, McGrew, & Leforgee, 2001; Fiorello, Hale, & Snyder, 2006; Floyd et al., 2003; Floyd, Keith, & Taub, 2007; Floyd, McGrew, & Evans, 2008; Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001). Processing abilities, then, can be crucial factors in eligibility decisions as well as instructional planning. Current changes in eligibility procedures (IDEIA, 2004) are leading practitioners away from considering students' processing strengths and weaknesses,

New Approaches to Eligibility Decision-Making

The IDEIA (2004) regulations do not prohibit the use of a discrepancy formula but provide options for identification. The options are (a) states may prohibit the use of a severe discrepancy between ability and achievement in making a determination of eligibility under the SLD category [§300.37(a)(1)], (b) local educational agencies may not be required to take into consideration a severe discrepancy between achievement and intellectual ability [§614(b)(6)(A)], and (c) districts may use response to scientific, research-based interventions as part of the determination process [§300.309(a)(2)(i)]. From these federal guidelines, states make their own regulations. Given the above options, some states may develop guidelines which in effect nullify the construct of SLD, which continues to include a disorder in one or more of the psychological processes. Thus, states that choose to focus on how students respond to interventions without requiring measures of

ability or cognitive processes as part of a comprehensive evaluation have no process in which to identify strengths and weaknesses in psychological processes.

Locally, the Oregon Administrative Rules (OAR) describe two approaches to eligibility decision making: (a) Response to Intervention (RTI) and (b) Pattern of Strengths and Weaknesses (PSW) evidenced in classroom performance and/or academic achievement relative to age, Oregon grade-level standards, or intellectual development. These approaches shift the focus from the discrepancy model, in which a child's academic difficulties are viewed as a "within child" issue, to examining the educational environment and then documenting how students' respond to research-based instruction and interventions. These two options, however, have the potential for convergence.

Approach One: Response to Intervention

The process based on the child's response to scientific, research-based intervention outlined in IDEIA (2004) is commonly referred to as Response to Intervention (RTI). This is the practice of screening all children early in their education to identify those who are not responding to instruction in their classroom, providing support using research-based curriculum and interventions at increasing levels (tiers) of intensity and monitoring progress frequently (Batsche et al., 2005). Instructional and educational decisions are based on continuing analysis of the data collected. RTI models typically have three or four tiers. While this approach is excellent at providing support for students early in their educational program, the

controversy lies in the final tier. In some models, students may be placed into special education programs at the final tier under the assumption that since the student has not made adequate progress with the increasingly intense interventions in the lower tiers, this data should be sufficient to qualify a student for special education. The problem with this practice is that it conflicts with the central concept and federal definition of SLD because a basic disorder in one or more psychological processes is never identified nor considered.

Approach Two: Identifying Patterns of Strengths and Weaknesses

The second approach is identifying a pattern of strengths and weaknesses (PSW) between classroom performance/academic achievement and cognitive development. Three theories can be found in the literature and each includes the concept of consistency between weaknesses in specific cognitive processes related to specific academic areas based on norm-referenced tests in an otherwise normal ability profile (Fiorello et al., 2006; Flanagan et al., 2007; Naglieri, 1999). As discussed, cognitive processing strengths and weaknesses can only be identified through standardized testing which remains controversial, particularly when the examinee is an ELL student. Identifying strengths and weaknesses could be conducted at the final tiers of RTI after a student has progressed through the earlier tiers. Thus, standardized assessment, coupled with the multiple data sources gathered through the RTI process, is a model in concordance with the SLD federal definition.

Reconciling the Approaches

In the school psychology literature, many are writing about reconciling the two camps – RTI-only and traditional psychoeducational evaluation approaches to SLD eligibility. As Hale, Kaufman, Naglieri, and Kavale (2006) pointed out

Seemingly opposing factions have called for either a response-to-intervention (RTI) approach or one that includes comprehensive evaluation of basic psychological processes prior to classification of children with specific learning disabilities (SLD). These apparently disparate approaches should not necessarily lead to a politicized professional schism. Instead, both positions should be scrutinized for their individual merits and limitations, with the result being a model that incorporates the best tenets of both perspectives in a balanced practice model that maximizes SLD diagnostic accuracy and optimizes educational outcomes for this heterogeneous and enigmatic population. (p. 753)

It makes sense, then, that the field can benefit from the strengths of both models. RTI includes a review of students' educational history, provides early intervening services, and documents interventions and progress. RTI cannot, however, provide insights into why some students do not respond. In this case, cognitive assessment may help solve the puzzle as well as provide guidance to appropriate educational interventions based on assessment results. As Willis and Dumont (2006) wrote "It seems odd that anyone would define the two approaches as mutually exclusive since the two are, in fact, complements to each other" (p. 902). Thus, even within this time of change, the practice of and need for administering standardized cognitive assessments cannot be minimized.

Standardized Tests Performance and Diverse Populations

Throughout the last century, research has shown performance differences on tests of intelligence between racial or ethnic groups and mainstream groups (Figueroa, 1990; Samuda, Kong, Cummins, Pascual-Leone, & Lewis, 1991; Valdes & Figueroa, 1994). "Test-score differences are usually in the range of three-quarters to one standard deviation" (Brooks-Gunn, Klebanov, & Duncan, 1996, p. 396). Differences in IQ scores between Black and White populations were seen as far back as the first Stanford-Binet IQ test in 1932 (Williams & Ceci, 1997) and then later on tests administered to recruits during World War I.

Throughout this century, Whites have outscored Blacks and Hispanics on IQ tests as well as standardized achievement tests. The gap most commonly reported is approximately one standard deviation. On the most widely used individual IQ test, the Wechsler series, one standard deviation translates into a 15-point gap between Blacks and Whites, with Hispanics falling midway between these two groups. (Williams & Ceci, 1997, p. 1228)

Because of these differences, assumptions have been made that diverse groups are simply not as smart as the middle-class, mainstream population (Gordon, 1980; Jensen, 1974). However, researchers failed to understand that many of the abilities measured by tests develop in *culturally specific ways* and, as such, are *manifested in culturally specific ways*. Intelligence tests, unfortunately, do not allow opportunities for demonstrating abilities in culturally responsive ways (Green, McIntosh, Cook-Morales, & Robinson-Zanartu, 2005).

The Impact of Cultural Experiences

Much of the research on IQ test bias reports that IQ tests are not biased (Cleary, Humphreys, Kendrick, & Wesman, 1975; Jensen, 1982; Reynolds, 1982) because they predict who will be successful in an American classroom “as well, if not better, for ethnic students as for Anglo students” (Valdes & Figueroa, 1994, p. 153). Valdes and Figueroa (1994), however, asserted that bias has not been found in IQ tests because of inappropriate research procedures which examined the predictive power of a test across racial/ethnic and socioeconomic groups. In these cases, correlation and regression analyses were generally used and procedures using a large sample and averaging a wide range of ability and achievement resulted in the loss of precision in the data.

To date, most studies on test bias using these procedures have failed to find racial/ethnic bias in psychometric tests. The common assertion among psychometricians is that well-normed and standardized tests, such as the WISC-R and the Binet, predict as well, if not better, for ethnic students as for Anglo students. (p. 153)

“Both with respect to culture and language proficiency (in the ethnic and societal languages), the psychometric literature has failed to adequately control for these independent, robust variables” (Valdes & Figueroa, 1994, p. 95).

To understand performance differences between mainstream and diverse groups takes deep probing into the causal factors. Tests of intelligence and cognitive ability reflect the culture from which they are designed and are based on the cultural and class values of test authors (Valdes & Figueroa, 1994). Thus, these tests measure the degree to which the examinee has acquired and can access the

culturally specific information reflected in test structures. In other words, an individual's exposure to the test's underlying culture affects test performance because standardized tests assume that test-takers have equivalent levels of *acculturation* across the variables of age or grade (Cummins, 1984; Figueroa, 1990; Matsumoto, 1994; Valdes & Figueroa, 1994). Knowledge about test-takers acculturation levels is a crucial factor in assessing how well a standardized instrument may measure an individual's abilities.

Acculturation

Acculturation is the process of change individuals undergo when they encounter and begin to incorporate a new culture into their own heritage culture. This process is usually measured by the level in which an individual incorporates the new culture's values, mores, language and traditions into their existing cultural base (Casas & Pytluk, 1995). Until recently, it was described as a developmental, gradual, one-way process correlated with the length of time an individual was exposed to a new culture (Monzó & Rueda, 2006). Newer models of acculturation, however, present a multidirectional process in which individuals carry knowledge of their heritage culture while at the same time accessing new and diverse cultural patterns of the dominant society (Monzó & Rueda, 2006; Padilla & Perez, 2003). Thus, as an individual's acculturation increases, performance on standardized tests should also improve as mainstream cultural capital and knowledge increases.

Standardized tests, whether normed on a monolingual English or Spanish population, cannot control for the variation of acculturation levels in bilingual Spanish-speaking students living in the United States. Therefore, this is problematic because living within a monocultural or multicultural environment may affect assessment outcomes (Acevedo-Polakovich et al., 2007). Jensen (1974) described an interesting way to conceptualize the impact of culture on test performance related to acculturation. Rather than claiming bias, he says that standardized tests are not culturally biased but they are culture *loaded*.

Tests are Culture Loaded

Culture loaded does not mean the same as *culture biased*. Tests and test items can be ordered along a continuum of culture loading, which is the specificity or generality of the informational content of the test items. The narrower or less general the culture in which the test's information content could be acquired, the more culture loaded it is. A test may contain information that could only be acquired within a particular culture. This can usually be determined simply by examination of the test items. The specificity or generality of the content corresponds to its cultural loading. A test requiring the respondent to name three monuments in the United States, is, in this sense, more culture loaded than the question, "How many hours are there in a day?"

In other words, tests may "measure a lower range of a bilingual's ability because the test may not have sampled the cultural content that is part of the

cognitive repertoire and processes available to the bicultural individual” (Valdes & Figueroa, 1994, p. 99). “Culture also may influence values and beliefs in ways that might influence test performance” (Brooks-Gunn et al., 1996, p. 404). For example, some cultures emphasize group cooperation rather than individual achievement. Thus, some children may underperform so as not to appear as boastful. The second critical factor is to consider students’ language differences.

The Impact of Language Experiences

The second major concern for diverse individuals on standardized tests relates to the language demands of the tests. Language development, whether it is the first, second, and so forth, is also experientially based and follows a sequential developmental course (Cummins, 1984). Cummins found that test items that are primarily language-based do not measure incidental learning equally well compared to tasks that are more visual or perceptual in nature. Tasks on standardized tests, therefore, vary according to their language demand. For ELL students, their language backgrounds, and not just their language proficiency in their native language and English, must be taken into account “in every facet of assessment such as test development, selection, administration, and interpretation” (Figueroa, 1990, p. 94). Because of linguistic differences, ELL students often score lower on a wide variety of tests, not because they have less innate ability, but because their linguistic differences may impede comprehension and communication (Flanagan & Ortiz, 2001). For instance, in many American middle-class families,

mothers and caregivers typically carry on running dialogues with children from early infancy to socialize them into language and literacy (Espinosa, 2005; Rogoff, 1990). In contrast, other cultures in which young children are cared for and learn through observation of adult's actions (Heath, 1983). Further, it is "possible for people to speak another language with fluency without understanding the nuances of culturally based nonverbal behavior" (Acevedo-Polakovich et al., 2007, p. 378).

Native Language Standardized Tests

The issue of linguistic and cultural differences of ELL students and standardized assessments are not solved when native-language tests are used because ELL students living in the United States are not appropriately reflected in the norming samples. Because of the great variation in acculturation levels and language proficiencies in English and native-language, it is highly unlikely that ELL students can ever be appropriately reflected in any norming sample even when native-language instruments are developed in the U.S. (Flanagan & Ortiz, 2001). The assumptions underlying standardized tests are that all test-takers have similar experiences, exposure to curriculum, general motivation toward test-taking, and language learning opportunities (Mercer, 1979). Therefore, these tests, while having acceptable psychometric properties, are not normed on individuals like U.S. ELL students who live within two language systems (Valdes & Figueroa, 1994) and usually within two underdeveloped language systems. Rich linguistic opportunities are the contexts which are taken for granted in a middle class

discourse but usually not present in working class and linguistic minority communities (Macedo, Dendrinos, & Gounari, 2003). Thus, standardized testing in L1 or L2, in the end, may not capture the students' full cognitive potential just as an English standardized test also will not adequately assess potential. The measurement of innate potential of ELL students is governed by environmental determinants that differ to a large extent from middle class ways of making meaning in any language. Therefore, one is left to wonder what to do when the focus of concern is an ELL student.

Current Practices in Testing ELL and Diverse Learners

There is a shortage of bilingual professionals in fields related to special education. Thus, a common approach is to use interpreters during the assessment process. Unfortunately, interpreters are rarely provided with adequate training. Instructional assistants are often asked to interpret during a standardized test administration without adequate preparation as to the content or the procedures for administering standardized tests (S. Haghighi, personal communication, November 8, 2005).

A second common practice is to administer non-verbal standardized assessments. However, nonverbal tests, while oral or expressive language may not be required, a high level of nonverbal receptive communication is needed to comprehend the assessor. (Flanagan & Ortiz, 2001). Also, the type of nonverbal communication that may be required for administration of non-verbal tests often

carries more culturally- and class-based implications than does verbal communication (Ehrman, 1996). The other problem of using non-verbal tests for measuring an individual's cognitive ability is that these tests do not measure the full range of abilities as identified in modern theories of intelligence such as CHC (Flanagan & Ortiz, 2001). Administering non-verbal tests only is potentially problematic in that a student's inherent strengths or weaknesses may not be measured nor identified. As Gunderson and Siegel (2001) stated regarding nonverbal tests, "An IQ test is not culture free, because background is important, nor is it language free, because it requires knowledge of English" (p. 49). In sum, neither the use of interpreters to translate English tests nor administering non-verbal tests result in nondiscriminatory assessment.

Bainter and Tollefson (2003) surveyed 500 school psychologists in eight states to examine the methods used to assess the cognitive ability of language minority students. They found that 85% of the respondents indicated that a bilingual school psychologist administering tests in both English and the child's native language is the best practice. However, 87% rated as acceptable the practice of administering traditional intelligence tests in English when a student is dominant in, or prefers using, English. Overall, the study found that there was no complete agreement as to what "best practice" is. Given the lack of understanding of what constitutes best practice in the intellectual/cognitive assessment of minority

students, there is a critical need for guidance on what constitutes nondiscriminatory assessment.

Nondiscriminatory Assessment

Although assessors may be mandated by IDEA to conduct nondiscriminatory assessments, Helms (1997) contended that psychometric models are not sophisticated enough to factor out the roles that race, culture, and social class play on students' responses to test stimuli. Given these issues, a framework for selecting, administering and interpreting standardized cognitive assessment data in a systematic manner based on research on how culture and language impact test performance for ELL students, both in English and in native language, is critically needed. Flanagan and Ortiz (2001) developed a framework that holds promise for nondiscriminatory assessment and interpretation when ELL students are assessed in English. This framework, however, needs to be studied when students are assessed on a Spanish-language cognitive test to investigate its appropriateness for use in helping to interpret scores in L2 testing.

Theoretical Framework for this Study

The Culture-Language Test Classifications and the Culture-Language Interpretive Matrix

Flanagan and Ortiz (2001; Ortiz & Flanagan, 2002) organized tests of cognitive ability according to three test characteristics: (a) the broad and narrow abilities they measure (according to Carroll-Horn-Cattell or CHC abilities), (b) the degree of cultural loading, and (c) the degree of linguistic demand. The

classifications are based on an evolutionary process, limited research on the use of intelligence tests in diverse populations (Cummins, 1982; Jensen, 1974; Sanchez, 1934; Vukovich & Figueroa, 1982; Yerkes, 1921) *and* expert consensus. For example, Mercer (1979) studied WISC-R scores of 700 “English proficient” Hispanic students. She found depressed scores in their verbal-factual subtest scores (i.e., Information, Similarities, Vocabulary, and Comprehension). In a review, Valdes and Figueroa (1994) found that the majority of language minority groups showed low verbal IQ and high nonverbal IQ profiles. Sandoval (1979) found performance differences in ethnic groups.

The pattern of children’s responses to the test is very similar regardless of the children’s cultural background. Despite the fact that the black and Mexican American children are different from each other, as well as from the general Anglo-American culture, the same items (59 or 34% of WISC-R items) tended to be difficult for the black and Mexican American groups. (p. 925)

This finding is important because Sandoval demonstrated that performance differences observed in minority groups related to their lack of mainstream cultural capital and that scores were similar across minority groups. In other words, scores were consistently depressed across minority groups. Sandoval, Zimmerman, and Woo (1980) examined the issue again with similar results to those reported above. After carefully reviewing all existing data, Flanagan and Ortiz (2001) proposed their test classifications, the Culture-Language Test Classifications (C-LTC) to offer practitioners a method of interpreting subtest scores from intelligence/ cognitive batteries that accounts for the biasing factors of culture and language.

Within the C-LTC and Culture-Language Interpretive Matrix (C-LIM) framework, however, full scale or other broad intelligence/cognitive scores are not interpreted.

The original intent of the classifications, the C-LTC was to allow practitioners to select a group of tests that would be as nondiscriminatory as possible, while allowing for the measurement of a broader range of abilities than that typically found in a “nonverbal” or comprehensive cognitive battery. By selecting a battery of tests and subtests appropriate for a specific student based on their level of acculturation and linguistic proficiency in English (and their native language when appropriate), practitioners would comply with the mandate of the IDEA 2004 which states: “Tests and other evaluation materials used to assess a child under Part B of the Act are selected and administered so as not to be discriminatory on a racial or cultural basis” (IDEA, 1997; IDEIA, 2004).

In addition to assisting in non-biased test selection, Flanagan and Ortiz (2001) concurrently developed the C-LIM, a framework for evaluating the relative influence of cultural and linguistic factors on test performance. The C-LIM was designed to address the fundamental question (is the measured performance a reflection primarily of actual ability or simply one of cultural or linguistic difference?) in the evaluation of diverse learners.

The C-LIM provides a graphical representation of the patterns of expected performance for diverse individuals from the perspective of the test’s cultural and linguistic characteristics rather than from the perspective of the constructs measured. Williams (1970, 1971) found that because minority children’s

experiences with concepts and vocabulary are different from those of children from the mainstream culture, it follows that the pattern of responses for minority children is different than the pattern for majority children. Thus, this pattern was based on existing research on IQ tests and diverse groups (Cummins, 1982, 1984; Flanagan & Ortiz, 2001; Mercer, 1979; Valdez & Figueroa, 1994).

Figures 2 and 3 show a C-LIM for two intelligence/cognitive tests, the WISC-IV and the WJ-III. The subtests are arranged according to their degree of linguistic demand and cultural loading. "The three categories, (low, moderate, and high) for degree of linguistic demand span across the matrix from left to right and the similar categories for degree of cultural loading run down the matrix from top to bottom" (Flanagan & Ortiz, 2001, p. 247).

Figure 4 (Flanagan & Ortiz, 2001) shows how the expected pattern of performance is influenced by the increasing effect that language and cultural differences are likely to have on test performance. That is, as the linguistic demands of tests increase, it is anticipated that an ELL students' scores will systematically decrease. Figure 5 (Flanagan & Ortiz, 2002) shows the ranges of score that ELL students are likely to obtain on the subtests contained in each cell. Three different score ranges are provided. One is for students who are slightly different from the mainstream in acculturation and language proficiency, the second for students who are moderately different, and the third for those markedly different on those two characteristics.

