1998

Language Proficiency and Stress: Impact on Measured Intelligence and Anxiety for Latino Children

Marlene Baldizon de Naclerio

Seton Hall University

Follow this and additional works at: http://scholarship.shu.edu/dissertations

Part of the Developmental Psychology Commons, Linguistics Commons, and the Sociology Commons

Recommended Citation
http://scholarship.shu.edu/dissertations/1272
LANGUAGE PROFICIENCY AND STRESS: IMPACT ON MEASURED INTELLIGENCE AND ANXIETY FOR LATINO CHILDREN

BY

MARLENE BALDIZON DE NACLERIO

Dissertation Committee

Sandra S. Lee, Ph.D.
Juan Cobarrubias, Ph.D.
Sara Sparrow, Ph.D.

Submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy

Seton Hall University
1998
ACKNOWLEDGMENTS

I am indebted to my parents, César and Alba Baldizón, for their vision and faith...
"gracias por su amor". I thank my siblings, Florelia, César, Jorge, my nephews,
Joel and Julian, and my dear Ward and Sonia, for providing me with both experiential
and technical enlightenment as we grew bilingual. I am grateful to my friends and to "las
muchachas", specially Cristina Noguera and Mercy Casellas, and to my extended Naclerio
family, for their encouragement and ability to keep me on task.
I am also forever grateful to the Latino children of New Haven, Connecticut.
DEDICATION

I dedicate this project to my beloved husband and "alma gemela" Matthew Naclerio, for you are my inspiration and my life. You once offered the following to me:

If a man does not keep pace with his companions, perhaps it is because he hears a different drummer. Let him step to the music he hears, however measured or far away.

Thoreau

Thank you for your support and love as I stepped to the music I heard.
TABLE OF CONTENTS

ACKNOWLEDGMENTS .........................  ii
DEDICATION ................................... iii
LIST OF TABLES .............................. vii

Chapter

I INTRODUCTION .........................  1
   Statement of the Problem ..............  3
   Significance of the Problem ..........  4
   Significance of the Study .............  6
   Hypotheses ............................. 10
   Hypothesis 1 ........................... 11
   Hypothesis 2 ........................... 11
   Hypothesis 3 ........................... 11
   Hypothesis 4 ........................... 12
   Definition of Terms ................... 12
   Verbal and Performance Skills .......... 12
   Relative Language Proficiency ......... 12
   Testing Environment ................. 13
   Anxiety ............................... 13
   Language Shift ....................... 13
   Limitations of the Study ............ 13

II REVIEW OF THE LITERATURE .......... 18
   Intelligence and Bilingualism:
   1900 to 1970s ......................... 18
   Genetically Inferior or Language Handicapped? .. 20
   Psychometrics Vs Linguistics ....... 22
   Bilingual Children and Special Education .... 24
   Enrollment in Special Education Classes .... 24
   Bilingual Education Act .............. 24
   Impact on Assessment Practices ....... 26
Intelligence and Bilingualism:  
1970s to Present........ 26 
Studies with the WISC.... 26 
Verbal/Performance IQ Discrepancy... 29 
Reliability and Validity Studies 31 
Significance of the V/P discrepancy 36 
Limitations of Wechsler Scales Research... 38 
Procedural Errors... 38 
Theoretical Flaws... 44 
Linguistic Research........ 45 
Bilingualism Undefined?... 46 
Factors impacting on second language acquisition 47 
Bilingualism and Intelligence........ 48 
Bilingualism and the bilingual mind.. 51 
Bilingualism, Culture, and Stress..... 55 
Language, Stress, and Intelligence Tests... 58 
Integrated View of Bilingual Intelligence........ 61 

III METHODS....................... 62 
Overview and Design .......... 62 
Subjects......................... 63 
Procedure....................... 64 
Instruments.................... 66 
Wechsler Intelligence Scale for Children-III.... 66 
Woodcock-Muñoz Language Survey............. 68 
State-Trait Anxiety Inventory for Children...... 69 
Testing Environment.......... 70 

IV RESULTS.......................... 72 
Hypothesis 1................. 76 
Hypothesis 2................. 77
LIST OF TABLES

Table 1  Means and Standard Deviations for Woodcock-Muñoz Language Survey  73
Table 2  Means and Standard Deviations for the Woodcock-Muñoz Language Survey, English and Spanish  74
Table 3  Means and Standard Deviations for the WISC-III Sub-tests  75
Table 4  Summary of Stepwise Regression Analysis for Variables Predicting Verbal-Performance Discrepancy  77
Table 5  Pearson’s Correlation Coefficients for Variables Predicting Verbal-Performance Discrepancy  78
Table 6  Summary of Stepwise Regression Analysis for Variables Predicting Verbal-Performance Discrepancy  79
Table 7  Pearson’s Correlation Coefficients Between Woodcock-Muñoz-English (WME), Woodcock-Muñoz-Spanish (WMS) and Verbal-Performance Discrepancy  80
Table 8  Summary of Stepwise Regression for Variables Predicting Verbal Comprehension  83
Table 9  Summary of Stepwise Regression for Variables Predicting Perceptual Organization  83
Table 10  Summary of Stepwise Regression for Variables Predicting Freedom from Distractibility  83
Table 11 Summary of Stepwise Regression for Variables Predicting Processing Speed

Table 12 Pearson's Correlation Coefficient Between Woodcock-Huñoz-English (WME), Woodcock-Huñoz-Spanish (WNS) and WISC-III Factor Scores
Chapter I
INTRODUCTION

Language minority children in the United States are caught in a long-standing and fierce debate over the significance of their ethnic and linguistic background both for their own development (Gonzalez, 1994; Puente & Salazar, 1998) and for the growth of the nation (Crawford, 1989; Jones-Correa and Leal, 1996; National Coalition of Advocates for Students [NCAS], 1988; Reynolds, 1991). Bilingualism has either been revered or dishonored. Such arguments continue to profoundly impact education and legislation as well as public opinion and research contributing, as well, to the cultural milieu surrounding linguistic minorities. Since culture in turn is critical for the development of emotions (Matsumoto, 1993) and of self-esteem (Rosado, 1986), which, in turn, are known to enhance or deter second-language acquisition (Bialystok & Hakuta, 1994; Krashen, 1982) and academic success (Cummins, 1986, 1989; Rosado, 1986), it then follows that societal messages impact the ultimate adjustment of linguistic minorities (Malgady, Rogler, & Costantino, 1990; Quintana, 1995). In this study, the effect of environmental stress and language proficiency or measures of intelligence and anxiety will be explored.
Immigrant children, in general, and language minority children, in specific, comprise an increasing number of the United States population (NCAS, 1988). These children are part of the latest immigration wave that began in the late 1960s and that, by 1980, resulted in a national population of about 245 million. Of this population, 12% were African-American, 6.4% Latino-American, 1.6% Asian-American, and 80% Caucasian. By the year 2020, more than one third of the United States population will be non-Caucasian. Moreover, by the year 2050, the population is projected to grow to 300 million and to consist of 16% African-American, 15% Latino-American, and 10% Asian-American.

The population of language minority children is also on the increase: language minority children are those reared in a home in which a language other than English is used or whose first language is not English, using the definition of Chamot and O'Malley (1994). By mid-70s, 12.5% or twenty-eight million of the United States population had a language other than English (Ferguson & Heath, 1981), with the most commonly used languages being the European languages (Spanish, German, Italian, French, Greek, and Portuguese) and the Asian languages (Chinese languages, the Filipino languages, Japanese and Korean). The United States (Ferguson & Heath, 1981), in addition, has a substantial number of speakers of creole languages (Gullah, Louisiana Créole, and Hawaiian Creole), Black English Vernacular,
American Indian languages (approximately 200 different languages), Jewish languages (Yiddish, Hebrew, and Dzhidezmo), as well as Slavic languages (Polish, Czech, Slovak, Serbo-Croatian, Slovenian, Russian, and Ukrainian, among others). By 1990, among individuals older than 5, 14% of the United States population spoke a language other than English at home, representing 32 million individuals, according to Bialystok and Hakuta (1994), who also note that, in the 1990 Census, "virtually all languages of the world are represented" (p. 191). At present, 15% to 20% of all school children speak a language other than English at home (Geisinger & Carlson, 1992).

Statement of the Problem

Recognition of the educational and mental health needs of language minority children has been controversial. On the one hand, proponents of the United States as a pluralistic society value cultural and linguistic diversity as a national resource. Bilingual and multicultural education and the English Plus movements exemplify such view (Crawford, 1989; Price, 1992). Proponents of the United States as a melting pot, on the other hand, consider cultural and linguistic diversity as a threat to national unity. Support for the English Only movement and for "cultural conservatism" (Crawford, 1989, p. 63) has intensified.
A thread that has run among social and political debates over bilingualism has been its impact on a child's development and, more specifically, on intelligence. Studies on the intellectual skills of immigrants date back to the beginning of the century (Gregory & Lee, 1986; Jensen, 1983; Sattler, 1990; Schonemann, 1983). Although early findings of the negative impact of bilingualism on intelligence have been refuted (Bialystok & Hakuta, 1994; Gonzalez, 1994; Hakuta, 1986), the view of the intellectual inferiority of language minorities is pervasive and reflected in the following quote from Herrnstein and Murray (1994): "the kernel of evidence that must also be acknowledged is that Latino ... immigrants are, at least in the short run, putting some downward pressure of the distribution of intelligence" (pp. 360-361).