CULTURE-LANGUAGE INTERPRETIVE MATRIX (C-LIM) WORKSHEET - WISC IV SUBTESTS

Name of Examinee: _____ Age: _____ Grade: _____ Date: _____

DEGREE OF LINGUISTIC DEMAND

	LOW	MODERATE	HIGH
LOW	Test Name: _____ Score: () MATRIX REASONING (Gf) () Cancellation (Gs) () _____ () _____ () _____ ()	Test Name: _____ Score: () BLOCK DESIGN (Gv) () SYMBOL SEARCH (Gs) () DIGIT SPAN (Gsm) () CODING (Gs) () _____ ()	Test Name: _____ Score: () LETTER-NUMBER SEQ (Gsm) () _____ () _____ () _____ () _____ ()
	Cell Average = _____	Cell Average = _____	Cell Average = _____
	Test Name: _____ Score: ()	Test Name: _____ Score: ()	Test Name: _____ Score: ()
	_____ ()	ARITHMETIC (Gq) ()	_____ ()
	_____ ()	Picture Concepts (Gc) ()	_____ ()
MODERATE	_____ ()	_____ ()	_____ ()
	_____ ()	_____ ()	_____ ()
	_____ ()	_____ ()	_____ ()
	_____ ()	_____ ()	_____ ()
	_____ ()	_____ ()	_____ ()
HIGH	Cell Average = _____	Cell Average = _____	Cell Average = _____
	Test Name: _____ Score: ()	Test Name: _____ Score: ()	Test Name: _____ Score: ()
	Picture Completion (Gc) ()	_____ ()	INFORMATION (Gc) ()
	_____ ()	_____ ()	SIMILARITIES (Gc) ()
	_____ ()	_____ ()	VOCABULARY (Gc) ()
	_____ ()	_____ ()	COMPREHENSION (Gc) ()
	_____ ()	_____ ()	Word Reasoning (Gc) ()
	_____ ()	_____ ()	_____ ()
	_____ ()	_____ ()	_____ ()
	Cell Average = _____	Cell Average = _____	Cell Average = _____

DEGREE OF CULTURAL LOADING

Figure 2. Culture-language interpretive matrix (C-LIM) for the WISC-IV.

CULTURE-LANGUAGE INTERPRETIVE MATRIX (C-LIM) WORKSHEET - WJIII

Name of Examinee: _____ Age: _____ Grade: _____ Date: _____

DEGREE OF LINGUISTIC DEMAND

LOW		MODERATE		HIGH	
Test Name:	Score:	Test Name:	Score:	Test Name:	Score:
SPATIAL RELATIONS (Gr)	()	VISUAL MATCHING (Gs)	()	CONCEPT FORMATION (Gf)	()
	()	NUMBERS REVERSED (Gsm)	()	ANALYSIS SYNTHESIS (Gf)	()
	()		()	AUD WORKING MEM (Gsm)	()
	()		()		()
	()		()		()
Cell Average =		Cell Average =		Cell Average =	
Test Name:	Score:	Test Name:	Score:	Test Name:	Score:
Picture Recognition (Gr)	()	VIS-AUD LEARNING (Ghr)	()	MEMORY FOR WORDS (Gsm)	()
PLANNING (Gr)	()	Delay Rel-Vis And Learning (Ghr)	()	INCOMPLETE WORDS (Ga)	()
PAIR CANCELLATION (Gs)	()	RETRIEVAL FLUENCY (Ghr)	()	SOUND BLENDING (Ga)	()
	()	RAPID PICT NAMING (Ghr)	()	AUDITORY ATTENTION (Ga)	()
	()		()	DECISION SPEED (Gs)	()
Cell Average =		Cell Average =		Cell Average =	
Test Name:	Score:	Test Name:	Score:	Test Name:	Score:
	()		()	VERBAL COMPREHNSN (Gc)	()
	()		()	GENERAL KNOWLEDGE (Gc)	()
	()		()		()
	()		()		()
Cell Average =		Cell Average =		Cell Average =	
LOW		MODERATE		HIGH	
DEGREE OF CULTURAL LOADING		MODERATE		HIGH	

Figure 3. Culture-language interpretive matrix (C-LIM) for the WJ-III.

Figure 4. Pattern of expected test performance for Hispanic children.

General Guidelines for Expected Patterns of Test Performance for Diverse Individuals

DEGREE OF LINGUISTIC DEMAND

	Low	Moderate	High
L O W	Slightly Different: 3-5 points Moderately Different: 5-7 points	Slightly Different: 5-7 points Moderately Different: 7-10 points	Slightly Different: 7-10 points Moderately Different: 10-15 points
M O D	Slightly Different: 5-7 points Moderately Different: 7-10 points	Slightly Different: 7-10 points Moderately Different: 10-15 points	Slightly Different: 10-15 points Moderately Different: 15-20 points
H I G H	Slightly Different: 7-10 points Moderately Different: 15-20 points	Slightly Different: 10-15 points Moderately Different: 15-20 points	Slightly Different: 15-20 points Moderately Different: 20-30 points

DEGREE OF CULTURAL LOADING

Slightly Different: Includes individuals with high levels of English language proficiency (e.g., advanced BICS/emerging CALP) and high acculturation, but still not entirely comparable to mainstream U.S. English speakers. Examples include individuals who have resided in the U.S. for more than 7 years or who have parents with at least a high school education, and who demonstrate native-like proficiency in English language conversation and solid literacy skills.

Moderately Different: Includes individuals with moderate levels of English language proficiency (e.g., intermediates to advanced BICS) and moderate levels of acculturation. Examples include individuals who have resided in the U.S. for 3-7 years and who have learned English well enough to communicate, but whose parents are limited English speakers with only some formal schooling, and improving but below grade level literacy skills.

Highly Different: Includes individuals with low to very low levels of English language proficiency (e.g., early BICS) and low or very low levels of acculturation. Examples include individuals who recently arrived in the U.S. or who may have been in the U.S. 3 years or less, with little or no prior formal education, who are just beginning to develop conversational abilities and whose literacy skills are also just emerging.

Figure 5. General guidelines for expected pattern of test performance.

Using the C-LTC and C-LIM

On a subtest that has been identified as highly linguistically demanding, the performance of an ELL student is likely to be negatively impacted to a large degree. The scores on less linguistically demanding subtests will be less adversely impacted. The same will hold true for subtest scores based on their cultural loading. The diagonal arrow on the Matrix indicates the combined effect that cultural loading and language differences may have on test performance as a combined function of both factors.

When using the C-LTC and the C-LIM, practitioners first identify appropriate tests to administer that relate to the initial referral concern based on the CHC constructs. Then, assessors narrow those choices to the subtests that have the most appropriate level of cultural loading and linguistic demand for that student in consideration of their language proficiency levels (i.e., English and native language) and acculturation. After the assessment has been completed, the individual subtest scores are then recorded into one of the nine cells on the C-LIM. If tests from more than one battery are used, it may be necessary to convert the scores into a common metric (where the mean standard scores is 100 and the standard deviation 15).

After the data are calculated, assessors examine the average of each cell on the C-LIM to compare them across cells from left to right, and down cells from top to bottom. It is not the normative position of the scores (high, moderate, low) but

the relationships between the scores and the degree to which they form a pattern that is either consistent or inconsistent with the pattern of performance predicted by the matrix. Three general patterns may emerge: (a) scores decrease as the move down the cells in the matrix (the effect of cultural loading only), (b) scores decrease as they move across the cells from left to right in the matrix (the effect of linguistic demand only), or (c) scores in or near the upper left corner of the matrix may be higher than scores at or near the bottom right corner of the matrix (the overall effect of both culture and language). It is important to look for the interaction effect of acculturation and language proficiency because these factors are not perfectly correlated.

When patterns emerge from the data that are not consistent with the expected general pattern of performance for ELL students, then practitioners should look for inter- and intra-cognitive analyses conducted previously and base interpretation on results at that level. It is when the patterns diverge from the expected pattern that attenuated scores may not simply reflect an individual's cultural and linguistic differences, but may reflect some inherent weaknesses (Flanagari & Ortiz, 2001). Flanagan and Ortiz (2001) cautioned that the classifications are not necessarily definitive but are *subjective*. The classifications were derived from existing research, expert consensus and clinical judgment (Cummins, 1982; Jensen, 1974; Sanchez, 1934; Vukovich & Figueroa, 1982; Yerkes, 1921) and may change according to future research.

Using the classifications by themselves is *insufficient* to establish a comprehensive basis for the assessment of diverse individuals. Instead, they are *supplemental* to the assessment process and do not negate the need for interpreting test results within the context of thorough knowledge of the student's educational, experiential, socioeconomic, linguistic and cultural backgrounds. As Esters and Ittenbach (1997) wrote, this framework

offers clinicians a practical and expedient means by which to evaluate one aspect of the race, culture, and social class influence on tests and children's response patterns. While is it admittedly subjective and open to criticisms, their approach is indeed one place to begin the search for culturally aware response options at the service delivery level. (p. 217)

Although the framework appears to provide a systematic method for interpreting diverse students' standardized cognitive test scores, it is relatively new with little research to support its use, particularly when administering Spanish language cognitive assessments.

Brown (2005) examined the use of C-LIM (Flanagan & Ortiz, 2001) to determine if its use would reveal evidence of possible inappropriate diagnosis of a disability in ELL students. She found that in 52% of the cases examined, it was likely that ELL students had been inappropriately identified. To conclude, while this framework looks promising, more research is needed, particularly to examine its use with Spanish cognitive tests, particularly the *Bateria III*.

In summary, after an exhaustive review of the literature, it is apparent that there is little guidance on the appropriate assessment of intellectual/cognitive

abilities in ELL students. Since these students are immersed in two linguistic worlds, assessment in either L1 or L2 will not allow them to demonstrate the entire repertoire of their abilities. Yet, the literature is void of how bilingual students in the U.S. perform on a Spanish language cognitive test that is normed on monolingual Spanish-speakers. Further, the one potential framework for nondiscriminatory interpretation of test performance, the C-LTC and the C-LIM, has only looked at how ELL students perform on English cognitive tests. Current research links CHC abilities to academic achievement so it is also important to examine CHC clusters that may aide in intervention or program planning. This study examined the issues outlined here by responding to the following general research questions.

Research Questions

1. What is the difference in overall performance as measured by the Global Intellectual Ability (GIA), subtest, and cluster scores between the monolingual Spanish normative group and the U.S. bilingual comparison group?
2. What is the relationship between performance as measured by GIA and English language proficiency, Spanish language proficiency, and acculturation levels?
3. How do the patterns of subtest mean scores achieved by general education U.S. bilingual participants on the *Batería III* vary as a function of cultural

loading and linguistic demands of the subtests in the manner predicted by the WJ-III C-LIM?

4. How do participants perceive the experience of native language assessment?

CHAPTER III

METHOD

The following chapter describes the procedures used in this study. First the participants and criteria for inclusion in the sample groups are presented. Then, the procedures and instruments used for data collection, and the data analysis procedures are presented.

Participants

Six school districts in the Pacific Northwest region with large ELL student populations were contacted. Given the focus on NCLB and high-stakes testing, most districts declined citing they were not allowing students to miss instructional time. Originally, permission to conduct this study during the school day was sought. Ultimately, the participating school district arranged for the study to be conducted on three separate Saturdays.

One rural school district granted permission for the research. Two bilingual staff members agreed to support this project. One elementary school with the highest numbers of ELL students was targeted. This school had both a Native Language Development (NLD) program and ESL (English as a Second Language/English Language Development or ELD) program. The NLD program provided 90 minutes to two hours of language arts in Spanish to students in grades K-6 who qualified for the ELL program as a "limited English speaker." The ELD

program was provided for students whose parents declined native language instruction and consisted of 30 minutes of daily pull-out instruction to improve their English language proficiency.

Families of third and fourth grade ELL students were invited to a pizza dinner meeting for the researcher to explain the project, answer questions, and meet the community. There was, however, limited parent attendance, possibly due to weather conditions. As a follow up, letters in English and Spanish describing the study and inviting participation (see Appendices A and B) were sent to the homes of all third and fourth graders meeting the criteria (described below) in the school (and later the district). The letter also included information that either parents or children could decline participation in the study at any time with no repercussions or penalty. As the project progressed, the bilingual district staff assisted the researcher in contacting other parents and soliciting both permission and assessment appointments for the students.

Of the 5,144 students in the K-12 district, 493 students were enrolled in special education programs (9.58%) and 803 students were classified as English Language Learners (15.61%). The ELL student population in the district was predominantly Hispanic. The second group, though considerably smaller, was Russian-speaking students. Students within the district who met the following criteria were invited to participate:

1. Enrolled currently in either a Native Language Development (NLD) model or English as a Second Language (ESL) model of bilingual education
2. Had Spanish as a home language
3. Enrolled in grades 3 and 4
4. Enrolled in general education
5. Not enrolled in special education nor identified as having any type of disability

Table 3 shows the characteristics of the research sample. Eight students were born in Mexico while 26 were born in the United States. Most had received their entire education thus far in the participating school district and were representative of the relatively stable population within the close knit community.

Table 3

Characteristics of Hispanic Third and Fourth Grade Sample

Grade	Age Range	Median Age	Gender		Number
			M	F	
Gr. 3 & 4 (Total Sample)	8.2 – 10.11	9.02	15	19	34
Grade 3	8.2 – 10.0	8.8	8	11	19
Grade 4	9.5 – 10.11	9.9	7	8	15

Study Design

In this study, there were four independent variables (1) native language development instructional model (NLD) or English language development instructional model (ELD), (2) level of English language proficiency, (3) level of

Spanish language proficiency, and (4) level of acculturation. The effects of the independent variables on performance were examined in relation to the dependent variables of GIA score, subtest scores, and cluster scores (see Table 4).

Table 4

Independent and Dependent Variables

Independent Variables	Dependent Variables
Native language development instructional model (NLD)	Cognitive ability performance as measured by the GIA score.
English language development instructional model (ELD)	Cognitive ability performance as measured by the subtest scores.
Level of English language proficiency	Cognitive ability performance as measured by the cluster scores.
Level of Spanish language proficiency	
Level of acculturation	

Comparison Groups

To examine the performance of students living in the United States who are bilingual to differing degrees, four comparison groups were used. The first was the monolingual Spanish norm sample of the Batería (described below), the second was the NLD group, the third was the ELD group and the fourth group was the total bilingual group (NLD + ELD). Each group represented a different and unique population.

The Spanish norm sample population of the *Batería III* included 1,413 total monolingual Spanish-speakers. The 1,134 subjects were from Mexico, Costa Rica, Panama, Argentina, Colombia, Puerto Rico and Spain and the United States and 279 were from within the United States. The test manual indicated that strict criteria were used to verify that each subject's native and primary language was Spanish and Spanish was their dominant language as well (Schrack et al., 2005).

This study is quantitative in nature since the researcher attempted to quantify differences between the mean scores of the groups and relationships were investigated among variables. Miles and Huberman (1994) stated that quantitative research classifies features, counts them, and constructs statistical models in an attempt to explain what is observed. The strengths of quantitative research are that the methods generally produce quantifiable, reliable data that can be generalized to larger populations.

Data Collection

Six female bilingual school psychologists, including the researcher, with master's degrees and at least 5 years of experience in bilingual assessment, assessed two students per session. The assessors had received formal training on the WJ-III. Prior to the initial assessment session, the researcher trained the school psychology team on the specifics of the *Batería III*. The next section describes the three assessments that were used in this study.

Assessment Instruments

Batería III Woodcock-Muñoz Pruebas de Habilidades Cognitivas

The *Batería III* Woodcock-Muñoz Pruebas de Habilidades Cognitivas is the Spanish version of the Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III COG). Each subtest of the *Batería III* Pruebas de habilidades cognitivas (*Batería III* COG) was translated or adapted from the English WJ-III COG. The *Batería* COG, whose theoretical base is the CHC Theory, includes 20 subtests for measuring general intellectual ability, broad and narrow cognitive abilities (as identified previously in Figure 1), and aspects of executive functioning. The 20 tests are organized in two batteries – the Standard Battery (COG Tests 1-10) and the Extended Battery (COG Tests 11-20). Each of the 20 tests measures one or more narrow cognitive abilities and one measure of a broad cognitive ability identified in the CHC Theory (Figure 1). The 20 tests are organized into clusters. Raw scores are transformed into derived scores using a computer software program. The test yields an estimate of general ability or intelligence (GIA) as well as a standard score (where the mean score is 100 and the standard deviation is 15) with percentile ranks for each of the tests and clusters (Schrang et al., 2005). The general age range for this assessment is 5 to 95 years of age.

The *Batería III*, the third version of the Spanish cognitive battery, had input from numerous professionals from various regions and countries of the Spanish-speaking world. The *Batería III* tests are translations or adaptations of their parallel test from the WJ-III. All translations and adaptations were done by professionally

certificated Spanish translators who were native speakers of Spanish from Mexico, Puerto Rico, Spain, Argentina, Panama, Costa Rica, and Colombia. The manual stated that when all test items remained the same, items were translated. When direct translation was not possible, translation and adaptation occurred (Schrack et al., 2005).

The *Batería III* data were equated to the WJ-III norms. Each task on each test was scaled according to their difficulty on the parallel WJ-III test. Calibration data were collected on 1,413 subjects who were native Spanish speakers living in Latin America, Puerto Rico, Spain or from the United States. Of the 1,413 participants, only 279 lived in the United States. To be included in the sample, each participant from the United States was given an oral language test in both English and Spanish to verify Spanish dominance. Participants living outside of the United States were selected based on an informant's opinion of the examinees native and primary language use (Otero, 2006). Information on gender and age of the norm sample participants in the calibration sample were not available in the manual.

"The *Batería III* cognitive subtests have internal consistency reliabilities coefficients ranging from .88 to .94" (Otero, 2006, p. 88). Coefficients between .35 and .65 are considered to be moderate, and coefficients higher than .65 are considered to be high (Gay, Mills, & Airasian, 2006) indicating this instrument is a reliable and valid instrument for individuals whose profiles fit within the norm sample. Table 5 provides a description of the subtests administered.

Table 5

Description of 14 Subtests from the Bateria III Tests of Cognitive Abilities

CHC Factor	Subtest	Subtest Requirements
<i>Gc</i>	Verbal Comprehension/ <i>Comprensión verbal</i>	Examinees must name familiar and unfamiliar objects pictured, say words similar in meaning to word presented, say words that are opposites in meaning to the word presented, and complete phrases with words that complete analogies.
<i>Gc</i>	General Information/ <i>Información general</i>	Examinees must associate new visual symbols with orally presented words in order to translate the series of symbols into meaningful sentences.
<i>Glr</i>	Visual-Auditory Learning/ <i>Aprendizaje visual-auditivo</i>	Examinees must associate new visual symbols with orally presented words in order to translate the series of symbols into meaningful sentences.
<i>Glr</i>	Retrieval Fluency/ <i>Fluidez de recuperación</i>	Examinees must state as many words as possible in 1 minute from each of three specified categories.
<i>Gv</i>	Picture Recognition/ <i>Reconocimiento de dibujos</i>	Examinees must study images for 5 seconds and identify those images within a larger array of images after the initial images have been removed.
<i>Gv</i>	Spatial Relations/ <i>Relaciones espaciales</i>	Examinees must select the component parts of whole shapes.
<i>Gf</i>	Concept Formation/ <i>Formación de conceptos</i>	Examinees must identify the rules governing the organization of colored geometric figures when shown instances and non-instances of concepts.
<i>Gf</i>	Analysis-Synthesis/ <i>Análisis-Síntesis</i>	Examinees must analyze the components of incomplete logic puzzles and determine the missing components.
<i>Gs</i>	Visual Matching/ <i>Pareo visual</i>	Examinees must rapidly scan successive rows of six numbers and circle the two numbers in each row that are identical during a 3-minute period.
<i>Gs</i>	Decision Speed/ <i>Rapidez en la decisión</i>	Examinees must rapidly scan successive rows of images and circle the two images in each row that are most closely related during a 3-minute period.
<i>Ga</i>	Auditory Attention/ <i>Atención auditiva</i>	Examinees must detect differences discriminate words given via a tape recording under conditions of increasing distortion and point to the corresponding picture in a field.

Table 5 (continued)

Description of 14 Subtests from the Bateria III Tests of Cognitive Abilities

CHC Factor	Subtest	Subtest Requirements
<i>Gsm</i>	Numbers Reversed/ <i>Inversión de números</i>	Examinees must repeat a series of random numbers backward.
<i>Ga</i>	Sound Blending/ <i>Integración de sonidos</i>	Examinees must listen to a series of individual syllables, individual phonemes, or both that form words and name the complete words.
<i>Gsm</i>	Memory for Words/ <i>Memoria para palabras</i>	Examinees must listen to lists of unrelated words and repeat them in the exact order presented.

Source: Floyd, Bergeron, and Alfonso (2006).

The Language Assessment Scales (LAS) – English and Spanish

The Language Assessment Scales (DeAvila & Duncan, 2005) – Oral (LAS-O) in English and Spanish were administered to each participant. The LAS, for participants in grades 1 through 6, was normed on 1,671 English learners throughout the United States. It is an individually administered test that consists of vocabulary, listening comprehension, story retelling, minimal sound pairs and phoneme subtests. For the purposes of this research, the first three subtests that comprise the LAS-O were used to measure students' oral proficiency. These three subtests are commonly used for language proficiency monitoring in schools. The story retelling component is 50% of the score. The LAS generates scores from 1 (non-English or non-Spanish speaking) to 5 (fluent English speaker or fluent Spanish speaker). The test authors stated the following about the theoretical underpinnings of this assessment:

The development of the LAS was based on a view of language as

consisting of four linguistic aspects: phonology (phonemes, stress, rhythm and intonation), the lexicon (the “words” of the language), syntax (the rules for comprehending and producing meaningful utterances) and pragmatics (the appropriate use of language to obtain specific goals. (DeAvila & Duncan, 1990, p. 8)

The reliability correlation coefficients range from .87 to .88 and the manual stated evidence for construct validity. Again, according to Gay et al. (2006) this coefficient is considered high.

Acculturation Quick Screen

Each participant’s cumulative file was reviewed to gather information related to grade, age, country of birth and educational history, information needed to complete the Acculturation Quick Screen (AQS) (Collier, 1988). Currently, there are very few instruments that provide a measure of acculturation levels. These tools rely heavily on a student’s native and second language proficiency scores (Ortiz & Flanagan, 2002). The AQS, first developed by Collier in 1985, is a scale constructed to gauge a student’s adaptation to the mainstream school environment (see Appendix E). It is based on Padilla’s (1980) multidimensional theory of acculturation (Garcia-Vazquez, 1995). The items on the AQS address number of years in the U.S., in the current school district, model of bilingual programs, level of English and native language proficiency, ethnicity or nation of origin. For each item, the rater circles the appropriate choice. Then, the raw score is transformed into a scaled score which ranges from 8 (less acculturated) to 40 (more

acculturated). No validity or reliability data were reported nor were available from the author (C. Collier, personal communication, February 1, 2008) or through procuring and reading the dissertation that provided the initial data for constructing the AQS (Collier, 1986). Interestingly, this tool is commonly used in school districts' across the Northwest to assess students' acculturation levels. Garcia-Vazquez (1995) looked at the effects of acculturation on reading achievement in which the AQS was used. Garcia-Vazquez stated "No validity or reliability data are reported" (p. 309) for the AQS.

Assessment of Participants

Seven subtests from the standard battery and seven subtests from the extended batteries of the *Batería III* that comprise the seven broad clusters (14 subtests in all) were administered to each participant. Individual classrooms were used for the assessments to ensure quiet and relative privacy. After the testing session ended, each participant's individual test protocol was scored using the computerized program for the *Batería III*. The scores were then entered into a database using the SPSS computer program for later data analysis. After completing the *Batería*, participants were administered the oral portions of the English Language Assessment (LAS) Scales and the Spanish Language Assessment Scales (LAS) to determine language proficiency levels in both languages. The Story Retell section of both the English and Spanish LAS were tape recorded and later transcribed by each assessor. Each participant's cumulative file was reviewed

to extract the information necessary to complete the Acculturation Quick Screen (AQS) and calculate the acculturation level/score. Throughout the assessment time, participants were provided with healthy snacks, juices, and rest time. At the end of their assessment session, participants were given a book.

Data Analysis

Data were analyzed using SPSS Version 12 for all research questions. An alpha level of .05 was used. Effect sizes were reported for all significant results. See Table 6 for the research questions and methods of data analysis.

Ethical Considerations

Participation in this project was strictly voluntary so participants and their parents understood that there was no pressure. Individual testing sessions were scheduled on weekends so as not to disrupt instructional time. The bilingual researcher and assistants reminded participants and parents that confidentiality would be maintained and that they were free to withdraw from the study at any time without penalty or consequences.

Individual testing sessions were held at the child's school whenever possible. The researchers administered the test following the standard procedure and recorded the participant's responses and scores on the test protocol. The *Batería* sessions were not be audiotaped. The only portion audiotaped was the students' story retellings from the LAS English and Spanish Oral test. The students verbally agreed to this. The researchers were trained and experienced in working

Table 6

Research Questions and Methods of Data Analysis

Research Questions	Method of Data Analysis
Question 1: What is the difference in overall performance as measured by the Global Intellectual Ability (GIA) subtest, and cluster scores between the monolingual Spanish normative group and the U.S. bilingual comparison group?	
1.a. What is the difference in performance between the normative group ($\bar{X} = 100$) and the total bilingual group (NLD + ELD) as measured by the GIA scores?	Independent sample <i>t</i> test
1.b. What is the difference in performance between the normative group ($\bar{X} = 100$) and the total bilingual group (NLD + ELD) as measured by the 14 subtest mean scores?	Independent sample <i>t</i> test
1.c. What is the difference in performance between the normative group ($\bar{X} = 100$), the NLD and ELD groups as measured by the GIA scores?	ANOVA
1.d. What is the difference in performance between the normative group ($\bar{X} = 100$), the NLD and ELD groups as measured by the 14 subtest scores?	ANOVA
1.e. What is the difference in performance between the normative group ($\bar{X} = 100$) and the NLD group as measured by the CHC cluster mean scores?	Independent Samples <i>t</i> test
1.f. What is the difference in performance between the normative group ($\bar{X} = 100$) and the ELD group as measured by the CHC cluster mean scores?	Independent Samples <i>t</i> test
Question 2: What is the relationship between performance as measured by GIA and English language proficiency, Spanish language proficiency and acculturation level?	
2.a. What is the relationship between performance as measured by the GIA scores and level of English language proficiency in the NLD and ELD groups?	Pearson Correlation
2.b. What is the relationship between performance as measured by the GIA scores and level of Spanish language proficiency in the NLD and ELD groups?	Pearson Correlation
2.c. What is the relationship between performance as measured by the GIA scores and level of acculturation in the NLD and ELD groups?	Pearson Correlation
Question 3: How do the patterns of subtest mean scores achieved by general education U.S. bilingual participants on the <i>Batería III</i> vary as a function of cultural loading and linguistic demands of the subtests in the manner predicted by the WJ-III Culture-Language Test Classification?	
3.a. How do the <i>Batería's</i> 14 subtest mean scores for the ELD group follow the expected pattern of performance for diverse individuals when tested in English on the C-LIM?	Scores Plotted onto C-LIM
Question 4: How do participants perceive the experience of native language assessment?	
4.a. What languages do you think in?	Percentages
4.b. What language do you like to use?	
4.c. Describe what was easy about the testing.	
4.d. Describe what was hard about the testing.	

with children between the ages of 3-21 and provided a low-key environment for the assessment. Also, the researchers each spoke the children's native language and were familiar with the cultural nuances of the Latino population which added to the students comfort. Information and test protocols were coded to protect the identities of study participants. Confidentiality ensured that the participation had no impact on participants' educational program.

CHAPTER IV

PRESENTATION OF DATA

This study investigated the performance of ELL students on a Spanish cognitive test. The results of the data for the four research questions are reported. First, descriptive statistics provide information on the independent and dependent variables used in subsequent analyses. Second, inferential analyses, to test the research questions, included *t* tests, ANOVA, and correlation analyses. Independent *t* tests are used when scores come from different groups of people (Field, 2005). Because the number of participants in the *Batería III*'s normative sample ($n = 1,413$) and the research sample ($n = 34$) differ by a large amount and the independent *t* test assumes equal variances, Levene's test was consulted. The last analysis compared subtest mean scores of the ELD group to Flanagan and Ortiz's (2001) predicted pattern of performance on the C-LIM for diverse individuals when tested on the WJ-III. To aide the reader, Table 7 provides the names (English and Spanish) of the broad abilities and a brief definition and the subtests comprising the clusters. Although the subtests administered were in Spanish, the English names are used throughout for ease of reading.

Table 7

CHC Factors, Definition, and Subtests

CHC Broad Ability Factor	Definition	Subtests that Form Cluster
<i>Gf</i> <i>Fluid Intelligence</i>	Thinking used: <ul style="list-style-type: none"> with new tasks which cannot be performed automatically for reasoning, forming and recognizing concepts; comprehending implications for drawing inferences in reorganizing or transforming information for problem-solving and extrapolating 	Analysis-Synthesis/ <i>Análisis-Síntesis</i> Concept Formation/ <i>Formación de conceptos</i>
<i>Gc</i> <i>Crystallized Intelligence</i>	Knowledge: <ul style="list-style-type: none"> which has been acquired over time (general fund) of one's culture and the effective application of this knowledge (breadth and depth) that is primarily verbal- or language-based developed during life experiences and formal schooling 	Verbal Comprehension/ <i>Comprensión verbal</i> General Information/ <i>Información general</i>
<i>Gv</i> <i>Visual Processing</i>	Ability to: <ul style="list-style-type: none"> think about visual patterns and visual stimuli generate, perceive, analyze, synthesize, manipulate, and transform visual patterns and stimuli complete puzzles and interpret graphs or charts visualize stimuli not presented visual (the "mind's eye") 	Picture Recognition/ <i>Reconocimiento de dibujos</i> Spatial Relations/ <i>Relaciones espaciales</i>
<i>Gsm</i> <i>Short-Term Memory</i>	Ability to: <ul style="list-style-type: none"> apprehend and hold information in immediate awareness and then use it within a few seconds retain "chunks" of information (most individuals can only retain five to seven "chunks" at one time) 	Retrieval Fluency/ <i>Fluidez de recuperación</i>
<i>Glr</i> <i>Long-Term Retrieval</i>	Ability to: <ul style="list-style-type: none"> store information (e.g., concepts, ideas, items, names) in long-term memory and to fluently retrieve it later through association. (Does not represent <i>what</i> is stored in long-term memory but the <i>process</i> of storing and retrieving information.) 	Visual-Auditory Learning/ <i>Aprendizaje visual-auditivo</i> Retrieval Fluency/ <i>Fluidez de recuperación</i> Sound Blending/ <i>Integración de sonidos</i>
<i>Ga</i> <i>Auditory Processing</i>	Ability to: <ul style="list-style-type: none"> perceive, analyze, and synthesize auditory stimuli perceive and discriminate subtle nuances of patterns of sounds 	Auditory Attention/ <i>Atención auditiva</i>
<i>Gs</i> <i>Processing Speed</i>	Ability to: <ul style="list-style-type: none"> fluently perform simple clerical-type tasks quickly, especially when under pressure to maintain concentration and attention take simple tests that require simple decisions 	Decision Speed/ <i>Rapidez en la decisión</i> Visual Matching/ <i>Pareo visual</i>

Source: McGrew and Flanagan (1998).

Descriptive Statistics for the Independent Variables

GIA Scores

Table 8 shows the English LAS proficiency scores (1 through 5 with 1 being a Non-English speaker and 5 a Fluent English speaker) for the NLD group ($n = 16$) with the corresponding GIA (i.e., IQ) scores. The LAS scores were within the range of 3 to 5. Five participants scored at a LAS 3, Limited English Speakers. Ten received a LAS 4 score, and one participant scored at a LAS 5. LAS 4 and LAS 5 were Fluent English Speakers. Participants who scored a LAS 5 are considered to have an English proficiency level equivalent to native English speakers.

Table 8

GIA Scores by English Language Proficiency for NLD Group

GIA Score	LAS – Eng 1	LAS – Eng 2	LAS – Eng 3	LAS – Eng 4	LAS – Eng 5
71				1	
81				1	
82				1	
84			1		1
89			1		
90				1	
92			1	2	
95			1	2	
97			1		
101				1	
113				1	
Total	0	0	5	10	1
<i>(n = 16)</i>					

English Proficiency Scores

Table 9 shows the English LAS proficiency scores for the ELD group ($n = 18$) with the corresponding GIA (i.e., IQ) scores. The scores on the LAS ranged between LAS level 2 and 5. Two participants scored at a level 2 and five participants as LAS 3, Limited English Speakers. Eleven participants scored at LAS 4 ($n = 8$) and 5 ($n = 3$) are Fluent English Speakers. The three participants at LAS 5 are considered to have English skills comparable to a native English speaker.

Table 9

GIA Scores by English Language Proficiency for ELD Group

GIA Score	LAS – Eng 1	LAS – Eng 2	LAS – Eng 3	LAS – Eng 4	LAS – Eng 5
69					1
70			1		
73		1			
74					1
77				1	
78				1	
79			1		
81				1	
82			1		
84				1	
85				1	
89			1	1	
91		1			
92				1	
95				1	
106			1		
116					1
Total	0	2	5	8	3
<i>(n = 18)</i>					

Table 10 shows the frequency in which participants in both the NLD and ELD groups scored at each English proficiency score level. The largest portion of participants scored in the range of 3 to 4. No one scored a level 1. There were an equal amount of participants (5) who scored a LAS level 3 in English even though each group received instruction in a different language. Ten participants from the NLD group scored a level 4 while eight participants in the ELD group scored a level 4. There were more participants in the ELD group who scored as a Fluent English Proficient participant while only one participant in the NLD (Spanish instruction) group scored a 5.

Table 10

Frequency of English Language Proficiency Scores NLD and ELD Groups

LAS Score English	Count NLD	Frequency NLD	Count ELD	Frequency ELD
1	0	0%	0	0%
2	0	0%	2	11.1%
3	5	31.3%	5	27.8%
4	10	62.5%	8	44.4%
5	1	6.3%	3	16.7%
Total	16	100.0%	18	100.0%

Spanish Proficiency Scores

Table 11 shows the Spanish language proficiency scores for the NLD group ($n = 16$) with the corresponding GIA (i.e., IQ) scores. The LAS scores ranged from 1 to 5 (1 through 5 with 1 being a Non-Spanish speaker and 5 a Fluent Spanish

speaker). One participant scored at a LAS 1, “Non-Spanish Speaker.” One participant scored a LAS 2 and four participants scored a LAS 3. They are considered Limited English Speakers. Nine participants scored LAS 4 and one participant LAS 5. These participants are considered “Fluent Spanish Speakers.” The participant who scored LAS 5 is considered to have native-like fluency.

Table 11

GIA Scores by Spanish Language Proficiency for NLD Group

GIA Score	LAS-Span 1	LAS-Span 2	LAS-Span 3	LAS-Span 4	LAS-Span 5
71				1	
81		1		1	
82				1	
84	1			1	
90			1		
92			1	2	
95				2	1
97			1		
101			1		
113				1	
Total	1	1	4	9	1

(*n* = 16)

Table 12 shows the Spanish language proficiency scores for the ELD (*n* = 18) group which ranged from LAS 1 to 4. Two participants scored a LAS 1 and are “non-Spanish Speakers.” Five participants scored a LAS 2 and six participants a LAS 3. They are considered “Limited Spanish Speakers.” Five

participants scored a LAS 4 and are considered Fluent English Speakers. No participants scored a LAS 5.

Table 12

GIA Scores by Spanish Language Proficiency for ELD Group

GIA Score	LAS-Span 1	LAS-Span 2	LAS-Span 3	LAS-Span 4	LAS-Span 5
69		1			
70	1				
73			1		
74				1	
77			1		
78			1		
79			1		
81		1			
82				1	
84		1			
85		1			
89	1			1	
91				1	
92			1		
95			1		
106				1	
116		1			
Total	2	5	6	5	0

(n = 18)

Table 13 provides the frequency of each Spanish language proficiency score for both the NLD and ELD groups. The majority of the total sample scored a LAS level 4, followed by a LAS level 3 (58.3%), with the smallest group (6.3%) scoring

a LAS 5. One participant in the NLD group scored at the beginning stage (Level 1) in Spanish while two ELD participants also scored at Level 1. Most participants (a total of 24) scored a LAS level 3 and 4 or at the “Limited Spanish Speaker” level with more participants from the NLD group (9) scoring a 4 than from the ELD group (5). One participant from the NLD group scored as “Fully Spanish Proficient” while no one in the ELD group received this score.

Table 13

Frequency of Spanish Language Proficiency Scores of NLD and ELD Groups

LAS Score Spanish	Count NLD	Frequency NLD	Count ELD	Frequency ELD
1	1	6.3%	2	11.1%
2	1	6.3%	5	27.8%
3	4	25.0%	6	33.3%
4	9	56.3%	5	27.8%
5	1	6.3%	0	.0%
Total	16	100.0%	18	100.0%

Acculturation Levels

Table 14 shows the acculturation levels for the NLD group ($n = 16$) as measured by the AQS scores fell in the AQS 3 and 4 columns only. Eleven participants scored an AQS 3 indicating that they are “in transition.” Five participants scored an AQS 4 meaning they are “more acculturated.” No participants scored a level 5 or “fully acculturated.”

Table 14

GIA Scores by Acculturation Level for NLD Group

GIA Score	AQS 1	AQS 2	AQS 3	AQS 4	AQS 5
71			1		
81				1	
82				1	
84			3		
90				1	
92			3		
95			3		
97			1		
101				1	
113				1	
Total	0	0	11	5	0

(*n* = 16)

Table 15 shows that acculturation levels for the ELD group (*n* = 18) which mainly fall between 3 and 4. Eleven participants scored an AQS level 3 meaning they are “in transition” of acculturating. Six participants scored a level 4 meaning they are “more acculturated.” One participant, who had the highest GIA score, scored a level 5 indicating “highly acculturated.”

Table 15

GIA Scores by Acculturation Level for ELD Group

GIA Score	AQS 1	AQS 2	AQS 3	AQS 4	AQS 5
69				1	
70			1		
73			1		
74				1	
78				1	
79				1	
81			1		
82			1		
84			1		
85			1		
89			2		
91			1		
92				1	
95			1		
97				1	
106			1		
116					1
Total	0	0	11	6	1

(n = 18)

To summarize, the levels of English language proficiency between the NLD and ELD groups were very similar with the exception of two participants in the ELD group who scored a LAS 2, or lower than the rest of the participants in the two groups. The levels of Spanish language proficiency were also similar with two participants in the ELD group and one in the NLD group scoring as non-Spanish speakers. As a group, the NLD group received more scores in the LAS 3 and 4 range indicating higher overall Spanish language proficiency. The scores on acculturation as measured by the AQS also showed very similar scores across the two groups with the exception of one participant in the ELD group who scored a 5 which is considered highly acculturated.