**Significance of the Problem**

The impact of this bleak view of bilingual intelligence has been far from academic: persistent low scores on intellectual tests given in English to language minority children and persistent cultural insensitivity during testing (National Coalition of Advocates for Students [NCAS], 1988) have led to an over-identification of immigrant children as disabled (Crawford, 1989; Cummins, 1986, 1989; Durán, 1989; Figueroa, Fradl, & Correa, 1989; Geisinger & Carlson, 1992). Moreover, such testing
practices have also resulted in an under-representation of language minority children in gifted and talented programs (Constantino, 1992) and in higher education (Pennock-Román, 1992).

A great number of treatises and research studies emerged in the 70s and the 80s which explored testing bias (Cohen, Parmalee, Beckwith & Sigman, 1986; Elliot, 1988; Fourguean, 1987; Gutkin & Reynolds, 1980; Mishra, 1983a, 1983b, 1984; Oakland & Feigenbaum, 1979; Prewitt Diaz & Rivera, 1989; Reschly & Reschly, 1979; Reynolds, 1983; Reynolds & Gutkin, 1980; Sandoval, 1979; Sattler, 1990; Taylor, Ziegler, & Parteno, 1984; Zimmerman & Woo-Sam, 1972). This research, conducted in the United States with Spanish- and English-speaking Latino children, revealed that their obtained low overall IQ scores were associated with low Verbal IQ scores. A consistent pattern of significant discrepancies between verbal and non-verbal skills, with below average verbal scores in the context of near-average performance scores, emerged. The extent of the Verbal IQ score and Performance IQ score discrepancy (V/P discrepancy) obtained by Latino children has been found to be greater and more frequent than the discrepancy obtained by monolingual English-speaking children. Despite the lack of empirical explanation about this discrepancy, comparative studies of intelligence have, for the most part, ascertained that the psychometric properties of current intellectual tests
present no inherent source of bias for Latinos. The possibility that such persistent V/P discrepancy is an artifact of conceptually misconstrued or inappropriately used tests continues to be disregarded.

Albeit subtle, the findings of these research studies have promoted the view that bilingualism leads to low verbal intelligence. The scant acknowledgment of other factors impacting test scores, such as cultural bias, SES factors, and test bias, led to a series of recommendations in the assessment of language minority children (Sattler, 1990), such as changes in administration, the use of test translations and of translators, adjusting the IQ scores, using adaptive measures, using criterion reference testing, and the use of non-verbal tests. Tests themselves, however, have remained unchanged and the value of intellectual test scores for language minority children remains, for the most part, unchallenged.

Significance of the Study

The aim of the present study is two-fold. First of all, language proficiency will be incorporated. Thus far, empirical assessment of the role of two languages in the processing of cognitive tasks given in a child's second language, English for our purposes, has been omitted in the field of psychometrics described above. At the most basic level, sample selection has been critically flawed: Latino
children in these studies have been selected as long as their oral proficiency in English allows them to understand and respond to test directions. In this manner, the Latino sample has been assumed to be homogeneous and representative of the general population of Latino children. However, research in linguistics and bilingual cognition has long documented that oral proficiency in a second language (also referred to as Language 2 or L2) is a poor indicator of whether the child utilizes such language for higher-order thinking or whether the child processes such information in his/her first or native language (also referred to as Language 1 or L1) or in a combination of L1 and L2 (Cummins, 1989). At a deeper level, then, a critical question is not whether a child speaks the language of the test but whether the child thinks in the language of the test (Bialystok & Hakuta, 1994). And even such question is simplistic: bilingual children not only speak in two languages but also think in two languages. Again, research on bi-cognitive development has indicated that bilingual children fall along a complex continuum of proficiency in L1 and L2 and their relative mastery potentially impacts intellectual test scores. In addition, psycholinguistic research has begun to understand the level of second language development needed for the adequate processing of the complex information found in intellectual and academic tasks given in the child's second language (Chamot & O'Malley, 1994). Unless relative
language proficiency is included in empirical analyses of IQ scores for Latino children, the generalization of research results to the broad population of Latinos must be questioned.

Secondly, this study incorporates assessment of the emotional impact of evaluating intelligence through L2. On the one hand, empirical studies assessing the emotional impact of measuring a child's intelligence in his secondly acquired language are scarce (Murphy, 1990, is an exception). On the other hand, there is documentation of a relationship between language choice and the expression of emotion, although the nature of this relationship is inconclusive. Factors such as the qualities of L1 and L2, the order of acquisition of the two language, and the context, have been explored. Javier and Marcos (1989), found that stress influences which language a bilingual person utilizes to express himself or herself and that, in turn, the chosen language affects the intensity of the perceived stress. Guttfrand (1990), moreover, studied the effect of language use on the emotional experience of bilingual adults and concluded that, regardless of the level of proficiency in Spanish and English, most bilingual subjects preferred Spanish when expressing their emotions. The author thus documented that the qualities of the specific language, in this case Spanish, rather than the native language, had an impact on the emotional experience.
the greater the perceived distance, the less successful the learner will be in mastering the second language.

Do language minority children perceive a test given in L2 stressful? Is the test-taking situation, with its dependency on standardized administration procedures, traumatic for bilingual children because it constrains their cognitive and emotional resources? Are the scores obtained a reflection of the presumed measured skills or are they artifacts of the test administration itself?

In the present study, language proficiency and measures of anxiety were incorporated in the study of psychological test scores for Latino children. The testing environment was manipulated in terms of the incorporation or lack of cultural and linguistic sensitivity. With a better understanding of the role of linguistic and affective variables, then more valid and ethical assessment procedures could be developed. But the implications are more far-reaching: as Hakuta (1986) indicated, our understanding of the "wondrously complex phenomenon of bilingualism" (p. 240) would enhance our understanding of language, mind, and society.

Hypotheses

In this study, it was argued that the testing situation is a potentially stressful event for a language minority child. It deprives the child of utilizing the full range of
linguistic and cognitive repertoire. Also, it heightens the child's level of anxiety. The relative language proficiency of the child was expected to modulate this relationship. Measured language proficiency and the testing environment comprised the independent variables while scores on intellectual tests and measured anxiety as well as language shift encompassed the dependent variables.

**Hypothesis 1**

The discrepancy between measured verbal and performance abilities will vary according to measured relative language proficiency. Such discrepancy will be greater the greater the proficiency in Spanish versus English.

**Hypothesis 2**

The testing environment will influence the overall intellectual test scores of bilingual children: an empowering (supportive and bilingual) testing environment will lead to higher scores than those obtained in an assimilative (restrictive and monolingual) environment.

**Hypothesis 3**

The testing environment will influence measures of anxiety for bilingual children: an empowering (supportive and bilingual) testing environment will lead to lower scores
of anxiety than those obtained in an assimilative (restrictive and monolingual) environment.

**Hypothesis 4**

The testing environment will influence language use for bilingual children. Code switching (alternating use of Spanish and English) will be greater in an empowering (supportive and bilingual) versus an assimilative (restrictive and monolingual) environment.

**Definition of Terms**

**Verbal and Performance Skills**

The discrepancy between verbal and performance skills was defined as the relative difference between Verbal IQ scores and Performance IQ scores obtained from the Wechsler Intelligence Scale for Children - Third Edition (WISC-III). Overall intellectual test score was defined as the Full IQ score obtained from the WISC-III.

**Relative Language Proficiency**

Relative English and Spanish proficiency was defined as the comparison between the level of proficiency in English and in Spanish, based on the Woodcock-Muñoz Language Survey, English and Spanish forms (Woodcock & Muñoz-Sandoval, 1993).
of the subjects. Reichman (1997) also found that Hispanics, when questioned about stressful health issues, preferred to be interviewed in Spanish, indicating that Spanish was the language best suited to express their pain.

The idea that bilinguals switch between languages in stressful circumstances is far from new. Leopold (1978) had earlier pointed to language switching as a way to magnify the emotional content of a message. But the relationship between language choice and emotions is reciprocal; emotions also influence second language learning. Krashen (1982) reviewed the affective variables conducive to the acquisition of a second language: high motivation, self-confidence and a positive self-image, and low anxiety. Emotions, therefore, are critical in second language acquisition, as they are in most learning situations. As such, the emotional environment can foster or deter such learning because the environment can modulate affect. In her ethnographic research of self-image and cultural sensitive instruction, Abi-Mader (1990) alluded to the importance of enhancing cultural pride in the acquisition of a second language. Bialystok and Hakuta (1994) agree on the importance of the social context in the acquisition of a second language. In reviewing the acculturation model of second language acquisition, they note that the perceived social distance between the language learner and the second language community either enhances or deters acquisition:
Testing Environment

The testing environment was defined in terms of the instructions given to children prior to the standard administration of the WISC-III and is defined as follows: (a) Multi-cultural- children were encouraged and allowed to communicate and think in either or both languages; (b) Assimilative- children were instructed to communicate and think in English only; and (c) Neutral- no instructions regarding use of languages were provided.