Table 16 shows the mean scores and standard deviations for all 14 subtests for the total bilingual group. The range in the mean scores was 71.88 to 108.74 or a difference of 36.86 points. West, Finch, and Curran (1995) recommended skewness lower than or equal to 2 and kurtosis less than or equal to 7 to be acceptable. All scores fell into the acceptable range.

Table 16

Subtest Scores for Total Bilingual Group (NLD + ELD)

	Valid <i>N</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
Global Intellectual Ability (GIA)	34	87.44	11.77	.30	.44
Auditory Attention (<i>Ga</i>)	34	108.74	14.72	-.69	.23
Sound Blending (<i>Ga</i>)	33	108.09	12.50	-.53	.32
Picture Recognition (<i>Gv</i>)	34	105.38	9.56	.15	-.57
Decision Speed (<i>Gs</i>)	34	100.56	17.48	-.01	.40
Retrieval Fluency (<i>Glr</i>)	34	90.26	23.10	-1.6	5.80
Spatial Relations (<i>Gv</i>)	34	98.06	9.43	.60	.76
Analysis Synthesis (<i>Gf</i>)	34	98.24	12.73	.51	-.14
Visual Matching (<i>Gv</i>)	34	94.32	14.56	-.35	.09
Concept Formation (<i>Gf</i>)	34	93.94	12.78	.66	.50
Verbal Comprehension (<i>Gc</i>)	34	85.38	13.20	-.51	.39
Memory for Words (<i>Gsm</i>)	34	91.62	11.28	1.21	2.40
Numbers Reversed (<i>Gsm</i>)	33	87.18	15.78	-1.50	3.81
Visual-Auditory Learning (<i>Glr</i>)	34	86.12	17.50	-.71	-.34
General Knowledge (<i>Gc</i>)	34	71.88	11.53	-.97	1.1

Table 17 shows the mean GIA and 14 subtest scores for the norm, NLD and ELD groups. The range of the mean scores for the NLD group was between 76.44 (General Information) to 115.63 (Auditory Attention). The range of mean scores for the ELD group was between 67.83 (General Information) and 107.39 (Sound Blending).

Table 17

Means and Standard Deviations for All Groups

Subtests	<u>Normative</u>			<u>NLD</u>			<u>ELD</u>		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
GIA	100	15	1413	90.81	9.50	16	84.44	13.00	18
Verbal Comp. (<i>Gc</i>)	100	15	1413	93.0	9.03	16	78.61	12.78	18
Vis-Aud Lrng. (<i>Glr</i>)	100	15	1413	89.75	17.79	16	82.89	17.09	18
Spatial Rel. (<i>Gv</i>)	100	15	1413	99.38	10.96	16	96.89	7.96	18
Sound Blndg. (<i>Ga</i>)	100	15	1413	108.93	12.86	15	107.39	12.51	18
Concept Frm. (<i>Gf</i>)	100	15	1413	93.94	13.69	16	93.94	12.33	18
Visual Mtchng. (<i>Gs</i>)	100	15	1413	97.19	11.26	16	91.78	16.88	18
General Info. (<i>Gc</i>)	100	15	1413	76.44	5.36	16	67.83	13.99	18
Retrieval Fluency(<i>Glr</i>)	100	15	1413	99.56	13.25	16	82.00	26.97	18
Picture Recog. (<i>Gv</i>)	100	15	1413	104.50	10.27	16	106.17	9.10	18
Auditory Attn. (<i>Ga</i>)	100	15	1413	115.63	11.53	16	102.61	14.80	18
Dec Speed (<i>Gs</i>)	100	15	1413	102.88	19.71	16	98.50	15.52	18
Mem. For Words (<i>Gsm</i>)	100	15	1413	90.31	8.95	16	92.78	13.16	18
Sound Blndg. (<i>Ga</i>)	100	15	1413	108.93	12.86	15	107.39	12.51	18
Anal. Syn (<i>Gf</i>)	100	15	1413	98.38	11.91	16	98.11	13.76	18
Nmbrs Rvd (<i>Gsm</i>)	100	15	1413	89.81	14.15	16	84.71	17.23	17

Question One

What is the difference in overall performance as measured by the Global Intellectual Ability (GIA), subtests, and cluster scores between the monolingual

Spanish normative sample and the U.S. bilingual comparison group? The analyses for Questions 1A through 1F address Question One.

Question 1A

What is the difference in performance between the normative sample ($\bar{X} = 100$) and the total bilingual group (NLD + ELD) as measured by the Global Intellectual Ability (GIA) mean scores?

There was a significant difference between bilingual and normative sample, $t = 4.8, p < .001$, with the normative sample ($X = 100$) scoring significantly higher on the GIA than the bilingual group ($\bar{X} = 87.4$). A small effect size, Cohen's $d = .26$ was found (Cohen, 1988).

Question 1B

What is the difference in performance between the normative sample ($X = 100$) and the total bilingual (NLD + ELD) group as measured by the 14 subtest mean scores?

Table 18 shows the differences in scores between the normative sample and total bilingual group on the 14 subtests. To summarize, a significant difference was found between the GIA scores of the normative sample ($\bar{X} = 100$) and the bilingual (NLD + ELD) group ($\bar{X} = 97.4$). Significant differences were also found between the normative sample and bilingual group on eight subtests with the normative sample scoring significantly higher. Further significant differences were

found between the bilingual group and normative sample on three subtests in which the bilingual group scored significantly higher.

There were no significant score differences between the groups on Spatial Relations, Analysis-Synthesis, and Decision Speed.

Table 18

Means and Standard Deviations for Normative Sample and Bilingual Group

	<u>Normative Sample</u>		<u>Bilingual Group</u>		<i>t</i>	<i>D</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Global Intellectual Ability Score (GIA)	100	15	87.4	11.77	4.84***	.26
Subtests						
Verbal Comprehension	100	15	85.4	13.20	5.63*	.30
Visual-Auditory Learning	100	15	86.1	17.50	5.31***	.28
Spatial Relations	100	15	98.06	9.43	1.17	
Sound Blending	100	15	108.09	12.50	-3.07*	-.16
Concept Formation	100	15	93.94	12.78	2.33*	.12
Visual Matching	100	15	94.32	14.56	2.18*	.11
Numbers Reversed	100	15	87.18	15.78	4.85***	.26
General Information	100	15	71.89	11.52	13.94***	.02
Retrieval Fluency	100	15	90.26	23.10	2.45*	.13
Picture Recognition	100	15	105.38	9.56	-3.19**	.17
Auditory Attention	100	15	108.74	14.71	-3.36*	.18
Analysis-Synthesis	100	15	98.24	12.73	.68	
Decision Speed	100	15	100.56	17.48	-.21	
Memory for Words	100	15	91.61	11.28	4.35***	.22

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$

Question 1C

What is the difference in performance difference between the normative group ($\bar{X} = 100$), NLD and ELD groups as measured by the Global Intellectual Ability (GIA) scores?

There was a significant difference between the NLD, ELD and the normative samples on the GIA score, $F(2,1444) = 12.25, p < .05$. Games Howell tests revealed significant differences between the normative group ($\bar{X} = 100.0$) and the NLD group ($\bar{X} = 90.81$), and the normative group ($\bar{X} = 100$) and the ELD group ($\bar{X} = 84.44$) with the normative sample scoring significantly higher ($\eta = .13$).

Question 1D

What is the difference in performance between the between the normative group ($\bar{X} = 100$), NLD, and ELD groups as measured by the 14 subtest mean scores?

No significant differences were found between the norm, NLD and ELD groups on the Picture Recognition, Spatial Relations, Visual Matching, Analysis-Synthesis, Concept Formation and Decision Speed subtests.

Significant differences were found in Numbers Reversed between the norm, NLD, and ELD groups, $F(2, 1444) = 12.23, p \leq .05$. Tukey Post hoc tests revealed

that the normative sample ($\bar{X} = 100$) scored significantly higher than the NLD ($\bar{X} = 89.81$) and the ELD ($\bar{X} = 84.71$) groups ($\eta = .13$).

Significant differences were found in Retrieval Fluency between the norm, NLD, and ELD groups, $F(2, 1444) = 12.50, p \leq .05$. Games-Howell tests revealed that the normative sample ($\bar{X} = 100$) scored significantly higher than the ELD group ($\bar{X} = 82.00$) ($\eta = .13$).

There were significant differences found in Visual Auditory Learning between the norm, NLD, and ELD groups, $F(2, 1444) = 14.99, p \leq .05$. Tukey post hoc tests revealed that the normative sample ($\bar{X} = 100$) scored significantly higher than the NLD group ($\bar{X} = 89.75$) and the ELD group ($\bar{X} = 82.89$) ($\eta = .14$).

There were significant differences found in Auditory Attention between the norm, NLD, and ELD groups, $F(2, 1444) = 8.86, p \leq .05$. Tukey post hoc tests revealed that the NLD group ($\bar{X} = 115.63$) scored significantly higher than the normative sample ($\bar{X} = 100$) and the ELD group ($\bar{X} = 102.61$) ($\eta = .11$).

There were significant differences found in Memory for Words between the norm, NLD, and ELD groups, $F(2, 1444) = 5.35, p \leq .05$. Games-Howell tests revealed that the normative sample ($\bar{X} = 100$), scored significantly higher than the NLD group ($\bar{X} = 90.31$) ($\eta = .08$).

There were significant differences found in Sound Blending between the norm, NLD, and ELD groups, $F(2, 1444) = 4.77, p \leq .05$. Games-Howell tests

revealed that the normative sample ($\bar{X} = 100$), scored significantly lower than the NLD group ($\bar{X} = 108.93$) ($\eta = .08$).

There were significant differences found in General Knowledge between the norm, NLD, and ELD groups, $F(2, 1444) = 60.38, p \leq .05$. Games-Howell tests revealed that the normative sample ($\bar{X} = 100$), scored significantly higher than the NLD group ($\bar{X} = 76.44$) and the ELD group ($\bar{X} = 67.83$) ($\eta = .28$).

There were significant differences found in Verbal Comprehension between the norm, NLD, and ELD groups, $F(2, 1444) = 19.86, p \leq .05$. Tukey post hoc tests revealed that both the normative sample ($\bar{X} = 100$), and the NLD group ($\bar{X} = 93$) scored significantly higher than the ELD group ($\bar{X} = 78.61$) ($\eta = .6$). Figure 6 summarizes the above data and answers the question: What is the difference in performance between the normative sample ($\bar{X} = 100$) the NLD, and ELD group as measured by the 14 subtest mean scores?

Question 1E

What is the difference in performance between the normative sample ($\bar{X} = 100$) and NLD group as measured by the seven CHC cluster mean scores?

There were significant differences between the norm sample and the NLD group on four CHC cluster mean scores (see Figure 7). There was a significant difference between the normative sample and the NLD group, $t = -10.81, p < .05$, with the normative sample ($\bar{X} = 100$) scoring significantly higher on the Gc cluster

than the NLD group ($\bar{X} = 84.69$). A small effect size, Cohen's $d = .17$, was found (Cohen, 1988).

Figure 6. Differences in the 14 subtest scores for norm, NLD, and ELD groups.

There was a significant difference between the normative sample and the NLD group on the Glr cluster, $t = -2.35$, $p < .05$, with the normative sample ($\bar{X} = 100$) scoring significantly higher than the NLD group ($\bar{X} = 90.19$). A large effect size, Cohen's $d = .82$, was found (Cohen, 1988).

Similarly, there was a significant difference between the normative sample and the NLD group on the Gsm cluster, $t = -4.65$, $p < .05$, with the normative sample ($\bar{X} = 100$) scoring significantly higher than the NLD group ($\bar{X} = 88.25$). A small effect size, Cohen's $d = .42$, was found (Cohen, 1988). See Figure 7.

Figure 7. CHC cluster scores for normative sample and NLD group.

There was also a significant difference between the normative sample and the NLD group on the Ga cluster, $t = 2.23$, $p < .05$, with the NLD group ($\bar{X} = 110.75$) scoring significantly higher than the normative sample ($\bar{X} = 100$). A large effect size, Cohen's $d = .87$, was found (Cohen, 1988).

There were no significant difference between the normative sample and the NLD group on the Gv, Gf, or Gs clusters.

Question 1F

What is the difference in performance between the normative sample ($\bar{X} = 100$) and the ELD group as measured by the seven CHC cluster mean scores?

There were significant differences between the norm sample and the ELD group on four CHC cluster mean scores (see Figure 8). There was a significant difference between the normative sample and the ELD group, $t = -9.20$, $p < .05$, with the normative sample ($\bar{X} = 100$) scoring significantly higher on the Gc cluster than the ELD group ($\bar{X} = 72.89$). A small effect size, Cohen's $d = .22$, was found (Cohen, 1988).

Figure 8. CHC cluster scores for normative sample and ELD group.

There was a significant difference between the normative sample and the ELD group, $t = -4.34$, $p < .05$, with the normative sample ($\bar{X} = 100$) scoring significantly higher on the Glr cluster than the ELD group ($\bar{X} = 76.61$). A small effect size, Cohen's $d = .47$, was found (Cohen, 1988).

There was a significant difference between the normative sample and the ELD group, $t = -3.13$, $p < .05$, with the normative sample ($\bar{X} = 100$) scoring significantly higher on the Gsm cluster than the ELD group ($\bar{X} = 81.83$). A medium effect size, Cohen's $d = .69$, was found (Cohen, 1988).

There was a significant difference between the normative sample and the ELD group, $t = 2.15$, $p < .05$, with the ELD group ($\bar{X} = 106.56$) scoring significantly higher on the Ga cluster than normative sample ($\bar{X} = 100$). = 96). A large effect size, Cohen's $d = .96$, was found (Cohen, 1988).

There were no significant differences between the normative sample and the ELD group on the Gv, Gf, or Gs clusters (see Figure 8).

Question Two

What is the relationship between performance as measured by GIA and English language proficiency, Spanish language proficiency, and acculturation levels? The results of correlation analyses for Questions 2A – 2C are provided below. Effect sizes of $r = .1$ is small, $r = .3$ is medium and $r = .5$ is large (Cohen, 1988).

Question 2A

What is the relationship between performance as measured by the GIA scores and level of English language proficiency for the NLD and ELD groups?

There was no significant relationship between the GIA score and English language proficiency for either the NLD ($r = .12, ns$) or the ELD ($r = .16, ns$) group.

Question 2B

What is the relationship between performance as measured by the GIA scores and level of Spanish language proficiency in the NDL and ELD groups?

There was no significant relationship between the GIA scores and Spanish language proficiency for either the NLD ($r = .18, ns$) or ELD ($r = .12, ns$) group.

Question 2C

Is there a relationship between performance as measured by the GIA scores and level of acculturation in the NLD and ELD groups?

There was no significant relationship between the GIA scores and acculturation for either the NLD ($r = .24, ns$) or ELD ($r = .19, ns$) groups.

Question Three

Do the patterns of subtest mean scores achieved by general education U.S. bilingual participants on the *Batería III* vary as a function of cultural loading and linguistic demands of the subtests in the manner predicted by the WJ-III C-LIM? Figure 9 shows the predicted scores based on the WJ-III and the mean subtest

CULTURE-LANGUAGE INTERPRETIVE MATRIX (C-LIM) WORKSHEET - BATERIA III

Name of Examinee: _____ Age: _____ Grade: _____ Date: _____

DEGREE OF LINGUISTIC DEMAND

DEGREE OF CULTURAL LOADING	DEGREE OF LINGUISTIC DEMAND		
	LOW	MODERATE	HIGH
LOW	SOUND BLENDING (Ga) (107) Picture Recognition (Gv) (106) AUDITORY ATTENTION (Ga) (103) DECISION SPEED (Gs) (99) Predicted Level Average = 96 Cell Average = 104	Test Name: CONCEPT FORMATION (Gf) (94) ANALYSIS SYNTHESIS (Gf) (98) SPATIAL RELATIONS (Gv) (97) Predicted Level Average = 93 Level Average = 96	Test Name: NUMBERS REVERSED (Gsm) (85) Predicted Level Average = 91 Level Average = 89
	Test Name: MEMORY FOR WORDS (Gsm) (93) Predicted Level Average = 93 Level Average = 96	Test Name: RETREIVAL FLUENCY (Ghr) (82) VIS-AUD LEARNING (Ghr) (83) Predicted Level Average = 88 Cell Average = 83	Test Name: GENERAL KNOWLEDGE (Gc) (68) VERBAL COMPREHSN (Gc) (78) Predicted Level Average: 85 Cell Average = 73
	Test Name: _____ _____ _____ _____ 91 Level Average = _____	Test Name: _____ _____ _____ _____ 88 Level Average = _____	Test Name: _____ _____ _____ _____ 88 Level Average = _____
	Test Name: _____ _____ _____ _____ 91 Level Average = _____	Test Name: _____ _____ _____ _____ 91 Level Average = _____	Test Name: _____ _____ _____ _____ 91 Level Average = _____

Figure 9. Culture-language interpretive matrix (C-LIM) for the Bateria III.

scores achieved by the ELD group. Flanagan and Ortiz (2001), offer a range of predicted scores for individuals who are slightly, moderately, or markedly different for each of the nine cells on the C-LIM. The acculturation and language proficiencies of this research sample mostly fell within the middle ranges of both variables, therefore a predicted score from the Moderately Different range on Figure 5 was selected.

Question 3A

Do the *Batería's* 14 subtest mean scores (administered in Spanish) of the ELD group follow the expected pattern of performance on the C-LIM for diverse individuals when tested in English?

Figure 10 presents the C-LIM as a diamond. The cell containing the subtests that are the lowest in cultural loading and linguistic demand is labeled number one. The next two cells containing the subtests that have a medium degree of cultural loading and linguistic demand are labeled number two and so forth. The cell containing subtests that are high in both cultural loading and linguistic demand is labeled number 5. For each level of cells, a predicted score is provided. This predicted score was chosen based on Figure 5 in chapter 2. Although Figure 5 presents a range of scores, a single score from within the Moderately Different range was chosen to use here. Figure 10 shows the predicted scored (based on Figure 5) and the mean cluster scores for the ELD group for each scored (based on Figure 5) and the mean cluster scores for the ELD group for each cell level 1

through 5. Scores in Level 5 for the ELD group were lower for the ELD group than the predicted scores.

Predicted Scores by Levels of C-LIM

Figure 10. C-LIM predicted scores for ELL students tested in English and scores of ELD participants tested in Spanish on the *Batería III*.

Question Four

How do participants perceive the experience of native language assessment?

Question 4A

What language(s) do you think in?

Fifty percent of the participants thought predominantly in Spanish and 50% in English.

Question 4B

What language(s) do you like to use?

Sixty-five percent of the participants reported that they preferred to speak in Spanish, 5 % English, and 30 % in both languages.,

Question 4C

Describe what was easy about the testing?

Forty percent of the participants reported that tasks requiring them to find pairs were easy, 15% reported responding to taped items easy, 15% found reading tasks in Spanish easiest, 10% found tasks requiring them to respond by pointing easy while the remaining 10% reported naming items in Spanish easy

Question 4D

Describe what was hard about the testing?

Forty-five percent of the participants reported the task requiring them to discriminate words given from a tape recording under conditions of increasing distortion and pointing to the corresponding picture from a page of pictures was hard. Eleven percent reported puzzle tasks, remember words and repeating them, and repeating numbers in reverse to be hard. Twelve percent reported maintaining concentration was difficult.

CHAPTER V

SUMMARY

Discussion

The primary purpose of this study was to examine how U.S. bilingual students, commonly known as ELL students, performed on a Spanish language cognitive test, the *Batería III*. Unlike testing accomplished using English, empirical research on the use of Spanish cognitive tests with ELLs residing in the U.S. is lacking in the literature. Nonetheless, application and use of a Spanish language cognitive test is a common component of evaluations conducted on ELLs for the purpose of identifying disabilities and determining eligibility for special education programs and services (Ochoa, Rivera & Ford, 1997). However, without knowledge on how ELLs actually perform on tests administered in their native language (i.e., Spanish), there is no clear guidance regarding whether the assessment is appropriate, much less valid. It is reasonable to believe that there may well be an adverse impact on decisions regarding ELL students evaluated using native language testing that has yet to demonstrate empirically-based validity. Conversely, if the use of native language cognitive assessments could be shown to be scientifically valid and appropriate, it may be part of the solution to the misidentification of ELL students for special education.

Native Language Standardized Tests

The issue of linguistic and cultural differences of ELL students and standardized assessments are not solved when native-language tests are used because ELL students living in the United States are not appropriately reflected in the normative population. Because of the great variation in acculturation levels and language proficiencies in English and native-language, it is highly unlikely that ELL students can ever be appropriately reflected in any normative sample even when the instruments are developed in the U.S. (Flanagan & Ortiz, 2001). The assumptions underlying standardized tests are that all test-takers have similar experiences, exposure to curriculum, general motivation toward test-taking, and language learning opportunities (Mercer, 1979). Therefore, while tests may have acceptable psychometric properties, current instruments are not normed on individuals like U.S. ELL students who live within two language systems (Valdes & Figueroa, 1994) and usually within two underdeveloped language systems.

Characteristics of the Research and Normative Sample

One of the basic tenets of standardized assessment is that the scores generated provide information on an individual's standing on the measured construct in comparison to a normative reference (Gay et al., 2006). To the extent that the research sample is represented in the normative sample, the scores will be a gauge of the individual's standing on the construct (Graham, Naglieri, & Weiner, 2003). Several factors, however, must be considered in determining whether a

normative sample represents the target population. In this case, the research and normative samples differ in English proficiency, Spanish proficiency, and acculturation levels; significant factors that can impact performance.

Figures 11 and 12 provide a visual representation of the varying characteristics between the normative sample and the research sample. First, the research sample's bilingual world manifests in varying proficiency in both English and Spanish while the monolingual Spanish norm sample had access to a complete native language environment and opportunity to develop full proficiency. The research sample exhibits further duality in their lives, that is, acculturating to American culture. The participants walk in two cultural worlds. One is their heritage culture and the second, mainstream American culture. The majority of the normative sample, however, reside within their native countries (Woodcock et al., 2004) and most likely maintain the mainstream culture. In addition to the differences shown above, the research sample had two distinctly different populations. One group ($n = 16$) received native language literacy development (NLD) in a late-exit model while the second group ($n = 18$) received English language development (ELD) as a content with no instruction in native language. The normative sample most likely received Spanish instruction because of their monolingual Spanish status. Clearly, differences in these significant characteristics will impact measured test performance. The following sections review and discuss the differences found in this study.

Figure 11. Five characteristics of U.S. bilingual that impact test performance.

Figure 12. Three characteristics of normative sample that impact test performance.

Discussion of Study Findings

This study examined how U.S. bilingual students perform on the *Batería III*. Performance was evaluated on several indices, including overall cognitive functioning (GIA or IQ), functioning at the broad ability level (the cognitive ability and processing clusters that underlie general intelligence), and functioning at the narrow ability level (the specific cognitive ability or processes measured by a single subtest or task). Thus, it was possible to identify those tests or composite scores that may be inaccurate measures of the actual cognitive abilities of ELL children.

In general, the results from this study indicated that the bilingual research population (NLD and ELD groups) scored lower on three broad ability factors (*Gc*, *Glr*, *Gsm*) than the normative comparison and higher on one factor (*Ga*). Group subtest scores were examined using Flangan and Ortiz's (2001) C-LTC and C-LIM. The findings show that the patterns previously observed on the *Batería III*'s English parallel test, the WJ-III (Flanagan et al., 2007) were not entirely similar. For participants, performance differences appeared to be a function of whether ELLs were receiving instruction that fostered native language development (NLD) or those receiving instruction in an English language development (ELD) program only.

In the final analysis, while the *Batería III* may not be a perfectly accurate gauge of ELL students' abilities, this assessment, along with others in the

Woodcock-Johnson family of tests, is held in high esteem in the field and is commonly used. Thus, it is imperative to understand how the *Batería III* may be used in an appropriate manner with U.S. ELL students (a non-homogenous group) and what can be done to ensure that interpretation of results is made in as nondiscriminatory a manner as possible. The performance patterns and implications are discussed in the following section. This chapter concludes with a discussion regarding the limitations of the study and suggestions for future research.

Discussion of the Results

To review, the theoretical base of the *Batería III* is the Cattell-Horn-Carroll Theory of Intelligence (CHC). The *Batería III* scoring program generates scores for seven broad CHC abilities, subtests, and a GIA score. While the results of cognitive tests for any individual will show patterns of strengths and weaknesses, to determine the validity of these patterns assessors must examine them within the context of possible extrinsic reasons (e.g., lack of motivation, cultural and linguistic differences) for resulting patterns before searching for intrinsic ones (inherent weaknesses). The relevant extrinsic factors within this study are language proficiency in L1 and L2, language of instruction, and cultural differences.

To help illustrate the significance of language input and language of instruction, the data were analyzed for the total bilingual group (NLD + ELD), and the NLD and ELD groups separately. As will be seen, however, the NLD and ELD groups are two separate populations that score somewhat differently. For ease of

understanding the results and their implications, each section begins with the relevant research questions. Specific implications are discussed here with the broader implications discussed later.

Analysis of Functioning in General Cognitive Ability

Research Question 1: What is the difference in overall performance as measured by the Global Intellectual Ability (GIA), subtest, and cluster scores between the monolingual Spanish normative group and the U.S. bilingual comparison group?

Question 1.a. What is the difference in performance between the normative group and the total bilingual group (NLD + ELD) as measured by the GIA scores?

The *Batería*'s GIA score is synonymous to the more commonly known term Intelligence Quotient (IQ). It describes how individuals score on a test that measures their general or global cognitive abilities as compared to the general population, to the extent that the populations can be compared. The common metric is a standardized scale with 100 as the mean and a standard deviation 15. Thus, on the *Batería III*, scores between 85 and 115 are within the average range. In this study, there were differences in the GIA scores between the normative sample scores ($\bar{X} = 100$) and the total bilingual sample ($\bar{X} = 87.4$) of 12.6 points. The differences in the mean scores should not be interpreted as the bilingual sample

having lower innate cognitive abilities, but rather that the GIA scores are not appropriate to report for U.S. bilinguals.

Question 1.c. What is the difference in performance between the normative group, the NLD, and ELD groups as measured by the GIA scores?

When the bilingual group was separated by language of instruction, both the ELD ($\bar{X} = 84.44$) and the NLD ($\bar{X} = 90.81$) group scored significantly lower on the GIA score than the norm group. This concurs with earlier research on racial and ethnic minority groups when tested in English (Sanchez, 1934; Yerkes, 1921) showing that racial or ethnic groups usually perform at a range between three-quarters to one standard deviation lower than the mainstream population. In this research, bilingual individuals scored similarly on a Spanish language cognitive test when tested on English cognitive tests. On closer examination, the ELD group scored lower than the NLD group. This is likely a reflection of how native language instruction has aided the NDL group's linguistic and conceptual development.

The results of the GIA score differences are not surprising since the *Batería III*'s normative group is not representative of the levels of bilingualism, cultural experiences, and differences in education and acculturation that were present in the research group. The bilingual research sample demonstrated moderate proficiency in English and Spanish but did not include fluent Spanish speakers (except for one student). This is in contrast to the normative sample who were monolingual Spanish speakers and presumably demonstrated fluency in that language.

Another way to interpret this data might be to conclude that the research sample had lower cognitive abilities than Spanish speakers their age. This would be an erroneous assumption, however, because the language profile of the ELL research sample in this study is different from that of the normative sample, therefore, such a comparison is invalid.

The biasing effect from the use of psychometric instruments, therefore, operates whenever tests of intelligence and cognitive ability (developed and normed in the United States) are given to individuals whose cultural backgrounds, experiences, and exposures are not similar to or consistent with those of the individuals comprising the norm group against whom performance will be compared. In these cases, such tests will likely measure a lower range of ability in diverse individuals because the test samples only the cultural content related to mainstream experience and not the full or entire range of cultural content possessed by the individual. (Flanagan & Ortiz, 2001, p. 220)

Analysis of Functioning in Broad Cognitive Abilities and Processes

The following four research questions related to test performance on the seven CHC broad ability clusters and the 14 subtest scores. Following the questions, each broad ability, and the corresponding subtests are described when significant performance differences were found. The following summarizes the significant differences. Implications relative to that ability are then provided.

- 1.b. What is the difference in performance between the normative group and the total bilingual group as measured by the 14 subtest mean scores?
- 1.d. What is the difference in performance between the normative group, the NLD, and ELD groups as measured by the 14 subtest mean scores?

1.e. What is the difference in performance between the normative group and the NLD group as measured by the CHC cluster mean scores?

1.f. What is the difference in performance between the normative group and the ELD group as measured by the CHC cluster mean scores?

Although analyses were conducted to answer the four questions above, best psychometric practice guides assessors to interpret performance at the cluster level, so the cluster descriptions are provided first. Flanagan et al. (2007) reinforced this concept of clusters as two subtests of a construct while discussing the mixing of subtests across test batteries known as the Cross Battery Approach.

In general, the cross battery approach is based on a hierarchical model of interpretation, which emphasizes interpretations of broad ability/processing constructs (e.g., *Gf*) over narrow ability/processing constructs (e.g., Induction, General Sequential Reasoning) because they are typically more reliable and valid. That is, broad abilities/processes are represented by at least two qualitatively different indicators (subtests) of the construct, whereas narrow abilities/processes are typically represented by a single subtest. (p. 82)

The following shows the performance of each group on the cluster and subtest scores.

Comprehension Knowledge (Gc)

Gc ability, defined by Schrank and Flanagan (2003), includes the narrow abilities (a) language development (general development of native spoken language skills), (b) lexical knowledge (extent of vocabulary that can be understood in terms of correct word meanings, and (c) general information (range of general knowledge of a culture) (p. 6). Table 19 shows the significant differences on *Gc* by group.

Table 19

Significant Differences by Group on Gc Cluster and Subtests

Clusters	Subtests	Normative Group	Bilingual Group	NLD Group	ELD Group
Comprehension Knowledge (Gc)		$\bar{X} = 100$	No data	Lower	Lower
	Verbal Comprehension	$\bar{X} = 100$	Lower	Higher	Lower
	General Information	$\bar{X} = 100$	Lower	Lower	Lower

As evident from the table, both the ELD and NLD groups scored lower on the Gc cluster than the normative sample. The bilingual, ELD, and NLD groups scored lower on the General Information subtest than the normative sample. The bilingual and ELD groups scored lower on Verbal Comprehension but the NLD group scored higher than the normative sample.

In general, most research groups scored lower than the norm group on this broad factor (Gc). This ability is highly culturally loaded and is based on general life experiences that accumulate over time with informal adult interaction or mediation. Further, as discussed earlier, the U.S. bilingual research sample is a product of two language systems. Only one participant in the total research sample measured as a Fluent Spanish Speaker on the LAS. Thus, it stands to reason students who have not fully developed Spanish language proficiency will have difficulty on this cluster given the focus on language development as well as vocabulary (Cummins, 1984).

The General Information subtest asks “Where” (e.g., “In general, where would you find a caboose?”) and “What” (e.g., “In general, what do you do with a koto?”) questions that assume a broad range of mainstream life experiences in the country where the test was developed. ELL students, however, in the United States, some of who are immigrants, may not have had such experiences. Certainly their experiences differed from those of the normative population.

The third narrow ability forming the *Gc* broad ability is knowledge of the culture the assessment represents, in this case, mainstream American culture as the *Batería* was developed for use in the U.S. As indicated by the acculturation scores, only one student in the bilingual research sample was “highly acculturated” as measured by the AQS. Although the validity of the AQS was not available, for these purposes, the scores provide a general gauge of the research sample’s acculturation. Thus, since it appears the research sample is in the process of learning about mainstream culture or becoming acculturated, this aspect of *Gc* may be problematic for them (Cummins, 1984; Figueroa, 1990; Matsumoto, 1994; Valdes & Figueroa, 1994).

The NLD group scored higher than the norm and ELD groups on Verbal Comprehension, which assessed knowledge of vocabulary, synonyms, antonyms and analogies. The NLD group received the benefit of continuing native and English language development and this bilingual status may have helped their performance on this subtest.

Implications. This cluster is likely not a valid measure of ELL students' abilities. Rather it appears to highlight acculturation and language differences between the normative sample and bilingual research groups. All research groups scored the lowest on this cluster. Questions required a broad knowledge experience acquired by exposure to middle class, mainstream concepts. In addition, the questions included many with academic vocabulary that children may not have been exposed to, even in a native language curriculum at the third and fourth grade.

Long-Term Retrieval (Glr)

Glr, defined by Schrank and Flanagan (2003), includes the narrow abilities of (a) associative memory (i.e., ability to recall one part of an unrelated pair of items when the other part is presented), (b) meaningful memory (i.e., ability to recall items that are meaningfully related), and (c) ideational fluency (i.e., ability to rapidly produce items from a specified category). Table 20 shows the significant differences by group.

Table 20

Significant Differences by Group on Glr Cluster and Subtests

Clusters	Subtests	Normative Group	Bilingual Group	NLD Group	ELD Group
Long-Term Retrieval (<i>Glr</i>)		$\bar{X} = 100$	<i>No data</i>	Lower	Lower
	Visual-Auditory Learning	$\bar{X} = 100$	Lower	Lower	Lower
	Retrieval Fluency	$\bar{X} = 100$	Lower	<i>Not significant</i>	Lower

Both the ELD and NLD group scored lower on the Long-Term Retrieval (Glr) cluster. The bilingual, NLD, and ELD groups scored lower on the Visual-Auditory Learning subtest. Visual-Auditory Learning required students to learn the labels for rebuses and then asked them to read sentences consisting of the rebuses. Many of the labels, although in the participants' L1, were uncommon terms, particularly for bilingual students from a working class, rural community. These students likely lack exposure to the middle class discourse (Macedo et al., 2003) assumed by the test developers. This may have made it difficult for participants to retain the given label in their working memory.

On the Retrieval Fluency subtest, only the bilingual and ELD groups scored lower. On this subtest, participants were asked to name as many items as possible from a given category. Although the categories appeared general, many participants scored lowest on one category where broader world experiences may have helped them, although though they were allowed to respond in any language on this subtest (Cummins, 1984; Valdes & Figueroa, 1994). Thus, the low scores could indicate differing life experiences rather than memory difficulties. Similarly to *Gc* the NLD group's native language instruction appears to have benefited them to some degree since they scored in the same range as the normative sample on Retrieval Fluency.

Implications. The *Glr* cluster appears to be a poor measure of ELL participants' abilities in long term retrieval. While on the surface, the Visual-

Auditory Learning subtest, similar to decoding, seems to have reduced language demands, many of the vocabulary terms used were unfamiliar to the bilingual research sample. Also, one of the categories in Retrieval Fluency required a broader knowledge base than the research sample appeared to have.

Auditory Perception (Ga)

Ga ability, defined by Schank and Flanagan (2003), consists of the narrow abilities (a) phonetic coding (i.e., ability to blend small units of speech sounds into larger units, (b) speech-sound discrimination (i.e., ability to discriminate between different speech sounds, and (c) resistance to auditory stimulus distortion (i.e., ability to comprehend speech sounds that are masked by non-speech sounds). Table 21 shows the significant differences by research group.

Table 21

Significant Differences by Group on Ga Cluster and Subtests

Cluster	Subtests	Normative Group	Bilingual Group	NLD Group	ELD Group
Auditory Processing (<i>Ga</i>)		$\bar{X} = 100$	<i>No data</i>	Higher	Higher
	Sound Blending	$\bar{X} = 100$	Higher	Higher	<i>Not significant</i>
	Auditory Attention	$\bar{X} = 100$	Higher	Higher	<i>Not significant</i>

The NLD and ELD groups scored higher than the normative sample on the Auditory Processing (*Ga*) cluster. There were no differences found for the ELD

group on either subtest. The bilingual and NLD groups scored higher on both subtests, Sound Blending and Auditory Attention. One likely reason that the NLD group scored higher throughout this cluster is that because participants live within two linguistic worlds. Thus, their auditory processing skills may be more highly attuned than that of either the monolingual Spanish normative sample or the ELD group (Cummins, 1984). Also, the NLD group participants may be more used to blocking out auditory information since they use more Spanish in their daily lives than the other groups yet are continuously bombarded by the English language. A third reason may be the new curricular focus on phonemic awareness and training in sound blending may have aided the research sample. Thus, their higher scores also impacted the score of the total bilingual group boosting both subtest scores.

Implications. Participants appear to score higher on the *Ga* cluster and subtests. Being a product of two linguistic worlds appears to help them develop strong auditory processing skills. Additionally, students are now taught discrete skills (e.g., phonemic awareness) from kindergarten so they are familiar with perceiving and blending discrete sounds. Plus, since there are only 24 phonemes in Spanish, 19 of which are consonants and five vowels as compared to the 40 phonemes in English (Delfior & Serrano, 2005) recognition of sounds may be easier in Spanish.

Short-Term Memory (Gsm)

Gsm ability, defined by Schrank and Flanagan (2003), includes the narrow abilities of (a) working memory (i.e., ability to temporarily hold in mind and

mentally manipulate phonological stimuli to produce a response), and (b) memory span (i.e., the ability to attend to and immediately recall a series of phonological stimuli in their correct order). Table 22 shows the significant differences by group.

Table 22

Significant Differences by Group on Gsm Cluster and Subtests

Cluster	Subtests	Normative Group	Bilingual Group	NLD Group	ELD Group
Short-Term Memory (<i>Gsm</i>)		$\bar{X} = 100$	<i>No data</i>	Lower	Lower
	Numbers Reversed	$\bar{X} = 100$	Lower	Lower	Lower
	Memory for Words	$\bar{X} = 100$	Lower	Lower	<i>Not significant</i>

All groups scored lower on the cluster score as well as subtest scores, except for the ELD group who scored similarly to the normative mean score ($\bar{X} = 100$) on Memory for Words. All groups had difficulty with Numbers Reversed. The most likely reason is that they were less familiar with numbers in Spanish since both the NLD and ELD groups receive math instruction in English.

Memory for Words asked participants to repeat lists of unrelated words. Again, the vast majority of the sample was still developing Spanish proficiency even though it was their native language. Thus, they may have struggled retaining unfamiliar vocabulary. They also may have limited exposure and models to

academic Spanish since they live within an English-speaking country (Sandoval, Zimmerman, & Woo, 1980).

Implications. The *Gsm* cluster and subtests appeared to be difficult for the participants. The task required participants to retain chunks of information efficiently into their short-term memory. They may have had difficulty because they were not very familiar with some of the vocabulary. They also have not had math instruction in Spanish. Thus, there may have been interference if they had to translate the numbers in their head to create a visual representation and then state the numbers in reverse order. The crucial point here is that assessors must know the curricular areas in which students receive native language or English instruction as it may help explain why the students score as they did.

Clusters with Significant Differences in Only One Group

Fluid Reasoning (Gf)

Gf ability, defined by Schrank and Flanagan (2003), includes the narrow abilities of (a) induction (i.e., ability to identify the concept or rule that underlies a problem or set of stimuli, and (b) general sequential reasoning (i.e., ability to start with stated rules and engage in steps to reach a solution to a novel problem. Table 23 shows the significant differences by group.