Anxiety

Anxiety was defined as the scores from the State-Trait Anxiety Inventory for Children (Spielberger, Edwards, Montouri, & Lushene, 1970).

Language Shift

Code switching is defined as the number of times a child shifted from one language to the other during testing.

Limitations of the Study

One of the expected limitations of the study was the possible relationship between one of the independent variables, relative language proficiency, and one of the dependent variable, intellectual scores. A significant relationship between language proficiency and intellectual
scores could result from the shared variance contributed by
the child's true level of verbal intelligence. This
possibility would then preclude inferences about the
predictive power of the independent variable. This
difficulty was addressed by the nature of the sample: the
sample was randomly selected and assigned to treatment
conditions; it is therefore assumed that the level of innate
verbal intelligence was randomly distributed across the
sample. In addition, children who, due to a history of
academic, learning, and behavioral difficulties, were, at
any point, referred to special education services, were
excluded from the analysis.

A second limitation of the study was the possible
relationship between language proficiency, intellectual
scores, and anxiety scores, to a degree, on some dimension
which is not accounted for in the study, such as
socio-economic factors, gender, or age. Again, the random
assignment of the subjects to the treatment condition was
expected to evenly distribute potential sources of error.
Gender and age were included in the analyses.

A third limitation of this study was the power to
generalize these findings to the general population of
language minority children. The SES status of the present
sample was expected to be restricted as children came from
an urban community which is considered, mostly, of low SES.
Its SES homogeneity entailed both an advantage as well as a
disadvantage: on the one hand, factors associated with SES status were circumscribed, limiting their contribution to the interaction between the dependent and the independent variables. On the other hand, the generalizability of the results to the broad population of Latino Americans and of language minority children from a diverse SES background was beyond the scope of the present study.

Another limitation of this study was related to the psychometric properties of the intellectual tests used and, to a certain degree, the language proficiency tests used. Ample research (and a point of contention in this study) has suggested that the norms of intellectual tests are, in numerous ways, problematic for bilingual children. For example, the homogeneity of the Latino sample included in the normalization sample, and the extent to which the sample is representative of the general population of Latino children, is questionable. The dilemma of how, and if, to use the available norms represented a seemingly insurmountable double bind: on the one hand, how can we justify the continuing use of norms suspected to be biased? On the other hand, if we calculate descriptive statistics, including norms, for our sample alone; how do we draw inferences about the study? Of greater concern, how do we present the results in the context of current practices and research? Given such dilemma, one solution would be to both analyze test scores based on available norms and to
calculate descriptive statistics for the sample after the contribution of relative English proficiency is statistically removed from the test scores. In this manner, these two approaches can be compared and contrasted.

In summary, in the present study, language proficiency and measures of anxiety were incorporated in the study of psychological test scores for Latino children. The testing environment was manipulated in terms of the incorporation or lack of cultural and linguistic sensitivity.

A historical review of the intellectual assessment of bilingual children will be intertwined with a review of comparative psychometric studies of intelligence, with attention to the research conducted with the Wechsler scales and the Wechsler Intelligence Scale for Children-Revised. Linguistic research with bilingual children will also be reviewed. Finally, an integrated view of bilingualism and intelligence will be proposed with attention to the need to ensure sensitive assessment procedures.

It must be noted that there is great controversy in the literature as to which is the correct or preferred term to refer to Spanish-speaking populations in the United States (Jones-Correa & Leal, 1996). The present author would prefer the term "Latin American-Origin populations" but agrees with Jones-Correa and Leal to use the term "Latino". In the following review of literature, however, the original
labels utilized by the authors will be used and the term Latino will be used whenever appropriate.
Chapter II
REVIEW OF THE LITERATURE

The psychological assessment of linguistic minority children has represented a formidable challenge for educators and researchers in the United States for over 50 years (Figueroa, Sandoval, & Merino, 1984; Gregory & Lee, 1986; Jensen, 1983; Paivio, 1991; Fuente & Salazar, 1998; Sattler, 1982, 1990; Schonemann, 1983). Traditionally, Latino children who are speakers of Spanish and English, obtain lower than average scores on tests of intelligence given in English (Figueroa, 1989; Palij & Homel, 1987). Given the reliance on intellectual test scores for educational placement (Garcia, 1983; Ortiz & Yates, 1987), these low scores have been instrumental in the over-representation of Latino children in special education classes for the Mentally Retarded and, increasingly, in programs for the Learning Disabled.

Intelligence and Bilingualism: 1900 to 1970s

The controversy over the psychological testing of bilingual children has an extensive and at times convoluted history which dates back to the beginning of the 20th century, to the emergence of the first studies of
bilingualism and intelligence; this history is linked to social and political forces which often shaped its course. By the end of the 19th century (Crawford, 1989; Heath, 1981), the tradition of linguistic and cultural diversity, common since colonial America, gave way to fears of political separation within the United States, leading both to efforts to establish English as the national language and to efforts to restrict the maintenance and the teaching of other languages. Efforts to restrict naturalization and immigration also emerged. Hakuta (1986) indicated that the movement to restrict immigration coincided with the emergence, in Europe, of instruments to assess intelligence. Translated into English, these tests were used to assess the intelligence of immigrants and, in such manner, limit the entry of immigrants deemed of low ability. The misuse of intellectual tests for these linguistic minorities is poignantly described by Hakuta (1986) in the following description of the practices of the time:

In one study, Goddard (1917) took the English-language version of the Binet test to Ellis Island, the point of entry for newly arrived immigrants. In testing thirty adult Jews through an interpreter, he assessed twenty-five of them as "feeble-minded." (p. 19)
Genetically Inferior or Language Handicapped?

Cultural factors, limited English proficiency, and improper procedures were disregarded as possible explanation of the low scores obtained by the immigrants. From similar studies, the notion of the genetic inferiority of linguistic minorities emerged early in the century and persisted through the next four decades, in the context of an explosion of intellectual tests (see Sattler, 1982, for a historical review of intelligence tests).

Gradually, the theory of the presumed genetic weakness of linguistic minorities was challenged by proponents of the "language handicap" of bilinguals (Hakuta, 1986) and learning, not genes, became the culprit. This discussion, in turn, must be understood in the context of the nature versus nurture controversy which permeated psychology and the sciences during the first part of the century. The controversy focused on the immutability of genetic characteristics, on the one hand, and the absolute contribution of learning in human development, on the other hand (see LaBarba, 1981, for review of the nature-nurture controversy). Proponents of the language handicap theory, therefore, argued that bilinguals performed poorly on intellectual tests not due to their genetically inherited inferiority but because the learning of two languages, that is bilingualism itself, led to cognitive confusion and poor development of verbal skills (López-Laguerre, 1989).
By the 1960’s (Reynolds, 1991) it had become apparent that earlier studies on the negative impact of bilingualism on intelligence had been characterized by numerous methodological problems, such as sampling issues, definition of variables, questionable test instruments, poor statistical controls, and disregard of demographic variables. In 1962, however, a pivotal study was published by Peal and Lambert. The authors discovered that 10 year-old bilingual children performed significantly better than monolingual children on most cognitive tests given. The impact of such study, according to Reynolds (1991) was three-fold. First of all, it set the methodological standards that were lacking in prior studies and which were replicated in later studies. Secondly, it had a political impact on the issue of bilingual education, as it dispelled the notion that bilingualism was detrimental to the development of intelligence. Thirdly, it coincided with Vygotsky’s publications on the relationship of thought and language and his postulation that bilingualism allows a person to be aware of his or her own cognitive functioning (as cited in Bialystok and Hakuta, 1994; see also Reynolds, 1991).

Research on the relationship of bilingualism and intelligence and the mind, as well, continue to date (Bochner, 1996; Miranda & Valencia, 1997; Paivio, 1991;
Pearson, Fernandez, Lewedeg, & Oller, 1997; Vaid & Hall, 1991) and will be reviewed later.

**Psychometrics vs Linguistics**

It must be noted that the controversy over the intelligence of linguistic minorities and the relationship between bilingualism and intelligence followed, by the middle of the century, two complementary yet, unfortunately, often parallel paths: linguistics and psycholinguistics on the one hand and psychometrics and psychology on the other. While developments in linguistics from the 1940's to the 1960's had begun to point to the positive effects of bilingualism on intelligence, psychology had continued to focus, to a significant degree, on why bilingual children performed poorly in intellectual tests. An illustration of such analysis is provided by Holland (1960) in his study of Spanish-speaking children. Concerned that, despite the previous 40 years of psychometric development, linguistic and cultural factors continued to be disregarded in the intellectual assessment of Spanish-speaking children, Holland employed a bilingual administration of the Wechsler Intelligence Scale for Children (WISC). In this manner, Holland administered the WISC items in English and, if the child was unable to comprehend the question, the item was repeated in Spanish. He then computed the following formula
"Language Barrier - Bilingual Verbal IQ - English Verbal IQ" (p. 44). Holland's conclusion merit close examination:

1. A child has a language barrier when his knowledge of Spanish is greater than his knowledge of English. This condition originates in lack of acculturation.

2. There was an average language barrier of 4.6 IQ points per student. In eight cases the language barrier was very serious, in seven cases it was serious, and in 18 it was moderate. Only three students were completely free of language barrier....