Table 23

Significant Differences by Group on Gf Cluster and Subtests

Cluster	Subtest	Normative Group	Bilingual Group	NLD Group	ELD Group
Fluid Reasoning (Gf)		$\bar{X} = 100$	<i>No data</i>	<i>Not significant</i>	<i>Not significant</i>
	Concept Formation	$\bar{X} = 100$	<i>Lower</i>	<i>Not significant</i>	<i>Not significant</i>
	Analysis-Synthesis	$\bar{X} = 100$	<i>Not significant</i>	<i>Not significant</i>	<i>Not significant</i>

The bilingual group (NLD + ELD) was the only group to score lower in the area of Fluid Reasoning (Gf) on the subtest Concept Formation. This is a controlled learning task requiring examinees to identify and state rules governing a set of colored geometric figures. There is a strong oral language component to this subtest. Participants had to comprehend lengthy oral directions, process them, follow the instructions and give an oral response. The NLD and ELD groups may not have demonstrated a significant difference because dividing the bilingual group into two smaller groups. The total numbers in each of the smaller groups may have been too small to show a significant difference.

Implications. This cluster is linguistically demanding and depressed scores of minority groups in this area have been found in past research (Mercer, 1979; Sandoval, 1979; Valdes & Figueroa, 1994). Yet, in general this cluster appeared to be a fair measure of Gf for all groups.

Processing Speed (Gs)

Gs ability, defined by Schrank and Flanagan (2003), includes the constructs of (a) perceptual speed (i.e., ability to rapidly search for and compare visual symbols or patterns and make a simple response, and (b) semantic processing speed (i.e., ability to rapidly identify basic conceptual relationships among stimuli. Table 24 shows the significant differences by group.

Table 24

Significant Differences by Group on Gs Cluster and Subtests

Cluster	Subtest	Normative Group	Bilingual Group	NLD Group	ELD Group
Processing Speed (Gs)		$\bar{X} = 100$	<i>No data</i>	Lower	<i>Not significant</i>
	Visual Matching	$\bar{X} = 100$	Lower	<i>Not significant</i>	<i>Not significant</i>
	Decision Speed	$\bar{X} = 100$	<i>Not significant</i>	<i>Not significant</i>	<i>Not significant</i>

Only the NLD group scored lower than the other groups on the *Gs* cluster. The bilingual group scored lower on the Visual Matching subtest. Visual Matching was a timed task in which participants had to rapidly locate and circle numbers from a given row of numbers. The subtest does not have high linguistic demand since it involves numbers. Although the numbers were on paper and not given orally, some participants may have still been internally translating the numbers.

Implications. In general, the *Gs* cluster and subtests do not have high linguistic demands nor require a high degree of cultural knowledge so may be fair measures of ELL students' abilities in this area (Flanagan & Ortiz, 2001).

Processing speed is the ability to perform mental tasks fluently and automatically, particularly under a timed condition. Some participants, however, may have tried to read the numbers in Spanish as they were scanning the row and lost time thus causing a lower score.

Visual Perception (Gv)

Gv, defined by Schrank and Flanagan (2003), includes the narrow abilities (a) spatial relations (i.e., ability to rapidly construct specified visual patterns or to understand how visual stimuli relate to each other, (b) visualization (i.e., ability to hold visual stimuli and spatial forms in mind and to alter them in some way, and (c) visual memory (i.e., ability to retain representations of visual stimuli in the mind and to recognize or recall them soon afterward). Table 25 shows the significant differences by group.

Table 25

Significant Differences by Group on Gv Cluster and Subtests

Cluster	Subtests	Normative Group	Bilingual Group	NLD Group	ELD Group
Visual Perception (<i>Gv</i>)		$\bar{X} = 100$	<i>No data</i>	<i>Not significant</i>	<i>Not significant</i>
	Spatial Relations	$\bar{X} = 100$	<i>Not significant</i>	<i>Not significant</i>	<i>Not significant</i>
	Picture Recognition	$\bar{X} = 100$	Higher	<i>Not significant</i>	<i>Not significant</i>

The bilingual group scored higher on the Picture Recognition subtest. Picture Recognition required participants to study pictures for five seconds and then identify them within a larger array of pictures. Spatial Relations, a type of puzzle, asked participants to select parts of whole shapes. Neither subtest had extraordinary linguistic demands or cultural implications. Accordingly, there were no differences between any other groups.

Implications. It appears that the *Gv* cluster and subtests may be a fair measure of ELL students' abilities on this factor. Participants were strong in reversing and rotating shapes visually to complete puzzles and attended well to both subtests.

Relationships Between Variables

This section addresses relationships between the study's variables and discusses Questions 2 through 2c.

Question 2: What is the relationship between performance as measured by the GIA and English language proficiency, Spanish language proficiency, and acculturation level?

2.a. What is the relationship between performance as measured by the GIA scores and level of English language proficiency in the NLD and ELD groups?

2.b. What is the relationship between performance as measured by the GIA scores and level of Spanish language proficiency in the NLD and ELD groups?

2.c. What is the relationship between performance as measured by the GIA scores and level of acculturation in the NLD and ELD groups?

In this study, no significant relationships were found between GIA scores, English language proficiency, Spanish language proficiency, or acculturation in either the NLD or ELD groups. These results are puzzling because it would be reasonable to conclude that greater fluency in Spanish may result in higher performance on a Spanish language assessment. However, this was not the case. One reason may be the relatively small research sample ($N = 34$) and the language proficiency scores of the bilingual participant groups did not contain much variability. The scores of both the NLD and ELD groups were mostly between LAS 3 and LAS 4.

Implications

Since there were no relationships between language proficiency or acculturation to GIA scores, the results here may suggest that this test is not just a measure of an examinee's proficiency in Spanish but measures the actual constructs it purports to measure. Yet, the GIA, cluster, or subtest scores *cannot* be reported as true measures of an ELL student's abilities but must be interpreted with an understanding of how this population performs on the *Batería III* as compared to the monolingual normative sample. Using the framework developed by Flanagan and Ortiz (2001), described below, is one possible method of appropriate test interpretation.

Using the C-LIM and the *Batería III*

Question 3: Do the patterns of subtest mean scores achieved by general education U.S. bilingual participants on the *Batería III* vary as a function of cultural loading and linguistic demands of the subtests on the C-LIM in the manner predicted by WJ-III Culture-Language Test Classification?

The Culture-Language Interpretive Matrix

The third area examined was the pattern of performance on subtest mean scores achieved by the ELD group as compared to the C-LIM's predicted pattern of performance for ELL students on English cognitive tests. For this study, the pattern of performance was analyzed only for the ELD group because this is the population (i.e., ELL students receiving English instruction) for which the C-LIM was developed (Flanagan & Ortiz, 2001). The subtest scores from the ELD group's C-LIM were averaged and plotted to compare the scores to the predicted scores for the WJ III subtests (see Figure 12). There appears to be a similarity in performance patterns between the ELD group and the predicted pattern of performance for diverse individuals when the English WJ III cognitive test was administered.

To most clearly illustrate the similarities and differences based on the theoretical framework, the C-LIM (see Figure 10) is presented here as a diamond to illustrate the interpretive levels (levels 1-5). The appropriate use of the C-LIM is revisited here. The C-LIM takes the obtained standard scores of a cognitive test battery or subtests from more than one test battery (i.e., the cross-battery approach)

and calculates an average for all subtest scores that are classified for each of the nine cells. Each cell contains subtests which have been classified according to the level of cultural loading and linguistic demand (Flanagan & Ortiz, 2001). Each of the subtests have been classified by Flanagan and Ortiz (2001) and found in their C-LTC. Each subtest has been classified to fit into one of the cells in the matrix. The matrix is arranged from low culture/low linguistic (the uppermost left hand cell) to high culture/high language (the lowest right hand cell). When testing in English, individuals who are bilingual, research has consistently demonstrated that their performance will be highest in the low/low cells and lowest in the high/high cells (Mercer, 1979; Sanchez, 1934; Yerkes, 1921). The greatest decline has been quantified to be up to one standard deviation (15 points). Thus, performance is closest to average ($SS = 100$) in the low/low cell, and the lowest in the high/high cell ($SS = 85$). Since the matrix was developed to account for the influence of cultural and linguistic differences, when scores are entered, it is possible to evaluate the results to determine whether the results were more influenced by cultural factors (i.e., level of acculturation) and linguistic factors (i.e., level of English or Spanish language proficiency) or actual intrinsic ability.

When an individual's scores decline systematically from the top left cell to the bottom right cell, this indicates that the primary influence on test performance should be attributed to cultural and linguistic difference more so than to actual

ability level. This is the *typical* pattern of declining performance found in bilingual individuals who do not have a disability.

In contrast, if the pattern of scores does not follow this systematic decline, then it is likely that cultural or linguistic factors were *not* the primary influences on test performance. Thus, performance may be ascribed to intrinsic factors rather than extrinsic ones. When potential confounding factors (e.g., lack of motivation, fatigue, incorrect scoring) have been ruled out. Then, it can be concluded that the data are valid and may be interpreted as one would for any other individual. This means that scores within a deficient range could support the possibility of a learning disability. Flanagan and Ortiz (2001) remind assessors, however, that information from the C-LIM is only one data source and must be supported by additional data gathered through a thorough developmental history, observations, authentic assessments, and input from family.

It is important to note that the C-LIM does not alter the scores in any way but rather allows for a systematic review of the data to determine whether the scores are valid (and interpretable), or not valid (because they are more a reflection of cultural and linguistic influences than ability). To analyze the scores from this study, both the predicted scores (Flanagan & Ortiz, 2001) and the mean level scores for the ELD groups are shown in Figure 8. In levels 1, 2, and 3, the ELD group's scores are very close to the predicted scores. In levels 4 and 5, however, the scores differ from the predicted scores. The ELD group's mean score is higher for level 4

than the predicted scores. In level 5, the ELD group's average scores were substantially lower than the predicted score.

Implications

Based on the above scores, there may be a better arrangement for the *Batería's* subtests. Figure 10 illustrates suggested changes. These changes are suggestions based on the data from this study and must be interpreted with caution and validated by other research. It is presented here as a potential starting point to establish a framework for interpreting scores from the *Batería* when used with ELL students

Participants' Perceptions

Question 4: How do participants perceive the experience of native language assessment?

4.a. What language do you think in?

Language is said to be “a medium, a mediator, and a tool of thought” (Nelson, 1996, p. 21). Language acquisition and conceptual development are intricately woven processes. Because of this symbiotic relationship, participants were asked to reflect on the process by answering four questions. Evidence is strong that cross-language facilitation in language acquisition occurs from a solid first language base but a basic level of competency is required before this process is cognitively beneficial (Baker, 2003; Bialystok, 2001). Based on the Spanish

language proficiency scores of the participants, 67.6% of the NLD group and 61.1% of the ELD group demonstrated moderate to advanced Spanish proficiency.

When asked what language they think in, 50% of the participants reported that they thought in Spanish. Knowing what language children think in helps us understand their connection to the culture represented by that language (Nelson, 1996). The participants represented an equal number of students who had greater connections to either Spanish or English and their representative cultures. This information supports the thesis that the participants are not represented in the monolingual (and likely monocultural) norm sample of the *Batería III*.

4.b. What language do you like to use?

In response, 65% of the participants reported preferring to use Spanish. Five percent reported they preferred English and 30% liked to use both. Language preference is often considered an individual's dominant language. Thus, although more participants in the sample were instructed in English only, some of those students remained Spanish dominant. This implies that building on students' dominant and preferred language may ultimately aide their academic achievement since research clearly demonstrates fluency and literacy in a first language forms the basis of academic achievement in the second language (Bialystok, 2001; Cummins, 1984; Thomas & Collier, 1997, 2002).

4.c. Describe what was easy about the testing.

4.d. Describe what was hard about the testing.

Forty percent of the participants reported that finding pairs, a visual task with no auditory input, was the easiest. The most common response as to the hardest aspect of testing was a listening task with background noise requiring sound discrimination. Although participants live within two linguistic worlds and are continually discriminating between two languages, this auditory task still required significant concentration. Yet, all groups scored above 100 (i.e., mean) on this subtest. Therefore, it is important to understand students' perceptions on language use.

Implications

Revisiting the Mandates of IDEIA 2004

While recognizing the critical role language plays in all aspects of life, it is untenable to disregard all the language systems that children possess, particularly during assessment processes. Returning to the earlier topic of RTI, it remains to be seen exactly how future models will incorporate standardized assessment. The present study found that both substantial as well as subtle information can be acquired through the use of native language tests. At the final RTI tier, whether in a three or four tier model, standardized assessments (in both languages) provide valuable information as to why some students have not responded to interventions in the previous tiers.

Policy Implications

The general summary that can be made from this study is that the NLD group tended to score higher than the ELD group on all factors. Thus, native language instruction appeared to benefit performance. Results from this study and previous research (Crawford, 2004; Thomas & Collier, 1997, 2002; Tse, 2001) illustrate the benefits of native language instruction. Yet, the English-only movement is rapidly spreading across the country. This, despite the fact that data from California, Arizona and Massachusetts indicate ELL students are falling further behind in academic areas and dropping out of school at higher rates than ever (Crawford, 2004). Yet, our society needs a highly educated populace to compete today and tomorrow in the global market.

The second policy implication concerns nondiscriminatory assessment. IDEIA (2004) has clear language about using testing and evaluation materials and procedures that are not racially or culturally discriminatory and the directive of using the native language when feasible. However, there are few guidelines from the federal or state level on how to conduct such assessments. Clearly, as the demographics of the country change, policymakers must establish clear guidelines based on empirical data from this and other similar studies.

Finally, the new RTI paradigm holds every educator accountable for the achievement of all students. This change could provide impetus for all service providers from general education, special education, gifted education and English

Language Learner educators to collaborate and thus expand everyone's professional knowledge base.

Implications for Practice

This research has implications for practice. First, the initial stages of the evaluation process must include a thorough developmental history, review of records, information from the family, and observations across settings. Assessors must have knowledge of the child's first and second language proficiency, and acculturation level. The instructional setting should be examined to determine the fit between the child's language proficiency in the instructional language and the linguistic level the instruction demands. Unless the child scores as a Fluent English Speaker on the basis of a valid language proficiency test, assessments in both languages are strongly recommended.

Second, when interpreting cluster and subtest scores from the *Batería III*, assessors should consider the results of this study. That is, U.S. bilingual/ELL students may score lower on the GIA score, the Comprehension Knowledge (*Gc*), Long-term Retrieval (*Glr*), and Short-term Memory (*Gsm*) clusters and their subtests, and higher on the Auditory Perception (*Ga*) cluster and subtests relative to the normative population in which the average score is 100. Thus, scores of between three quarters to one standard deviation ($SD = 15$) may indicate average performance for this population. Consequently, standardized scores should not be

reported. Instead, the patterns of strengths and weaknesses in these abilities should be reported instead.

Third, it is recommended that Flanagan and Ortiz's (2001) C-LTC and C-LIM framework be used for interpreting English language cognitive assessments when assessing ELL students. The revised C-LIM for the *Batería III* suggested here could be used to examine the rearrangement and resulting patterns. If both English and Spanish cognitive assessments were conducted, the C-LIM and revised C-LIM should be used to aide in the interpretation of the WJ-III and the *Batería III*.

Completely parallel assessments, however, are seldom appropriate and practitioners should conform to best practices by structuring an assessment plan that responds to the initial referral questions and allows for the bilingual child to demonstrate their abilities in both languages [e.g., cross-battery approach, Flanagan and Ortiz (2001)]. Since the ultimate goal of assessment and certainly the RTI process is to plan appropriate and effective educational interventions for struggling students, readers are referred to Table 2 to empirical links between the CHC factors and academic achievement.

Limitations

This study examined the performance of ELL participants on a Spanish language cognitive assessment, an area with little previous research. While the topic is important, the small research sample ($N = 34$) makes it difficult to generalize the results of this study. There was only one school district (rural) that

agreed to allow the research, thus the sample was not equally distributed across geographic areas. Additionally, the sample was not equally represented by gender, grade levels, language(s) of instruction, language proficiency in L1 and L2, and acculturation. Socioeconomic status was not considered. Differences in language proficiencies and acculturation levels alone can cause performance differences between students (Flanagan & Ortiz, 2001; Valdes & Figureoa, 1994).

Participants' bilingual instructional models were native language literacy instruction and English language development in a push in/pull out model. To obtain a more complete picture, participants should be included that represent the varying bilingual education models and who have varying levels of both Spanish and English proficiency.

Suggestions for Future Research

There is a critical need to collect further data in response to these and similar research questions. After an extensive review of the research, no information was found on how ELLs perform on the *Batería III*. Yet, conducting English only assessments on bilingual students will only provide part of their cognitive profile. Much more research is needed examining the performance of U.S. bilingual students on the *Batería III* as well as the other Spanish language test available, the WISC-IV. Certainly, similar research should include ELL students from the continuum of bilingual education models and varying amounts of native language instruction to more clearly view the relationship between performance

and first and second language proficiency. Another interesting line of research might include students who are not only bilingual, but those whose third language may include an indigenous language in which they may or may not have had formal instruction.

Increasing the empirical data examining the links between CHC abilities and interventions would benefit the educational field. To bolster the existing data, investigating the links between individual's CHC strengths and weaknesses identified on the *Batería III* and Spanish interventions would provide important information within RTI models.

The final implications discussed relate to the states that choose eligibility criteria for SLD that includes examining students' cognitive strengths and weaknesses. Interpreting standardized assessment results in the manner described here will comply with state regulations. Once again, much research is needed regarding the link between strengths and weaknesses and effective interventions.

Conclusion

In conclusion, this research attempted to provide a first step in solving the 40-year-old problem of disproportionate representation of cultural and linguistic minority students in special education that result from the misidentification of students. While limited in sample size, significant results were found that need further investigation. All students have the right to have their individual educational needs met in the most appropriate setting and least restrictive

environment. This can only be accomplished by ensuring that eligibility decisions are based on nondiscriminatory assessment processes and that the interpretation of testing results be made in a systematic and legally defensible manner. Only when eligibility teams learn to reliably determine difference from disability will children like Felipe receive the services needed regardless of their English proficiency level.

REFERENCES

- Acevedo-Polakovich, I. D., Reynaga-Abiko, G., Garritto, P. O., Derefinko, K. J., Wimsatt, M. K., Gudonis, L. D., & Brown, T. L. (2007). Beyond instrument selection: Cultural considerations in the psychological Assessment of U.S. Latinas/os. *Professional Psychology: Research and Practice*, 38(4), 375-384.
- Artiles, A. J., Aguirre-Munoz, A., & Abedi, J. (1998). Predicting placement in learning disabilities programs: Do predictors vary by ethnic group? *Exceptional Children*, 64(4), 543-559.
- Artiles, A. J., & Ortiz, A. (Eds.). (2002). *English language learners with special education needs: Identification, assessment, and instruction*. McHenry, IL: Delta Systems.
- August, D., & Hakuta, K. (Eds.). (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.
- 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(6), 599-602.
- Baker, C. (2000). *Foundations of bilingual education and bilingualism* (3rd ed.). Clevedon, United Kingdom: Multilingual Matters.
- Baker, M. (2003). *Lexical categories: Verbs, nouns, and adjectives*. New York: Cambridge University Press.
- Bateman, B. (1965). An educational view of a diagnostic approach to learning disorders. In J. Hellmuth (Ed.), *Learning disorders* (vol. 1; pp. 219 – 239). Seattle, WA: Special Child Publications.
- Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovalesski, J. F., Prasse, D., & et al. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: NASDSE.