3. Language barrier is a very important factor in the lower academic achievement of many Spanish-speaking children recommended for testing in the Tucson Public Schools.

4. Aside from language barrier, the sub-standard verbal development of Spanish-speaking children is probably the result of being bilingual (italics added) and having to forfeit a more thorough knowledge of one language for partial familiarity with two. (p. 50)

Reacting to these results, Holland also concluded that "Bilingual education for bilingual children might prove to be a worthwhile experiment" (p. 50). In spite of his foresight, Holland's study, unfortunately, demonstrated the
methodological and conceptualization flaws of the psychometric research of the time and the pervasive bleak view of bilingualism and intelligence.

**Bilingual Children and Special Education**

**Enrollment in Special Education Classes**

By the 1960's, consternation over the over-representation of ethnic minorities in programs for the mentally retarded, and in the context of the era's social and political interest in racial equality (Almanza & Mosley, 1980; Bersoff, 1992; Braden, 1987; Brandt, 1984; Cleary, Humphreys, Kendrick, & Wesman, 1975; Figueroa, 1989; Flaugher, 1978; Gordon & Terrell, 1981; Prasse & Reschly, 1980), the potential bias of intellectual tests and procedures became under heated debate. This debate also occurred in the context of increased concerns about the rights and needs of linguistic minorities in education.

**Bilingual Education Act**

Such concerns led to the eventual enactment of the Bilingual Education Act, in 1968, a step in which, according to Crawford (1989) "the US government signaled its first commitment to addressing the needs of students with limited English skills" (p. 32).
The history of bilingual education, in itself intricately intertwined with social and political forces, is beyond the scope of this presentation (for a review, refer to Crawford, 1989). By the seventies, it is important to note, however, that the special educational needs of linguistic minorities fell to intensive scrutiny, including the inadequacy of traditional assessment and intervention practices; practices considered, even two decades later (Cummins, 1989; Durán, 1989; Figueroa, 1989; Figueroa, Fradd, & Correa, 1989; Valencia, 1984; Vance, Hankins, & Brown, 1988; Vance, Huelsman, & Wherry, 1976; Vázquez Nuttall, 1987) unethical and not in compliance with federal mandates. A review of The Education for All Handicapped Children Act (1975), also known as Public Law 94-142, is essential; it stipulated that children should receive suitable education in the least restrictive environment. Regarding linguistic minorities, the act also stipulated that testing materials or procedures need to be administered in the child's native language or mode of communication. Such directives are repeated in the latest revision of P.L. 94-142, now known as Individuals with Disabilities Education Act (IDEA) or Public Law 105-17 (1997).
Impact on Assessment Practices

However, as indicated by Figueroa (1989), such mandates did little to acknowledge the unique needs of language-minority children. Specific directives, such as the definition of native language or the nature of materials and procedures, were not delineated. Thus, without these directives, it was left to the individual states to interpret the federal law. Few states, to date, have established guidelines. Current procedures, therefore, are believed to often isolate Latino children in special programs they do not need or, paradoxically, exclude them from needed services (Cummins, 1989).

Responding to the increased criticism of intellectual tests, the psychological research on intelligence in the United States, for the past two decades, has focused on comparative studies of minorities within the United States. A review of the research conducted with the Wechsler Intelligence Scale for Children-Revised follows.

Intelligence and Bilingualism: 1970s to Present

Studies with the WISC

The Wechsler Intelligence Scale for Children-Revised (WISC-R), developed in 1974 (Wechsler, 1974) is the revision of the Wechsler Intelligence Scale for Children (WISC) that was developed originally in 1949 (Sattler, 1982). Like the
WISC, the WISC-R, for ages 6 to 16, is devoted to assess the cognitive skills presumed to be essential for effective and analytical thought and behavior. Twelve sub-tests, organized into two sets, consisting of verbal tasks and perceptual-motor tasks, yield a Verbal IQ score and a Performance IQ score; the aggregate of the sub-test scores yields a Full Scale IQ score. The Verbal Scale has been presumed to depend on a child's accumulated experience and to reflect a child's level of verbal ability and crystallized experience. The Performance Scale, in turn, has been postulated to depend on the child’s current problem-solving skills and to reveal nonverbal ability and fluid intelligence. The WISC-R has been amply used in research and in comparative studies of intelligence. Of note, a comprehensive review of the WISC properties and of WISC research (Zimmerman & Woo-Sam, 1972), for the decade of 1960 to 1970, had raised concerns over the inadequate representation of bilingual children in the standardization sample of the WISC and the disregard of bilingualism in empirical research with bilingual children. Despite such criticism, it is important to note that the standardization group of the WISC-R included children categorized only as White and Non-white (Kaufman, 1976). In addition, only children that were able to speak and understand English were included in the sample (Wechsler, 1974). Dean (1979) noted that the sample included only 2% of Mexican American
children, although they constituted a population of 5 million.

In the past two decades, comparative research conducted with the WISC-R has attempted to empirically analyze the psychometric properties of the scale and the pattern of verbal and performance skills. Although a consistent and significant Verbal IQ score/Performance IQ score discrepancy (V/P discrepancy) has been documented for Latino children, with lower Verbal IQ scores in the context of average Performance IQ scores, reliability studies and validity studies, in general, have reported adequate psychometric properties.

**Verbal/Performance IQ discrepancy**

The research conducted in the United States with the WISC-R indicates that Latino children commonly obtain unusual statistical discrepancies between verbal and non-verbal skills. The nature of such discrepancy must be analyzed in the context of the average V/P discrepancy score for children in the standardization sample of the WISC-R. In general, a 12-point discrepancy between Verbal and Performance IQ scores has been determined to be significant at a confidence level of .05 whereas a 15-point discrepancy has been determined to be significant at a confidence level of .01 (Sattler, 1982, 1990). Kaufman (1976) has also indicated that the average V/P discrepancy for the
standardization sample of the WISC-R was of 9.7 and that 25% of the sample had V/P discrepancies of 15 or more points.

WISC-R research has indicated that large V/P discrepancies are commonly found in Latino samples. Sattler (1990), quoting research conducted between 1971 and 1980, notes that "Studies almost universally report that Hispanic-American children obtain higher Performance IQs than Verbal IQs" (p. 586). Sattler also indicated that such discrepancy "is probably associated with the language difficulties [italics added] experienced by Hispanic-American children" (p. 586). Denotation of bilingualism as an example of a cultural handicap is also seen in Gilger and Geary (1985).

More recent research has also corroborated large V/P discrepancies for Latino children: Cohen, Parmalee, Beckett, and Sigman (1986) studied the extent to which biological and environmental factors predicted cognitive development of pre-term infants, from English-speaking and Spanish-speaking homes, at age 8. The results indicated several differences between the predictive value of identified factors for both groups which led to the conclusion that "In the Spanish language group, we have been less successful in identifying factors associated with recovery" (p. 109). One striking finding was that the average IQ scores obtained by the two groups: the English group obtained an Verbal IQ score of 109.1 (SD=15.3), a
Performance IQ score of 106.4 (SD=17.1), and a Full IQ score of 108.8 (SD=16.3); the Spanish group obtained a Verbal IQ score of 80.1 (SD=11.9), a Performance IQ score of 103.5 (SD=8.7), and a Full IQ score of 94.7 (SD=9.6). The difference between the V/P discrepancy for the English group, equal to 2.7 points, and the Spanish group, equal to 15.4, was found to be significant at the .05 level. The authors noted that "The lower verbal intelligence scores of the children from the Spanish-speaking families parallels that found in Hispanic full-term children of lower socioeconomic status and is not a function of pre-term birth" (p. 108-109).

Furthermore, Taylor, Ziegler, and Partenio (1984) investigated the WISC-R Verbal-Performance IQ discrepancies for African American, Latino, and White children, ages 6 to 11. The sample consisted of a stratified sample of 555 children, of which 184 were Latino. The authors found a mean V-P discrepancy of 8.91 for African American children, of 9.63 for White children, and of 13.72 for Latino children; the discrepancy found for the Latino children was significantly larger than the one obtained by the other two groups. Such finding is even more significant given that all the groups, as reported by the authors, obtained higher mean IQ than the ones reported in the standardization sample. The Latino sample, in this fashion, obtained a mean VIQ score of 95.15 (SD=18.18), a PIQ score of 105.67.
(SD=14.86), and a FIQ score of 100.39 (SD=16.05). However, approximately 50% of the Latino sample obtained a V-P discrepancy of 12 or more points. Finally, the authors noted that the Latino group obtained a V-P discrepancy that was 4 points larger than the average discrepancy obtained by the standardization sample of the WISC-R.

The literature on referred children has also revealed large V/P discrepancies: an average V/P discrepancy of 21.34 was documented by Fourqurean (1987) for "Latino" (p. 16) children; Mishra (1984) obtained a 15.18 V/P discrepancy for "Mexican-American children" (p. 1486); and a study by Reynolds and Gutkin (1980) revealed a V/P discrepancy of 13.26 for "Chicanos" (p. 239). An exception is noted in a study by Gutkin and Reynolds (1980) which revealed a V/P discrepancy of 2.7 for "Chicanos" (p. 35).