- Bialystok, E. (1991). Letters, sounds and symbols: Changes in children's understanding of written language. *Applied Linguistics*, 12(1), 75-89.
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy and cognition*. New York: Cambridge University Press.
- Bilingual Education Act (1968).
- Bradley, R., Danielson, L., & Hallahan, D. P. (Eds). (2002). *Identification of learning disabilities: Research to practice*. Mahway, NJ: Lawrence Erlbaum Associates Publishers.
- Brooks-Gunn, J., Klebanov, P. K., & Duncan, G. J. (1996). Ethnic differences in children's intelligence test scores: Role of economic deprivation, home environment, and maternal characteristics. *Child Development*, 67, 396-408.
- Brown, J. E. (2005, September). *Use of the culture-language interpretive matrix: A framework for non-biased assessments*. Poster session presented at the annual LASER Conference, Tampa, FL.
- Brown v. Board of Education of Topeka, 347 U.S. 483, 74 S. Ct. 686, 91 L.ED 873 (1954).
- Casas, J. M., & Pytluk, S. D. (1995). Hispanic identity development: Implications for research and practice. In J. G. Ponterotto, J. M. Casas, L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (pp. 155-180). Thousand Oaks, CA: Sage.
- Castaneda v. Pickard, 648 F. 2d, 989, 5th Cir. (1981).
- Chambers, J. G., Shkolnik, J., & Perez, M. (2003). *Total expenditures for students with disabilities: Variation by disability*. Retrieved January 22, 2005, from <http://www.csef-air.org/>
- Civil Rights Act, 42 U.S. C. S 2000d-2000d-7 (1964).
- Cleary, T. A., Humphreys, L. G., Kendrick, S. A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist*, 15, 461-470.
- Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed.). Mahway, NJ: Lawrence Erlbaum.

- Cole, M., & Engestrom, Y. (1993). A cultural-historical approach to distributed cognition. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 1-46). New York: Cambridge University Press.
- Collier, C. (1986). *Acculturation quick screen*. Ferndale, WA: Crosscultural Development Education Services.
- Collier, C. (1988). *Acculturation quick screen*. Ferndale, WA: Crosscultural Development Education Services.
- Collier, C. (2000). *Acculturation quick screen*. Ferndale, WA: Crosscultural Development Education Services.
- Cosentino de Cohen, C., & Clewell, B. C. (2007). Putting English language learners on the educational map: The No Child Left Behind Act implemented. *Education in Focus Policy Brief*. Lansing, MI: The Education Policy Center.
- Crawford, J. (2004). *Educating English learners: Language diversity in the classroom* (5th ed.). Los Angeles, CA: Bilingual Educational Services, Inc.
- Cruickshank, W. M., Bentzen, F. A., Ratzeburg, F. H., & Tannhauser, M. T. (1961). *A teaching method of brain-injured and hyperactive children*. Syracuse, NY: Syracuse University Press.
- Cummins, J. C. (1982). Tests, achievement, and bilingual students. *NABE Journal*, 4(3), 1-19.
- Cummins, J. C. (1984). *Bilingual and special education: Issues in assessment and pedagogy*. Austin, TX: Pro-Ed.
- DeAvila, E., & Duncan, S. (1990). *LAS, Language assessment scales: Oral technical report, English, Forms 1C, 1D, 2C, 2D, Spanish, Forms 1B, 2B*. Monterey, CA: CTB MacMillan McGraw-Hill.
- DeAvila, E., & Duncan, S. (1991). *Language assessment scales*. San Rafael, CA: Linguametrics.
- DeAvila, E., & Duncan, S. (2005). *Language assessment scales*. New York: McGraw-Hill.

- DeValenzuela, J. S., Connery, M. C., & Musanti, S. I. (2000). The theoretical Foundations of professional development in special education: Is sociocultural theory enough? *Remedial and Special Education*, 21(2), 111-120.
- Delfior, S. K., & Serrano, F. (2005). The initial development of spelling in Spanish: From global to analytical. *Reading and Writing*, 18(1), 81-98.
- Diana v. State Board of Education, No. C-70-37, N.D. CA (1970).
- Donovan, S., & Cross, C. (2002). *Minority students in special and gifted education*. Washington, DC: National Academies Press.
- Dunn, L. M. (1968). Special education for the mildly retarded – is much of it justifiable? *Exceptional Children*, 23, 5-21.
- Education for All Handicapped Children Act of 1975, P.L. 94-142, 20 U.S.C. §1400 *et seq.* (1975).
- Education of the Handicapped Act P.L. 91-230 (1970).
- Ehrman, M. E. (1996). *Understanding second language learning difficulties*. Thousand Oaks, CA: Sage.
- Elementary and Secondary Education Act of 1965, P.L. 89-10, 79 Stat. 27 (1965).
- Espinosa, L. M. (2005). Curriculum and assessment considerations for young children from culturally, linguistically, and economically diverse backgrounds. *Psychology in the Schools*, 42(8), 837-853.
- Esters, I. G., & Ittenbach, R. F. (1997). Today's IQ tests: Are they really better than their historical predecessors? *School Psychology Review*, 26(2), 211-224.
- Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2001). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review*, 31(2), 246-262.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

- Figueroa, R. A. (1990). Assessment of linguistic minority group children. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Vol. 1: Intelligence and achievement*. New York: Guilford.
- Fiorello, C. A., Hale, J. B., & Snyder, L. E. (2006). Cognitive hypothesis testing and response to intervention for children with reading problems. *Psychology in the Schools, 43*(8), 835-853.
- Fiorello, C. A., & Primerano, D. (2005). Cattell-Horn-Carroll cognitive assessment in practice: Eligibility and program development issues. *Psychology in the Schools, 42*(5), 525-536.
- Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment*. New York: John Wiley & Sons, Inc.
- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross battery assessment* (2nd ed.). New York: John Wiley & Sons, Inc.
- Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Dynda, A. M. (2006). Integration of response to intervention and norm-referenced tests in learning disability identification: Learning from the Tower of Babel. *Psychology in the Schools, 43*(7), 807-825.
- Fletcher, J. M., Lyon, G. R., Barnes, M., Stuebing, K. K., Francis, D. J., Olson, R. K., Shaywitz, S. E., & Shaywitz, B. A. (2002). Classification of learning disabilities: An evidence-based evaluation. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 185-250). Mahway, NJ: Lawrence Erlbaum Associates.
- Floyd, R. G., Bergeron, R., & Alfonso, V. C. (2006). Cattell-Horn-Carroll cognitive ability profiles of poor comprehenders. *Reading and Writing: An Interdisciplinary Journal, 19*(5), 427-456.
- Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measure of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools, 40*(2), 155-170.

- Floyd, R. G., Keith, T. Z., & Taub, G. E. (2007). Cattell-Horn-Carroll Cognitive Abilities and their effects on reading decoding skills: g has indirect effects, more specific abilities have direct effect. *School Psychology Quarterly*, 22(2), 200-233.
- Floyd, R. G., McGrew, K. S., Evans, J. J. (2008). The relative contributions of the Cattell-Horn-Carroll cognitive abilities in explaining writing achievement during childhood and adolescence. *Psychology in the Schools*, 45(2), 132-144.
- Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice*, 18(3), 157-171.
- Garcia-Vazquez, E. (1995). Acculturation and academics: Effects of acculturation on reading achievement among Mexican-American students. *The Bilingual Research Journal*, 19(2), 305-315.
- Gartner, A., & Lipsky, K. (1987). Beyond special education: Toward a quality system for all students. *Harvard Education Review*, 57, 367-395.
- Gay, L. R., & Airasian, P. W. (1996). *Educational research*. Upper Saddle River, NJ: Pearson Higher Education.
- Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational research: Competencies for analysis and applications* (8th ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- Gordon, R. A. (1980). Labeling theory, mental retardation, and public policy: Larry P. and other developments since 1974. In W. R. Gove (Ed.), *The labeling of deviance* (2nd ed.). Beverly Hills, CA: Sage.
- Gottlieb, J., Alter, M., Gottlieb, B. W., & Wishner, J. (1994). Special education in urban America: It's not justifiable for many. *Journal of Special Education*, 27(4), 453-465.
- Graham, J. R., Naglieri, J. A., & Weiner, I. B. (Eds.). (2003). *Handbook of psychology* (vol. 10). New York: John Wiley and Sons.

- Gravois, T., & Rosenfield, S. A. (2006). Impact of instructional consultation teams on the disproportionate referral and placement of minority students in special education. *Remedial and Special Education, 27*(1), 42-52.
- Green, T. D., McIntosh, A. S., Cook-Morales, V. J., & Robinson-Zanartu, C. (2005). From old schools to tomorrow's schools: Psychoeducational assessment of African American students. *Remedial and Special Education, 26*(2), 82-92.
- Greene, J. P. (1998). *A meta-analysis of the effectiveness of bilingual education*. Claremont, CA: Tomas Rivera Policy Institute.
- Greenfield, P. M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve* (pp. 81-124). Washington, DC: American Psychological Association.
- Gunderson, L., & Siegel, L. S. (2001). The evils of the use of IQ tests to define learning disabilities in first- and second-language learners. *The Reading Teacher, 55*(1), 48-55.
- Hakuta, K. (1990). Language and cognition in bilingual education. In A. M. Padilla, H. H. Fairchild, & C. M. Valadez (Eds.), *Bilingual education: Issues and strategies* (pp. 47-59). Newbury Park, CA: Sage.
- Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Hoepfner, J. A. B., & Gaither, R. A. (2001, Spring). WISC-III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable? *School Psychology Quarterly, 16*(1), 31-55.
- Hale, J. B., Kaufman, A., Naglieri, J. A., & Kavale, K. A. (2006). Implementation of IDEA: Integrating response to intervention and cognitive assessment methods. *Psychology in the Schools, 43*(7), 753-770.
- Hallahan, D. P., & Mercer, C. D. (2001, August). *LD: Historical perspectives*. Paper presented at the LD Summit, U.S. Department of Education, Washington, DC.
- Hallahan, D. P., & Mercer, C. D. (2002). Learning disabilities: Historical perspectives. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 1-73). Mahway, NJ: Lawrence Erlbaum Associates, Publishers.

- Hallahan, D. P., & Mock, D. R. (2003). A brief history of the field of learning disabilities. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 16-29). New York: The Guilford Press.
- Hargrove, M. L. (1982). Learning disabilities: Early identification signals for parents and teachers. *Education*, 102(4), 366-368.
- Harry, B., & Klingner, J. (2007). Discarding the deficit model. *Educational Leadership*, 64(5), 16-21.
- Harry, B., Klingner, J., Sturges, K. M., & Moore, R. F. (2002). Using qualitative methods to investigate disproportionality: Of rocks and soft places. In D. J. Losen & G. Orfield (Eds.), *Racial inequity in special education* (pp. 71-92). Cambridge, MA: Harvard Civil Rights Project.
- Heath, S. B. (1983). What no bedtime story means: Narrative skills at home and school. *Language and Society*, 11, 49-76.
- Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, 47, 1083-1101.
- Helms, J. E. (1997). The triple quandary of race, culture and social class in intelligence testing. In D. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Beyond traditional intellectual assessment: Contemporary and emerging theories, tests, and issues* (pp. 517-532). New York: Guilford.
- Herrnstein, R. J., & Murray, C. (1994). *Bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Heufner, D. S. (2000). *Getting comfortable with special education law: A framework for working with children with disabilities*. Norwood, MA: Christopher-Gordon.
- Heward, W. L. (2003). Ten faulty notions about teaching and learning that hinder the effectiveness of special education. *The Journal of Special Education*, 36(4), 186-205.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57(5), 253-270.

- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). New York: Guilford Press.
- Hosp, J. L., & Reschly, D. J. (2003). Referral rates for intervention or assessment: A meta-analysis of racial differences. *The Journal of Special Education*, 37(2), 67-80.
- Individuals with Disabilities Education Act of 1990, P.L. 101-476 (1990).
- Individuals with Disabilities Education Act of 1997, P.L. 105-117 (1997).
- Individuals with Disabilities Educational Improvement Act of 2004, P.L. 108-446 (2004).
- Jensen, A. R. (1974). How biased are culture-loaded tests? *Genetic Psychology Monographs*, 90, 185-244.
- Jensen, A. R. (1982). Reaction time and psychometric g. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93-132). New York: Springer.
- Katsiyannis, A., Yell, M. L., & Bradley, R. (2001). Reflections on the 25th anniversary of the individuals with disabilities education act. *Remedial and Special Education*, 22(6), 324-335.
- Kauffman, J. M. (1996). Research to practice issues. *Behavioral Disorders*, 22(1), 55-60.
- Kavale, K. A., & Forness, S. R. (1994). Learning disabilities and intelligence: An uneasy alliance. In T. E. Scruggs & M. M. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (pp. 1-63). Greenwich, CT: Jai Press.
- Kavale, K. A., Holdnack, J. A., & Mostert, M. P. (2005). Responsiveness to intervention and the identification of specific learning disability: A critique and alternative proposal. *Learning Disability Quarterly*, 28, 2-16.
- Kirk, S. A. (1963, April 6). Behavioral diagnosis and remediation of learning disabilities. *Proceedings of the First Annual Meeting of the ACLD Conference on Exploration into the Problems of the Perceptually Handicapped Child*, Chicago, IL.

- Kranzler, J. H. (1997). Educational and policy issues related to the use and interpretation of intelligence tests in schools. *School Psychology Review*, 26(2), 150-162.
- Krashen, S. (1999). *Condemned without a trial: Bogus arguments against bilingual education*. Portsmouth, NH: Heinemann.
- Laosa, L. M. (2001). School segregation of children who migrate to the United States from Puerto Rico. In G. V. Glass (Ed.), *Education policy analysis archives, 2001* (pp. 1-57). Retrieved September 10, 2005, from <http://olam.ed.asu.edu/epaa/>
- Larry P. v. Riles, 495, F Supp. 926, N.D. CA (1979).
- Lau v. Nichols, No. 72-6520, 414 U.S. 56 (1974).
- Losen, D. J., & Orfield, G. (Eds.). (2002). *Racial inequity in special education*. Cambridge, MA: Harvard Education Press.
- Lyon, G. F., & Chhabra, V. (2002). *The current state of the science and the future of specific reading disabilities*. Denver: Colorado Reading Center. Retrieved March 15, 2008, from <http://www.coloradoreading.com/nichad.htm>
- Macedo, D., Dendrinos, B., & Gounari, P. (2003). *The hegemony of English*. Boulder, CO: Paradigm Publishers.
- MacMillan, D. L., & Forness, S. R. (1998). The role of IQ in special education placement decisions: Primary and determinative or peripheral and inconsequential? *Remedial and Special Education*, 19(4), 239-253.
- Mather, N., & Gregg, N. (2006). Specific learning disabilities: Clarifying, not eliminating, a construct. *Professional Psychology: Research and Practice*, 37(1), 99-106.
- Matsumoto, D. (1994). *Cultural influences on research methods and statistics*. Pacific Grove, CA: Brooks/Cole.
- McDermott, R., Goldman, S., & Varenne, H. (2006). The cultural work of learning disabilities. *Educational Researcher*, 35(6), 12-17.

- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 136-182). New York: Guilford.
- McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference: Gf-Gc cross-battery assessment*. Boston, MA: Allyn and Bacon.
- McHatton, P. A., & Correa, V. (2005). Stigma and discrimination: Perspectives from Mexican and Puerto Rican mothers of children with special needs. *Topics in Early Childhood Special Education*, 25(3), 131-142.
- Mellard, D. (2004). *Understanding responsiveness to intervention in learning disabilities determination*. Retrieved July 1, 2005, from www.nrld.org/html/information
- Mercer, J. R. (1979). *The system of multicultural pluralistic assessment: Technical manual*. New York: Psychological Corporation.
- Miles, M., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Mills v. D.C. Board of Education, 348 E. Supp. 866, D.D.C. (1972).
- Monroe, M. (1932). *Children who cannot read*. Chicago: University Park Press.
- Monzó, L. D., & Rueda, R. (2006). A sociocultural perspective on acculturation: Latino immigrant families negotiating diverse discipline practices. *Education and Urban Society*, 38(2), 188-203.
- Munoz-Sandoval, F. G., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *Batería III Woodcock-Munoz*. Itasca, IL: Riverside Publishing.
- Naglieri, J. A. (1999). How valid is the PASS theory and CAS? *School Psychology Review*, 28(1), 145-162.
- National Association of State Directors of Special Education, Inc. (2006). *Response to Intervention: Policy considerations and implementation*. Alexandria, VA: Author.

- National Research Council. (2002). *Minority students in special and gifted education*. (M. S. Donovan & C. T. Cross, Eds.). Washington, DC: National Academy Press.
- Nelson, K. (1996). *Language in cognitive development: The emergence of the mediated mind*. New York: Cambridge University Press.
- Ochoa, S. H., Rivera, B., & 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.
- O'Connor, C., & Fernandez, S. D. (2006). Race, class, and disproportionality: reevaluating the relationship between poverty and special education placement. *Educational Researcher*, 35(6), 6-11.
- Office of Civil Rights. (1970, May 25). *Office of Civil Rights memo*. Washington, DC: Department of Health, Education, and Welfare, Office of the Secretary.
- Ofiesh, N. (2006). Response to intervention and the identification of specific learning disabilities: Why we need comprehensive evaluations as part of the process. *Psychology in the Schools*, 43(3), 883-888.
- Onwuegbuzie, A. J., & Daley, C. E. (2001). Racial differences in IQ revisited: A synthesis of nearly a century of research. *Journal of Black Psychology*, 27(2), 209-220.
- Ortiz, S. O., & Flanagan, D. P. (2002). Best practices in working with culturally diverse children and their families. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (vol. 2, pp. 337-351). Bethesda, MD: The National Association of School Psychology.
- Orton, S.T. (1939). A neurological explanation of the reading disability. *The Educational Record*, 20(Suppl. 12), 58-68.
- Otero, M. (2006). Test review: Bateria III Woodcock-Muñoz. *The School Psychologist*, 60(2), 86-89.
- Padilla, A. M. (1980). The role of cultural awareness and ethnic loyalty. In A. M. Padilla (Ed.) *Acculturation: Theory, models and some new findings* (pp. 47-84). Boulder: Westview Press.

- Padilla, A. M., & Perez, W. (2003). Acculturation, social identity, social cognition: A new perspective. *Hispanic Journal of Behavioral Sciences*, 25(1), 35-55.
- PASE (Parents in Action on Special Education) v. Joseph P. Hannon, 506 F. Supp. 831, N.D. Ill (1980).
- Pasternack, R. H. (2002, March). *The demise of IQ testing for children with learning disabilities*. Keynote Address presented at the annual convention of the National Association of School Psychologists, Chicago.
- Pennsylvania Association for Retarded Citizens (PARC) v. Commonwealth of Pennsylvania, 334 E Supp. 1257(E.D. Pa. 1972).
- Peterson, K. M., & Shin, M. R. (2002). Severe discrepancy models: Which best explains school identification practices for learning disabilities. *School Psychology Review*, 31, 459-476.
- Polloway, E. A. (2002). The profession of learning disabilities: Progress and promises. *Learning Disabilities Quarterly*, 25, 103-112.
- Poplin, M. S. (1988). Holistic/constructivist principles of the teaching/learning process: Implications for the field of learning disabilities. *Journal of Learning Disabilities*, 21(4) 389-400.
- Proctor, B. E., Floyd, R. G., & Shaver, R. B. (2005). Cattell-Horn-Carroll broad cognitive activity profiles of low math achievers. *Psychology in the Schools* 42(1), 1-12.
- Ramirez, J. D. (1992). Executive summary. *Bilingual Research Journal*, 16(1-2), 1-62.
- Rehabilitation Act of 1973, P.L. 93-112, 87, Stat. 394 (1973).
- Reid, D. K., & Knight, M. G. (2006). Disability justifies exclusion of minority students: A critical history grounded in disability studies. *Educational Researcher*, 35(6), 18-23.
- Reschly, D. J. (1997). Diagnostic and treatment utility of intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 437-456). New York: Guilford.