Reliability and validity studies. The definitions of test bias, according to Flaugher (1978), are many. One is test bias as mean differences between groups and, another, the existence of differential validity for different groups. These definitions of bias have influenced the analysis of potential sources of bias in the WISC-R. For the past two decades, consequently, researchers have focused on whether the psychometric properties of the WISC-R are similar for children of various ethnic background (Sandoval, Zimmerman, & Woo-Sam, 1983).
Discussed in detail in the following sections, such research has found little indication of a systematic error of measurement. However, sampling errors limit the validity of such conclusions and will be discussed in later sections. Oakland and Feigenbaum (1979) analyzed the results obtained by children from various ethnic backgrounds on the WISC-R. Several statistical procedures were utilized to assess internal consistency, namely, item difficulty, correlations of item-total correlations, concurrent validity (with the California Achievement Test), and construct validity. The sample included 180 White children, 119 African American children, and 137 Mexican American children, ages 7 to 14. No indication of SES background was given. The results indicated that with few exceptions, no consistent source of test bias could be documented. One exception related to the significant difference in item difficulty of the WISC-R for the Mexican Americans when compared to the White sample.

Sandoval (1979) compared the internal performance of children from various ethnic groups on the WISC-R. The sample included 1,050 children, representing a random sub-sample of the standardization sample for the System of Multiculture Pluralistic Assessment (SOMPA). The children ranged from 5 to 11 years and were equally distributed among three racial groups: White, African American and Mexican American. The author noted that "Children in the Mexican-American group who could not understand English were
dropped from the analysis" (p. 921). Analyses of alpha reliability, rank-order correlations and multivariate analysis of variance pointed to essentially equal reliability indexes across the three groups. However, the Vocabulary subtest was found to be significantly more difficult for Mexican American children.

Mishra (1983) examined the validity of the IQ scores and the factor scores from the WISC-R in their power to predict academic achievement on the Wide Range Achievement Test (WRAT). The sample consisted of 64 Mexican American children from low socioeconomic backgrounds, mean age of 10.13 (SD = 3.24). The children "... predominantly spoke Spanish at home as well in their conversation with friends and peers" (p. 443). The results revealed low correlation coefficients between the WISC-R factor scores and achievement scores. The WISC-R IQ scores were found to have low to moderate relationship with WRAT scores.

Gutkin and Reynolds (1980) studied the factorial similarities of the WISC-R for White and Mexican American children. The sample consisted of 142 Chicano American children and 78 White children, with a mean age of 10-6 (SD = 2-6), from a referral population and were mostly from a lower-middle class background. Separate principal factor analysis were conducted for each ethnic group and then both compared via coefficients of congruence. The authors reported a similar two-factor solution for both racial
groups; this is, the results supported the distinction
between Verbal Comprehension and Perceptual Organization
factors for both groups. Of note, support for the third
factor, Freedom from Distractibility, was reported to be
questionable. Limitations of this study, relating to the
sample size and the homogeneity of the sample, are worth
noting. The sample include a small sample size and unequal
sample sizes for each group. As Tabachnick and Fidell
(1983) indicate, for a factorial analysis, a sample size
below 100 is poor and one between 100-200 is questionable.
A sample size of such magnitude is adequate when subjects
are homogeneous and the number of variables is small. The
homogeneity of the Chicano American sample is questionable
given a possible wide variation of level of bilingualism, a
likely source of variance for the sample.

Reynolds and Gutkin (1980) examined the validity of the
WISC-R to predict academic achievement on the Wide Range
Achievement. The sample consisted of 174 Chicano American
and 94 White children, mean age of 11-1 (SD = 2-9) who had
been referred for special education services. The children
were mostly from lower-middle-class backgrounds. Regression
equations for the prediction of achievement (Reading,
Spelling and Arithmetic), for each group, were then
compared. With one exception, the regression lines were
comparable for both group of children. The authors
concluded that the WISC-R predicted academic scores independently of race.

Lewis and Lorentz (1994) compared the Leiter International Performance Scale and the WISC-R and the Wechsler Preschool and Primary Scale of Intelligence-Revised, for African-American and Latino children and concluded that the former may provide a more valid assessment of the Latino children's intelligence.

Recent research with the WISC-III and ethnic-minorities is scant (Puente & Salazar, 1998). Glutting, Oakland, and Konold (1993), however, explored the relationship between the Guide to the Assessment of Test-Session Behavior for the WISC-III and WIAT (GATSb) and the WISC-III. In general, higher problematic behaviors were associated with lower IQ scores, a relationship that was expected and that held true for Anglo, Black, and Latino children. For Latinos, however, this relationship held true only for average and above average IQs; at lower IQs, Latino children were rated as displaying better test behaviors than Anglo children.

Of further note, there was a discrepancy of 9.1 points between their VIQ ($\bar{M} = 88.9$) and their PIQ ($\bar{M} = 98$). Although not statistically significant (according to the values reported by Kaufman, 1994), the V-P discrepancy follows the trend of previous research, when compared to the V-P discrepancy of 1.7 for the Anglo sample and of -.6 for Black children.
In general, researchers have concluded that the WISC-R has similar psychometric properties across ethnic groups and little evidence of systematic test bias. However, the unusual and significant V/P discrepancy scores obtained by Spanish-speaking children continues to raise concerns about the validity of such scores. This discrepancy has emerged, as well, in WISC-III studies.

**Significance of the V/P discrepancy.** Significant V/P discrepancies (between 11-14 points at the .05 significance level and over 15 points at the .01 significance level, for all ages), traditionally, have been considered of clinical importance. Kaufman (1979; 1994) indicates that, although research on the significance of IQ discrepancy has been inconclusive and controversial, large V/P discrepancies have been associated with organic impairment, asymmetrical left- and right-brain functioning, learning disabilities, and psycho-linguistic deficiencies, among others. Sattler (1990) also indicates that several hypothesis can be raised when Performance IQ scores are significantly larger than Verbal IQ scores, including the following:

a. Performance skills are better developed than verbal skills.

b. Visual-motor discrimination skills are better developed than auditory-vocal processing skills.
c. Immediate problem-solving ability is better developed than knowledge acquired as a result of accumulated experience.

d. The examinee may have difficulty with reading and academic achievement.

e. A language deficit may exist.

f. Limitations in auditory conceptual skills and auditory processing skills may be influencing performance.

g. Difficulties may be experienced in working effectively without time pressure. (p. 173)

Empirical research on the nature and meaning of the V/P discrepancy for Latino children is almost non-existent, an unfortunate consequence of present comparative studies of intelligence. As previously discussed, these studies have, in general, supported the psychometric value of the WISC-R for Latino children; consequently, the significant V/P discrepancy has been adopted, in the WISC-R literature, as a valid representation of Latino children intellectual attribute. Wariness about the V/P discrepancy, in all fairness, has promoted the development of alternative tests and procedures (Figueroa, 1989; Sattler, 1990) but has produced no empirically-supported guidelines for the understanding and use of the V/P discrepancy.

More recently, Kaufman (1994), in his review of the Wechsler Intelligence Scale for Children-III (WISC-III),
revisited the existence and the significance of the V/P discrepancy for bilingual children. Hispanic children as well as American Indian children continue to exhibit a consistent discrepancy of 10 to 15 points, in favor of Performance IQ over Verbal IQ. In his review, Kaufman acknowledges that empirical explanations for such discrepancy are still lacking and he cautions against the use of Full IQ scores.

Limitations of Wechsler Scales Research

In summary, intellectual tests have been found psychometrically reliable, on the one hand, but still reflect an unequal pattern of scores. How could we reconcile such findings? One answer is that most studies have failed to incorporate the construct of bilingualism, and the related concept of language proficiency, in studies evaluating the obtained IQ scores of bilingual children on the Wechsler scales. This omission is a potential source of procedural and theoretical flaws.

Procedural errors. In the study of test bias, most researchers have focused on the reliability and validity of the WISC-R for Latino children vis-à-vis White and African American children. However, most have confounded the issue of bilingualism with ethnic group membership (Mishra, 1983; Oakland and Feigenbaum, 1979; Reynolds & Gurkin, 1980; Taylor & Ziegler, 1987; Taylor, Ziegler, & Partenio, 1984).
Language proficiency for the Latino children in these studies, variously described as Mexican American, Chicanos, or Hispanics, is usually not reported. Furthermore, the variety of demographic labels used for the Spanish-speaking population is in itself problematic, because it is assumed that all of these groups are homogeneous, a belief challenged, among others, by Jones-Correa and Leal (1996) and Puente and Salazar (1998). In specific, four procedural limitations are noted as follows:

1. The extent to which Spanish-speaking groups represent a single ethnic category with shared cultures and histories. According to Jones-Correa and Leal (1996), Spanish-speaking people living in the United States prefer to identify themselves in national-origin terms (that is Puerto Ricans, Cuban, and so forth). Panethnic identifiers (such as Latino, Hispanic, Latinoamerican, among others), which are larger groupings than national origins, are controversial. According to the authors, these terms are "a creation" (p. 216) of state and federal agencies to describe the diversity of Spanish-speaking people in the United States. There is further discussion of which is the "correct" or the preferred term. What is known, however, is that panethnic identification varies with national-origin, generation, age, and education, and that it may have little effect on perceived cultural commonalities. In terms of
research, therefore, such cultural variation might have been disregarded or misinterpreted.

2. The extent to which the sample is randomized on the factor of language. Bilingualism is a complex entity referring to the development of two language systems. The interplay of two language systems results in a continuum of relative language proficiency. Woodcock and Muñoz-Sandoval (1993), for example, define a "Comparative Language Index" (p. 38) as the proficiency in Spanish versus the proficiency in English. Before determining such a ratio, however, the proficiency in each language must be assessed. In each language, the proficiency can range as follows: Advanced, Fluent to Advanced, Fluent, Limited to Fluent, Limited, Very Limited to Limited, Very Limited, Negligible to Very Limited, and Negligible. As such, the language proficiency of the Latino sample in comparative studies of intelligence could be quite homogeneous or heterogeneous. Without such information, the extent to which a sample is randomized must be questioned. Both potential sources of error and the generalizability of the results to the general population of bilingual children are suspect.

3. The extent to which a particular sample is comparable to the norm sample. It was previously mentioned that the normalization sample of the WISC-R included only bilingual children believed to speak and understand English (Wechsler, 1974). There is no information about the norm
sample's relative language proficiency: were these bilingual children, using Woodcock and Muñoz-Sandoval's definition, Negligible Spanish speakers and Advanced English speakers or Advanced Spanish and English speakers or Negligible speakers of both? Without such information, descriptive and inferential statistics for a particular sample should be questioned.

4. The extent to which linguistic factors are involved in the assessment of bilingual children's intelligence. Relevant to this discussion is a study by Braden (1984) who studied whether language differences, between standard American English and non-standard dialects, could explain IQ differences between White and African American children. As such, he compared the discrepancy in Performance IQ (PIQ) scores on the WISC-R of White and deaf children, who are, presumably, isolated from a linguistic environment. He then compared such differences against existing PIQ differences between White and African American children. He found the latter twice as large than the former and concluded that the low scores obtained by African American children could not be a result of linguistic differences. But, how was the WISC-R administered to the deaf children? Were the children lip readers and as such fluent in Standard English? Or did the evaluators mimic the instructions and, in this manner, the administration of the test was non-standard? Were both the evaluators and the deaf
therefore, the test given in the language of the child not in standard English? These procedures, not described by Braden in his article, then invalidate the claim that linguistic factors are not involved in IQ differences among groups. The interplay of subtle linguistic factors in the assessment of bilingual children’s intelligence must be carefully considered.

In summary, most researchers have confused the factors of bilingualism and ethnicity. Another confounding factor which may have also been overlooked in such research is the interplay of bilingualism, ethnicity and gender on intellectual test scores. Ethnicity and gender, on the one hand, have been explored (Johnson, 1994; Sellers, 1990; Wessel & Potter, 1994; Wilkinson, 1993). In general, the effect of gender has been small. In his review, Kamphaus (1993) notes that boys outperformed girls on most Performance subtests, with the exception of Coding, in which girls obtain higher scores. In general, however, the author indicates that these gender differences are found in studies with large sample sizes and may be statistical, not clinically, relevant.

The effects of bilingualism, ethnicity and gender on intellectual test scores, on the other hand, however, have not been explored in most studies of test bias conducted in the 1980’s (Cohen, Parmalee, Beckwith, & Sigman, 1986; Fourqurean, 1987; Gutkin & Reynolds, 1980; Mishra, 1984;
Reynolds & Gutkin, 1980; Taylor, Ziegler, & Partenio, 1984). In contrast, Boulon-Diaz (1992) examined the effects of intelligence and gender, among other factors, on the school achievement of Puerto Rican Children. She found that parental characteristics and social factors had an effect on boys’ IQ and Grade Point Average, a relationship not found for girls. Other research studies conducted in the 1990’s (Glutting, Oakland, & Konold, 1993; Lewis & Lorentz, 1994) have likewise included gender and ethnicity in the study of intelligence, but, however, have not reported the interaction of gender and ethnicity on intelligence. Furthermore, the interaction of gender and bilingualism and its effect on intelligence has yet to be explored.

**Theoretical flaws.** In addition to procedural errors, most validity studies have been based on flawed theoretical assumptions. The assumption has been that, as long as the child understands English and expresses himself or herself in English, then the child is comparable to the norm group and the evaluation, in English, of his or her intelligence is appropriate. Or, as troubling, is the assumption that if a child is more proficient in Spanish, a test translation should be used. Two problems arise with this practice:

1. **Monolingual assessment of a bilingual child’s intelligence.** This practice neglects the body of literature, to be discussed in the next section, which indicates that thought and behavior are different in the
indicates that thought and behavior are different in the monolingual mind versus the bilingual mind (Alatis & Staczek, 1985; Bialystok & Hakuta, 1994; Gonzalez, 1994; Hakuta, 1986). According to research, bilingual children may have developed cognitive strengths in their native language, their second language, or both. Therefore, a test given exclusively in one of their languages, does not fully assess their cognitive development. Ideally, the best intellectual assessment for these children would be a tool, simultaneously given in the languages spoken by the child, that could identify which cognitive skills are shared by the two languages and what is particular to either language.

2. Reliance on test translations. Although a widely used intellectual test such as the Wechsler scales have been translated in several languages, among others, Cantonese, French, Spanish (Chan, 1984; Dague, 1982, 1983, Wechsler, 1992), Dutch, and German (for a review, refer to Kroonenberg & ten Berge, 1987), appropriate norms have often not been developed and they have been found controversial (Rousey, 1990; Tamayo, 1990). For example, Prewitt Diaz, Rodriguez, and Rivera Ruiz (1986) studied the 1982 translation of the WISC-R, the Escala de Inteligencia Wechsler para Niños-Revisada (EIIWN-R), for which there are no norms developed. When compared to the USA standardization sample, the Full IQ mean of their sample was lower. The sample of Puerto Rican children living in Puerto Rico obtained a mean
obtained a mean IQ score of 85.2. Norms for Puerto Rico were, furthermore, not developed until 1992 (Wechsler, 1992) for the WISC-R (named the WISC-R-PR). In the meantime, however, a third revision of the WISC test emerged, the Wechsler Intelligence Scale for Children—Third Edition (WISC-III) (Wechsler, 1991) for which there is no available translation or norms for Spanish-speakers.

**Linguistic Research**

The above discussion addressed the disparate findings regarding tests for Spanish-speaking children. On the one hand, no systematic source of bias has been documented. However, on the other hand, a persistent V-P discrepancy defies explanation. Procedural and theoretical flaws, specially linguistic factors, were offered as a likely explanation for the controversial findings. The following section is offered as elaboration of the impact of linguistic factors in testing.

While psychological and psychometric studies explored the potential bias in psychological tests, such as the WISC-R, for linguistic and ethnic minorities, linguistic and educational researchers were interested in the process involving the development of two languages such as the factors associated in the learning of a first language vis-à-vis a second language (Bialystok & Hakuta, 1994; Hakuta, 1986), simultaneous acquisition of two languages
versus second language acquisition (Hatch, 1978), the effect of the self and culture in second language learning (Bialystok & Hakuta, 1994), and the effect of two languages on adjustment and intelligence (McLaughlin, 1977; Rosado, 1986). For a review, refer to Reynolds (1991). In this context, the definition of what is bilingualism is critical, because its definition has been often both misunderstood and confusing.

**Bilingualism Undefined?**

This confusion arises partly from the fact that bilingualism is a relative concept. In general, it is a complex entity referring to the development of two language systems and speakers of two languages fall somewhere in that continuum at a given point in time. This continuum is far from linear or static because, as in monolingualism, several language components are involved, including form, content and use. Speakers of two languages might attain native-like proficiency in some or all of these language components in either, neither, or both languages. As Bialystok and Hakuta (1994) note, furthermore, "language is ultimately a system of knowledge that is represented in the mind" (p. 212). For second-language learner, some of this knowledge is like the knowledge they already have on their first language but some is unlike the knowledge on the first language which must, therefore, be learned. This development and learning could
occur at different times and at different rates and second-language learners might show different degree of performance and competence in different settings. Hence bilingualism is a relative concept and its definition constrained by the setting and circumstances in which it occurs.

Factors impacting on second language acquisition. Differential development in the two languages spoken by a person is likely and their relationship is expected to vary over time. Such differential development is associated with a complex host of intertwined factors. Hakuta (1986) and Krashen (1992) have delineated several of these factors as follows: (a) individual variations such as intelligence and language aptitude, personality, attitude, the individual context in which the person uses the language, and affective variables such as motivation, self-confidence and anxiety; (b) linguistic factors such as the level of proficiency in the first language and the influence of the native language, that is the similarity between both languages; (c) exposure to the second language such as the length of residence in the second language environment, the language of the parents, the community and the school; (d) age; (e) societal factors such as migration and acculturation, the relative prestige of the two languages spoken by the child, and the type of societal bilingualism, that is subtractive or additive bilingualism. In their extensive review of factors
impacting on second language acquisition, however, Bialystok and Hakuta (1994) indicated that factors such as these are likely to constraint second language acquisition but do not preclude it. What seems to be important is the match between the particular language learner and the circumstances in which this learning occurs. In a sense, then, the cultural context in which a second language is learned is critical in the process.