- Rethinking Schools. (2003, Spring). History of bilingual education. *Rethinking Schools Online*, 12(3). Retrieved August 2, 2007, from <http://www.rethinkingschools.org>
- Reynolds, C. R. (1982). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (pp. 178-208). New York: Wiley.
- Rogoff, B. (1990). *Apprenticeship in thinking*. New York: Oxford University Press.
- Rossell, C. H. (1999). *Mystery on the bilingual express: A critique of the Thomas and Collier study*. Retrieved February 18, 2006, from <http://www.read-institute.org/collier.html>
- Salend, S. J., Duhaney, G., & Montgomery, L. M. (2002). A comprehensive approach to identifying and addressing issues of disproportionate representation. *Remedial and Special Education*, 23(5), 289-300.
- Samuda, R. J., Kong, S. L., Cummins, J., Pascual-Leone, J., & Lewis, J. (1991). *Assessment and placement of minority students*. New York: D. J. Hogrefe/Intercultural Social Sciences.
- Sanchez, G. I. (1934). Bilingualism and mental measures. *Journal of Applied Psychology*, 8, 765-772.
- Sandoval, J. (1979). The WISC-R and internal evidence of test bias with minority groups. *Journal of Consulting and Clinical Psychology*, 47, 919-927.
- Sandoval, J., Zimmerman, I. L., & Woo, J. M. (1980, September). *Cultural differences on WISC-R verbal items*. Paper presented at the annual convention of the American Psychological Association, Montreal, Canada.
- Scarcella, R. (1990). *Teaching language minority students in the multicultural classroom*. Upper Saddle River, NJ: Prentice Hall Regents.
- Schrank, F. A., & Flanagan, D. P. (2003). *WJ III clinical use and interpretation*. San Diego: Academic Press.
- Schrank, F., McGrew, K., Ruef, M. L., Alvarado, C. G., Munoz-Sandoval, A. F., & Woodcock, R. W. (2005). *Overview and technical supplement*. Itasca, IL: Riverside Publishing.

- Skiba, R. G., Knesting, K., & Bush, L. D. (2002). Culturally competent assessment: More than nonbiased tests. *Journal of Child and Family Studies*, 11(1), 61-78.
- Skiba, R., Simmons, A., Ritter, S., Kohler, K., Henderson, M., & Wu, T. (2006). The context of minority disproportionality: Practitioner perspectives on special education. *Teachers College Record*, 108(7), 1424-1459.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.) (1998). *Preventing reading difficulties in young children*. Washington, DC: National Research Council.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research*, 21(4), 360-407.
- Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profiles of children with reading disabilities: A regression-based test of the phonological-core variable difference model. *Journal of Educational Psychology*, 86, 24-53.
- Swanson, H. L., Harris, K. R., & Graham, S. (Eds.) (2002). *Handbook of learning disabilities*. New York: The Guilford Press.
- Thomas, W. P., & Collier, V. P. (1997). *School effectiveness for language minority students*. Retrieved January 15, 2007, from <http://www.ncela.gwu.edu/ncbepubs/Resource/effectiveness/index.htm>
- Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement: Final report*. Retrieved January 15, 2007, from http://www.crede.ucsc.edu/research/llaa/1.1_final.html
- Tse, L. (2001). *Why don't they learn English: Separating fact from fallacy in the U.S. language debate*. New York: Teachers College Press.
- Urdan, T. G. (2001). *Statistics in plain English*. Mahway, NJ: Lawrence Erlbaum Association.
- U.S. Department of Education. (1999, March 12). Education department: Rules (34 C.F.R. 300.540-544). *Federal Register*, 54(121), 12541.

- U.S. Department of Education. (2004). Public teacher file: Data analysis system. Washington, DC: National Center for Education Statistics, Schools and Staffing Survey (SASS).
- Valdes, G., & Figueroa, R. A. (1994). *Bilingualism and testing: A special case of bias*. Norwood, NJ: Ablex.
- Vukovich, D., & Figueroa, R. A. (1982). *The validation of the system of multicultural pluralistic assessment: 1980-1982*. Unpublished manuscript, University of California at Davis, Department of Education, Davis, CA.
- Wasserman, J. D., & Tulsy, D. S. (2005). A history of intellectual assessment. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 3-22). New York: The Guilford Press.
- Weiner, I. B., Freedheim, D. K., Velicer, W. F., Schinka, J. A., & Lerner, P. M. (2003). The assessment process. In J. R. Graham, J. A. Naglieri, & I. B. Weiner (Eds.), *Handbook of psychology* (vol. 10; pp 3-27). New York: John Wiley and Sons.
- Weintraub, F. (2005). The evolution of LD policy and future challenges. *Learning Disability Quarterly*, 28(2), 97-99.
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Eds.), *Structural equation modeling: Concepts, issues and applications* (pp. 56-75). Thousand Oaks, CA: Sage Publications.
- Williams, R. L. (1970). Black pride, academic relevance and individual achievement. *Counseling Psychologist*, 2(1), 18-22.
- Williams, R. L. (1971). Abuses and misuses in testing black children. *Counseling Psychology*, 2(3), 62-73.
- Williams, W. M., & Ceci, S. J. (1997). Are Americans becoming more or less alike? Trends in race, class, and ability differences in intelligence. *American Psychologist*, 52(11), 1226-1235.
- Willis, J. O., & Dumont, R. (2006). And never the twain shall meet: Can response to intervention and cognitive assessment be reconciled? *Psychology in the Schools*, 43(8), 901-908.

- Woodcock, R. W. (1982). *Batería Woodcock psico-educativa en Espanol*. Itasca, IL: Riverside Publishing.
- Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery-Revised*. Itasca, IL: Riverside Publishing.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside Publishing.
- Woodcock, R. W., Muñoz, A. S., McGrew, K., Mather, N., & Schrank, F. (2004). *Batería III pruebas de habilidades cognitivas*. Itasca, IL: Riverside Publishing.
- Woodcock, R. W., & Muñoz-Sandoval, A. (1996). *Woodcock psico-educativa en Español - Revisada*. Itasca, IL: Riverside Publishing.
- Xu, Y., & Drame, E. (2008). Culturally appropriate context: Unlocking the potential of response to intervention for English language learners. *Early Childhood Education Journal*, 35, 305-311.
- Yell, M., & Rogers, D. (1998). The legal history of special education. *Remedial and Special Educators*, 19(4), 219-229.
- Yerkes, R. M. (Ed.). (1921). Psychological examining in the United States army. *Memoirs of the National Academy of Sciences*, 1, 252-259.
- Zettel, J. J., & Ballard, J. (1982). The education for all handicapped children act of 1975 (P.L. 94-142): Its history, origins, and concepts. In J. Ballard, B. Ramirez, & F. Weintraub (Eds.), *Special education in America: Its legal and governmental foundations* (pp. 23-40). Reston, VA: Council for Exceptional Children.
- Zhang, D., & Benz, M. (2006). Enhancing self-determination of culturally diverse students with disabilities: Current status and future directions. *Focus on Exceptional Children*, 38(9), 1-12.

APPENDIX A
INFORMED CONSENT

PLEASE GIVE PERMISSION FOR YOUR CHILD TO BE PART OF AN IMPORTANT PROJECT

Your child is invited to participate in a research study conducted by Julie Esparza Brown, a doctoral student in Educational Leadership in Special Education from Portland State University, Graduate School of Education. The study will help educators learn how bilingual students in the United States perform on assessments that test their abilities. Julie Esparza Brown is bilingual and a trained school psychologist qualified to give cognitive tests. The information from this study is very important to professionals who are responsible for determining the reasons why some bilingual students have academic difficulties. The difficulties might be because the student comes from a different culture and speaks a different language. As you can see, the results of the study will be very important to all educators. Your child's participation is voluntary. Your child does not have to take part in this study if he/she does not want to, and it will not affect his/her grade or educational program. If your child does participate in the project, he/she will receive a book in exchange for giving their individual time and effort for the study. You may choose to withdraw your child from this study at any time without affecting the relationship with the researcher, school personnel or other participants.

If you decide to allow your child to participate, you will be asked to:

- ◆ Allow Julie to administer an individual cognitive test (I.Q. test) in Spanish (the Batería III) to your child outside of the classroom setting for two to three hours (not during your child's literacy time).
- ◆ Give permission for the results to be combined with scores from all other participants in order to get a group average score.
- ◆ Give permission to access your child's confidential school records so the researcher can document information on your child's educational history including language(s) of instruction as well as your child's language proficiency information.
- ◆ Allow Julie to administer the Language Assessment Scales (LAS) – Oral in both English and Spanish.

In exchange for your child's participation, she/he will receive a reading book. Julie is an experienced educator who will make your child feel comfortable in the situation and will remind your child that they may choose not to participate at any point without any penalty or problem.

Information collected in connection with this study and that can be linked to your child or his/her identity will be kept confidential to guarantee your privacy. No individual information will be reported. To maintain confidentiality, all students will be given a number instead of using their name. Only the researcher will have access to the students' real names. All data will be kept secure in a locked file cabinet in the researcher's office. The data will be kept for only three years and will be destroyed after June 1, 2010. When information from this study is shared in the dissertation, articles or presentations, your name and school information will not be used and to protect your child's identity your child's subtest scores will be combined with all other student's scores to get a group

average score. Once again, individual scores will not be available as all data will be combined in order to get averages scores for the participating group.

If you have concerns or problems about your participation in this study or your rights as a research subject, please contact the Human Subjects Research Review Committee, Office of Research and Sponsored Projects, 111 Cramer Hall, Portland State University, (503) 725-4288. If you have questions about the study, please contact Julie Esparza Brown in room 608B of the Graduate School of Education, Portland State University, by telephone (503) 725-4704, or by email at jebrown@pdx.edu.

Your signature below means that you have read and understand the above information and that you agree to allow your child to take part in this study. Please understand that you may withdraw your consent at any time without penalty, and that, by signing, you are not waiving any legal claims, rights or remedies. You will receive a copy of this letter for your records. Thank you very much!

☐ Yes, I agree for my child to participate.

☐ No, I do not wish for my child to participate.

Signature of Parent

Date

Witness

Date

Name of the Child

APPENDIX B
TRANSLATED PARENT CONSENT

Favor de dar permiso que su hijo/hija participe en un proyecto muy importante.

Su estudiante está invitado a participar en un estudio académico conducido por Julie Esparza Brown, estudiante doctoral de la Universidad Estatal de Portland (*Portland State University*) en Educación Especial. Este estudio ayudará a educadores aprender como estudiantes bilingües en los Estados Unidos se desempeñan al tomar evaluaciones que examinan sus habilidades (inteligencia). Julie Esparza Brown es una psicóloga escolar con experiencia en administrar exámenes cognitivos (de inteligencia). La información que este estudio brindará será muy útil y importante a educadores del porque algunos estudiantes bilingües tienen dificultades académicas. Estas dificultades podrían ser conectadas con la cultura y lenguaje del estudiante bilingüe. **La participación de su estudiante es completamente voluntaria y confidencial.** Si su hijo o hija decide en cualquier momento que no quiere continuar tomando parte en este estudio académico no habrá ninguna repercusión. Su decisión de no continuar no afectará su relación con la escuela, los maestros, el programa educacional o la Sra. Esparza Brown. En agradecimiento por su participación, tiempo y esfuerzo, su estudiante recibirá un libro de lectura.

Su permiso incluye que:

- ❖ Julie Esparza Brown administre un examen cognitivo individual en español (la Batería III) a su hijo/hija. Éste examen es para medir la inteligencia. El examen durará 2 a 3 horas y será administrado para que no interfiera durante la enseñanza de literatura. Para reducir distracciones también será administrado fuera del salón de clase.
- ❖ Los resultados sean combinados con los resultados de todos los otros participantes para poder establecer un **promedio del grupo**.
- ❖ Julie Esparza Brown administre un examen del idioma inglés y uno de español (Language Assessment Scales, English and Spanish).
- ❖ Julie Esparza Brown tenga acceso a los archivos académicos de su estudiante para documentar información acerca de los lenguajes en cuales ha recibido instrucción académica y su preeficiencia en los lenguajes que el/ella usa.

Para garantizar la privacidad del estudiante y su información particular, se les dará a cada uno un número de identificación en vez de usar su nombre. Solo Julie Esparza Brown tendrá acceso a los nombres de los participantes y toda información colectada o archivos proporcionados por la escuela se mantendrán seguramente bajo llave. Los resultados serán destruidos después del 1 de junio 2010. Cuando los resultados de este estudio sean compartidos en presentaciones o artículos, los nombres de participantes e información escolar no serán usados. Adicionalmente, todos los resultados individuales se combinarán para establecer un **promedio del grupo**.

Como psicóloga escolar, la Sra. Esparza Brown tiene mucha experiencia en administrar estos tipos de evaluaciones y asegurará la privacidad y que su participante se sienta cómodo durante el examen.

Al marcar la cuadro abajo y firmar en la linea, ustedes me dan a saber que han leído y entendido la información contenida en este documento y que están dando permiso de participación en el estudio académico a su estudiante. Su firma **no** indica que esta renunciando a su derecho legal.

Favor de acordarse que en cualquier momento o por cualquier razón usted y su estudiante pueden dejar de participar sin ninguno problema.

☐ Sí, doy permiso para que mi hijo/a participe.

☐ No, no doy permiso.

Firma del Padre o Madre

Fecha

Testigo

Fecha

Nombre de Estudiante

Si tiene preguntas o preocupaciones acerca de su participación en este estudio académico o sus derechos como participante, favor de contactar al comité de Human Subjects Research Review, Office of Research and Sponsored Projects, 111 Cramer Hall, Portland State University, (503) 725-4288.

Si tienen preguntas acerca del estudio académico, favor de contactar directamente a Julie Esparza Brown por medio de teléfono (503) 725-4704, correo electrónico jebrown@pdx.edu o visítela en el cuarto #608B del edificio de la Escuela de Educación en Portland State University.

APPENDIX C

STUDENT ASSENT SCRIPT

Child Assent Script

My name is Julie Esparza Brown and I have been a teacher for many years. Right now I am also a student at Portland State University. Your mom and dad said it would be okay if you spend some time with me helping me with a very important project. I will be taking you to a quiet room where we will be doing some things together like looking at pictures and puzzles, listening to some stories and sounds, and seeing how fast you can match things. If you decide you don't want to come with me, you don't have to. Also, at any time if you don't want to help any more after we have already started, you can just say so and go back to your classroom. I do hope that you want to come and help out. To thank you for helping me, I will be giving you a book to keep. Does this sound okay with you?

(Child consents or does not consent to participate.)

Mi nombre es Julie Esparza Brown y ha estado una maestro por muchos años. Ahora, soy un estudiante in Portland State University. Tus padres son de acuerdos que pasaras tiempo conmigo en un proyecto importante. Vamos a otra sala para hacer algunos cosas como rompecabezas y escuchar cuentos. Si decides que no quieres acompañarme, no tienes que venir. Tambien, si ya no quieres ayudarme después de comenzar, dimélo y puedes regresar a tu clase. De versa espero que quieras venir para ayudarme. Para agradecerte por tu ayuda, te daré un libro. ¿Sueña bien contigo?

APPENDIX D

STUDENT DATA SHEET

Date: _____

Student Data Sheet

1. Student Identification Number (for this project only): _____
2. Grade: _____
3. Age: _____
4. Country of Birth: _____
5. Primary Language: _____
6. Dominant Language (the language the student chooses to use with researcher): _____
7. Language Proficiency Scores:
Native Language _____ Date _____
English _____ Date _____
8. Language(s) of Instruction:
Native Language _____ Duration _____
English _____ Duration _____
9. Years of schooling outside of the U.S.: _____
10. Attach completed C-LIM.

APPENDIX E

AQS

Acculturation Quick Screen (AQS)

©2001 Dr. Catherine Collier

Newcomer ☐
Continuing ☐

NAME/ID #: _____ SCHOOL: _____
DATE OF BIRTH: _____ SEX: _____ GRADE: _____ AGE AT ARRIVAL IN N.Amer: _____
LANGUAGE(S) SPOKEN AT HOME: _____

CULTURAL/ENVIRONMENTAL FACTORS	Information	Score
Number of years in the United States		
Number of years in the School District		
Number of years in ESL and/or Bilingual Education		
Bilingual Proficiency		
Native Language Proficiency		
English Language Proficiency		
Ethnicity/Nation of Origin		
% of Minority Language in Present School		
AQS Score Total:		

AQS SCALE SCORING GUIDELINES

1. NUMBER of YEARS IN U.S./Canada Under one year = 1 One to two years = 2 Over two, up to four years = 3 Over four, up to five years = 4 Over five, up to six years = 5 Over six years = 6	2. NUMBER of YEARS IN DISTRICT Under one year = 1 One to two years = 2 Over two, up to four years = 3 Over four, up to five years = 4 Over five, up to six years = 5 Over six years = 6	3. YEARS IN ESL/BILINGUAL PROGRAM Up to one year in directed instruction = 1 Over one, up to 1½ years = 2 Over 1½, up to two years = 3 Over 2, up to 2½ years = 4 Over 2½, up to four years = 5 Over four years = 6
4. NATIVE LANGUAGE PROFICIENCY Does not speak the language = 1 Has receptive comprehension [understands when spoken to] = 2 Limited social speaking ability only e.g. can carry on a basic social conversation [BICS only] = 3 Intermediate social speaking and limited academic thinking abilities [intermediate BICS, limited CALP] = 4 Intermediate social speaking and academic thinking abilities [intermediate BICS & CALP] = 5 Advanced social speaking and academic thinking abilities [fluent BICS & CALP] = 6		
5. ENGLISH LANGUAGE PROFICIENCY Does not speak the language – Lower than LAS, SOLOM 1 = 1 Has receptive comprehension [understands when spoken to] – LAS, SOLOM 1 = 2 Limited social speaking ability only e.g. can carry on a basic social conversation [BICS only] – LAS, SOLOM 2 = 3 Intermediate social speaking & limited academic thinking abilities [intermediate BICS, limited CALP] – LAS, SOLOM 3 = 4 Intermediate social speaking and academic thinking abilities [intermediate BICS & CALP] – LAS, SOLOM 4 = 5 Advanced social speaking and academic thinking abilities [fluent BICS & CALP] – LAS, SOLOM 5 = 6		
6. BILINGUAL PROFICIENCY Essentially speaks only one language [monolingual] = 1 Primarily speaks one language [fluent BICS], can speak some second language [beginning BICS] = 2 Advanced speaker [fluent BICS] in one, intermediate speaker [intermediate BICS] in other = 3 Basic academic thinking [CALP] in one, intermediate speaker [intermediate BICS] in other = 4 Most academic thinking [CALP] in one, some ability to think in other = 5 Bilingual in social speaking [BICS] and academic thinking [CALP] = 6		
7. ETHNICITY/NATIONAL ORIGIN American Indian/Native American/Indigenous Populations/First People = 1 Hispanic/Latino/Chicano = 2 African, East Asian (Countries around Bay of Bengal), Pacific Islander = 3 West Asian (Countries around Arabian Sea) or Middle Eastern = 4 Eastern European = 5 Western European = 6	8. PERCENT SPEAKING STUDENT'S LANGUAGE 81% - 100% of enrollment = 1 65% - 80% of enrollment = 2 45% - 64% of enrollment = 3 25% - 44% of enrollment = 4 11% - 24% of enrollment = 5 0% - 10% of enrollment = 6	

8-14 Significantly less acculturated • 15-22 Less acculturated • 23-31 In transition •
32-39 More acculturated • 40-48 Highly acculturated