**Bilingualism and Intelligence**

What is then the relationship between bilingualism and intelligence? Empirical studies on the impact of bilingualism on intelligence must be critically analyzed. It has been previously argued in the present study, that early findings of the negative impact of bilingualism on intelligence have been refuted (Bialystok & Hakuta, 1994; Gonzalez, 1994; Hakuta, 1986; Mohanty & Babu, 1983). However, as with comparative studies of intelligence, research involving the relationship between bilingualism and intelligence must be carefully reviewed for methodological flaws. One seemingly insurmountable problem is the appropriateness of the various tests used in these studies, including measures of bilingualism and intelligence. For example, Hakuta (1987) analyzed whether bilingualism was positively related to cognitive ability, defined as metalinguistic ability or the ability to objectively analyze
linguistic information (p. 13775). He found no such relationship. Problems with the tests administered are worth noting. For example, to measure bilingualism, Hakuta utilized the English Peabody Picture Vocabulary Test and a Spanish translation, which the author noted was based on monolingual English speakers. Given the lack of norms for this translation, its validity must be questioned, despite the fact that Hakuta validated the tests against other measures of English and Spanish. In addition, to measure cognitive ability, Hakuta utilized the Raven's Coloured Progressive Matrices Test and the Thurstone's Primary Mental Abilities Test. The author noted that this testing was conducted in Spanish which raises the question of how such alteration of standard administration procedures impact on the test results. Similar difficulties are found in the study by Lemmon and Goggin (1989). In comparing the performance of bilingual versus monolingual university students on cognitive ability, the authors found that the monolingual subjects had higher scores than bilingual subjects on cognitive skills. One test utilized to assess cognitive skills, that is, verbal concept formation, was the Similarity subtest of the Wechsler Adult Intelligence Scale-Revised. Again, the lack of norms for this test undermines the interpretation of results. Along these lines, Reynolds (1991), remains skeptical as to whether there is a causal relationship between bilingualism and intellectual
performance. Although he notes that most studies since 1962 have shown the bilinguals to be superior than monolinguals on cognitive tasks, he questions the methodological adequacy of most of the studies.

On the other hand, Gonzalez (1994) surmounted such methodological flaws both with a multi-faceted assessment of bilingualism and by the definition of cognitive skills. As such, bilingualism was defined by quantitative and qualitative indexes, including the IDEA Oral Language Proficiency Test, the child’s first and second language proficiency rated by the parents, and two informal measures, a Teacher’s Rating Scale and a Home Language Survey (p. 403). These scores were then converted to a 5-point scale. Cognitive skills were defined by the author as conceptual development defined by performance on a classification task which included labeling, defining, sorting, verbal justification for sorting, and category clue (p. 405). This performance was, in addition, classified into five developmental stages, based on Piaget’s stages of cognitive development. The results indicated that bilingual children obtained higher scores on nonverbal that in verbal classification tasks in both languages—and that, in both languages, their performance was at or above chronological and developmental age. More recent studies have compared the short-term memory of bilinguals on Spanish and English tests (Miranda & Valencia, 1997), the development of
vocabulary by bilingual infants (Pearson, Fernandez, Lewedeg, & Oller, 1997), and the learning strategies of bilinguals versus monolingual students (Bochner, 1996).

**Bilingualism and the bilingual mind.** From the above discussion, it is clear that the debate of whether bilingualism fosters or limits intelligence continues. Solid methodology and theory will continue to clarify the dilemma. In the meantime, a comparable area of research has addressed, not the quantitative relationship between bilingualism and intelligence, but their qualitative relationship.

In order to understand how are two languages represented in the bilingual mind, an understanding of language and thought is imperative. In itself, the study of language and thought has a long tradition which has been reviewed elsewhere (Bickerton, 1995; Bialystok & Hakuta, 1994; Segalowitz, 1977). This body of study has pointed to language as an innate, biological function unique to humans. It not only performs a social and communicative function but has an impact on our internal communication. According to Bickerton (1995), language frees our thoughts from external, here-and-now demands and allows us to ponder over past and future events or events remote in space. And even more powerful, language allows us to be aware of our own awareness, defining, therefore, our consciousness.
All of the world languages, although each defined by a different subset of all possible human speech sounds, which are in themselves a subset of an infinite range of possible sounds and limited by our articulatory and auditory systems, follow a mostly universal development (Bialystok & Hakuta, 1994). And all of the world languages bestow upon us the ability to communicate with others and with our selves.

The study of how two languages represented in the bilingual mind and how they influence thought has been controversial (for a review, refer to Gonzalez, 1994; see also Vaid & Hall, 1991). As Bialystok and Hakuta (1994) indicate, the bilingual mind must function in a way that discriminates between two languages but, at the same time, allows for interaction between the two. In a complex manner, the bilingual speaker has attached two sets of linguistic codes to one conceptual representation: bilingual thought, then, involves a unique process of pairing and contrasting which is unnecessary in the monolingual mind.

Falvio (1991) reviewed the question of whether two languages function independently or interdependently on the bilingual mind. According to his extensive review of the literature, the data is controversial and there is support to both views: this suggests that the question is a matter of degree, espoused in the theory of the bilingual coding theory. This theory views a bilingual verbal system (L1 and L2) which have multiple connections to the non-verbal
imagery system (or concepts). Each language, in turn, converge on common concepts but can also tap into different concepts.

If the bilingual mind was different from the monolingual mind only in that concepts were represented in two linguistic codes, the discussion would end here. But the relationship between language and concepts is more complex, and has been controversial (Bialystok & Hakuta, 1994). The controversy involved the nature of language, thought, and culture; is thought shaped by language and, as such, do different cultures view the world differently? Are concepts different in different languages? Benjamin Lee Whorf (as cited in Bialystok and Hakuta, 1994; Whorf, 1956) believed that the concepts that we form and the way that we categorize our perceptions are shaped by the language that we speak. For example, the present author experienced a sort of cognitive incongruity when, in learning English, she encountered various English words to described what she had always known as, plainly, "un huevo frito" (a fried egg): sunny-side up, over-medium, over-easy, and so forth. In a way, her Spanish language had determined the way she would perceive a fried egg while, for an English-speaker, a fried egg was perceived in a variety of states. The notion that language defines our perception and concepts was called linguistic determinism. This view has been controversial but, according to Bialystok and Hakuta, the current view is
that language influences the shape of thought but not the categories of thought. In other words, categorical concepts, such as concepts of color, shape, movement, and position, are universal but are organized differently by different languages. For example, in terms of the concept of spatial relationships, the relationship between two objects can be defined, in English, by terms such as put on, put in, put together. In Korean, the way that two items can be described can only be explained by the extent to which the two items are tight- or loose-fitting. Thus, a single Korean word can describe putting on a ring or putting puzzles together, because they result in a tight-fit. On the other hand, the same Korean term is used to describe putting on a hat and putting apples in a bowl, because they result in a loose fit (p. 106). This lends support to the idea that a particular language organizes the way that we organize our perceptions. This is seen in translations, in which the goal is to transfer the meanings of one language into the terms and structure of the other language, rather than literally translating word-by-word.

In summary, the bilingual mind is different from the monolingual mind in that concepts are organized differently according to each language and there are two linguistic codes for each concept. In learning a second language, the bilingual speaker has to learn not only a new word but, also, a new concept.
Also controversial is whether the acquisition of a second language follows the same process of maturation as for first language acquisition (Bialystok & Hakuta, 1994). There seems to be, according to Bialystok and Hakuta, "reasonable" evidence that there is a critical period for first language acquisition, and suggestions that it may be by age 5. A second language can be acquired at a later age although there seem to be a decline in ability with age. A second language is learned quicker if it resembles the native language. And, although older learners can develop native-like command of most linguistic features, they usually retain an accent in the second language.

**Bilingualism, Culture, and Stress**

So far, the discussion regarding bilingualism and intelligence has been esoteric. But, for the approximately 35 million individuals who speak a language other than English at home, these issues define their adjustment in an English-speaking, broader, community. Acculturation issues are discussed next.

The linguistic environment is as important as the ethnic and cultural context in shaping the adjustment of Latinos in the United States. This adjustment includes self-esteem, racial-ethnic identity, social interactions, and use of language, among others (Rosado, 1986). Acculturation stress (Quintana, 1995) refers to the changes
experienced by racial, ethnic, and cultural groups as a result of their interaction. Both the minority group as well as the majority group are influenced by such interaction. Berry (as cited in Quintana, 1995), proposed four acculturation models to describe the minority response to such interaction, namely, integration, assimilation, separation, and marginalization, and four for the majority group, multicultural, assimilation, segregation, and ethnocide.

The concept of acculturation and its measurement, in turn, are not without critics. Some authors (Magaña, de la Rocha, Amsel, et al. 1996) object to the measurement of acculturation along a linear continuum of exclusive identification with the native culture to total adoption of the host culture. Such view ignores the possibility of biculturalism. Marin and Gamba (1996), in response to such concerns, developed the Bidimensional Acculturation Scale for Hispanics (BAS). Chávez, Morán, Reid, and López (1997), in addition, validated the use of the Societal, Attitudinal, Familial, and Environmental Acculturative Stress Scale (SAFE) for its use with children.

Over and beyond issues of definition and measurement, acculturation is a reality to ethnic minorities. As such, culturally sensitive approaches have been found to enhance the adjustment of Latinos in a number of areas, including response to treatment (Malgady, Rogler, & Costantino, 1994),
education (Abi-Nader, 1990), and proper referral to specialized services (Collier, 1985). On the other hand, “cultural dissonance” (Costantino, Malgady, & Rogler, 1994, p. 14) is found when the cultural beliefs and language of Latinos differ from those of the majority. The importance of multicultural milieu for Latinos must be underscored given that the adjustment of Latinos in the United States has been problematic. The occurrence of depression, high-school dropout, social and behavioral difficulties, adjustment problems, as well as self-esteem problems for Latinos have been found to be the greatest of all ethnic groups (Malgady, Rogler, & Costantino, 1990). Dropout rates for Latino college students is high (Solberg, Valdez, Villarreal, 1994). Urban Latino children characteristically experience a multitude of environmental stressors associated with high levels of aggression (Attar, Guerra, & Tollan, 1994). Costantino, Malgady, and Rogler (1994), in addition, pointed to language as a factor in heightening the level of isolation of Latinos.

It is fair then to state that the linguistic environment, intrinsic to a cultural milieu, has the power to either promote or hinder the adjustment of linguistic minorities in the United States. As previously discussed, the cultural environment is critical for the acquisition of a second language. For linguistic minorities, environmental stress has an impact not only in the eventual development of
a second language but in the immediate choice of language and the manner in which the chosen language, in turn, modulates the level of stress. Javier and Marcos (1989) ascertained the interaction between degree of stress and language as complex: at a mild level of stress, bilingual adults shifted more often between languages (Spanish and English). At higher level of stress, such code-switching was reduced, apparently because high stress disrupts the generalization of information from one language to the other. As the authors note, their findings concur with prior findings in that stressful experiences are encoded in the language in which the experience first occurred.

Language, Stress and Intelligence Tests

If the linguistic and cultural environment either fosters or hinders adjustment, what is its effect in a defined situation such as a testing session? There is scant empirical analysis on the impact of stress on the psychological assessment of linguistic minority children as it is mediated by language proficiency. However, Murphy (1990) investigated the degree to which anxiety explained differences in WISC-R scores for bilingual children. The obtained scores, however, were not reflective of high anxiety. Because the language proficiency of his sample was not reported (although the children belonged to bilingual education classes), it precluded exploration of the
interaction between intellectual scores, anxiety, and language proficiency.

The context as well as the nature of psychological examinations are assumed to present an extraordinary challenge for all children. However, the nature of the standardization sample, upon which intellectual tests are developed, is assumed to equally distribute the variance attributed to anxiety. For linguistic minorities, however, the challenge is potentially greater: at the receptive level, they are asked to recognize, understand, retain, recall, analyze, and process verbal information. At the expressive level, they are asked to produce a verbal answer: the content of such answer is presumably what is evaluated during an intellectual assessment. However, the structure of the verbal answer, that is, syntax, phonology and morphology, also have an impact on the intelligibility of the answer and, consequently, on its measured correctness. Along these lines, Argulewicz and Sanchez (1992), pointed that pronunciation difficulties in English are often confused with decoding difficulties and that the reading skill level of bilingual children is often underestimated. Furthermore, for bilingual children, the demands of intellectual tasks are markedly more complex: at one extreme, their language proficiency might not go beyond the level of recognition and understanding. The inability to provide an answer could relate to their limited English
proficiency and not to intellectual limitations. At the other extreme, a bilingual child might be able to process verbal information which is given in English. However, the form of their verbal answer might be colored by their native language and such interplay might be misunderstood as lack of intellectual ability.

The level of second language proficiency potentially ameliorates the linguistic demands of a decontextualized task such as an intellectual test. Cummins (1980) has investigated the degree of communicative skills in a second language a child would need in order to process the kind of information found in intellectual tests given in that particular language. He makes a distinction between Basic Interpersonal Communicative Skills (BICS) and Cognitive/Academic Language Proficiency (CALP). BICS refers to the mastery of certain aspects of language which allows a person to communicate in a social context and involves pronunciation, grammar and vocabulary. Proficiency in BICS allows a person to orally communicate an idea, information or an opinion in a natural manner. On the other hand, CALP refers to the level of proficiency in a second language needed to manipulate or reflect upon linguistic information in decontextualized academic situations. The distinction between CALP and BICS has been found to associate with degree of metalinguistic awareness and learning.
It follows that a bilingual child's level of proficiency has a significant impact on the significance of intellectual test results. For a child who has not reached a high level of proficiency in the language of the test, the necessary to think and to process information in the language of the test, the test results can be a reflection of language development and not of intellectual skills. For these children, the construct validity of the WISC-R must be questioned.

Integrated View of Bilingual Intelligence

Our review of the literature has pointed to the complexities involved in the characterization of bilingual intelligence. Intellectual testing in the United States has proven to be methodological and theoretically flawed but instrumental in current views about bilingualism.

The present study aimed at improving upon previous studies by the incorporation of language proficiency in the intellectual assessment of bilingual intelligence. At the same time, the testing environment was manipulated in terms of the value given to the child's linguistic background. Measurements of anxiety were incorporated to better understand the emotional impact of assessing intelligence in a child's secondly-acquired language.
Chapter III

METHODS

Overview and Design

This study consisted of an experimental design in which Language Proficiency and the Testing Environment (which constituted the independent variables), were analyzed as they impacted on the dependent variables of IQ, Anxiety, and Language Switch. Sixty fourth- and fifth-grade children were randomly assigned to three experimental conditions: Multicultural, Assimilative, or Neutral. The children were administered the following: the Wechsler Intelligence Scale for Children-Third Edition (WISC-III); the Woodcock-Muñoz Language Survey-English Form (WMS) and the Spanish Form (WMS); and the State-Trait Anxiety Inventory for Children (STAIC). Number of language switch instances were recorded. A brief individual interview was conducted with the children following the testing.

The variables were analyzed via a stepwise regression model to assess the contributions of the independent variables on the dependent variables; Anova analyses were also utilized.
Subjects

The sample consisted of 60 children drawn from the New Haven, Connecticut, community. The names of all children entering 4th grade bilingual classes or entering 5th grade bilingual classes, along with the names of 4th and 5th graders exiting bilingual education were obtained from the New Haven Public Schools. Proper authorization was obtained from the school district. In addition, flyers (in Spanish and in English) announcing the project were posted in area community centers, clinics, and in one private school. Only children with current or prior history of bilingual education and no history of special education were selected. Most children and parents were originally contacted by phone by the evaluators. However, approximately 15 to 20 children were obtained by word of mouth. After such initial identification, parents were contacted by phone or by letters. Invitations to participate in the study were then mailed to the parents. The evaluators then conducted home visits to all the parents who were interested. At this time, the parents were presented with a parental consent form and with a demographics questionnaire. Transportation was provided to all families to a private office, where all testing was conducted. The assent of the child was also obtained. All children were given a coupon to an area restaurant for their participation.
The subjects consisted of 30 females and 30 males, ages 6 years to 12 years 4 months. The mean age was 9.83, with a standard deviation of .88. Thirty students were entering the 4th grade, 29 the 5th grade, and 1 student the 6th grade. Thirty subjects were born in the United States and 25 in Puerto Rico. There were also 2 subjects from Peru, 1 from Honduras, 1 from Mexico, and 1 from Nicaragua. Nine of the subjects had been retained at least one grade. The average number of years of bilingual education for the sample was 3.8 years (SD=1.8). Twenty subjects were randomly assigned to each of the experimental condition.

**Procedure**

When the parent and the child arrived for the testing appointment, the evaluators, who were bilingual in English and in Spanish, met with the parent and the subject to greet them, orient them, and to obtain basic information about the child. The evaluators were the present author and a Master level bilingual Latino psychologist. The evaluators then administered the Woodcock-Muñoz Language Survey, first in Spanish and then in English, to the subject. Standard administration procedures were followed, that is, each test was administered exclusively in the language of the test. Immediately after, to avoid subject attrition during
repeated testing sessions, the evaluator randomly assigned the subject to one of two experimental conditions or to a control group. The subject then heard a pre-recorded standard set of instructions given individually to the child:

1. **Neutral Condition**—consisted of a general discussion of cultures and of people (see Appendix A).

2. **Assimilative Condition**—consisted of a general discussion of cultures and of people, followed by a discussion of the importance of the English language in the United States. The subjects were then instructed to communicate only in English during the evaluation (see Appendix B).

3. **Multicultural Condition**—consisted of a general discussion of cultures and of people, followed by a discussion of the importance of all languages in the United States. The subjects were then instructed to communicate either in English, Spanish, or both (see Appendix C).

The evaluators were blind to the content of the tapes and the child listened to the tapes with headphones. Immediately after, the child was administered the Wechsler Intelligence Scale for Children-III. This evaluation was given strictly in English and standard procedures were followed. In all instances, on the protocols, the evaluator recorded changes from Spanish (S) to English (E) and from E to S. Next, the subjects were administered the State-Trait
Anxiety Inventory for Children (STAIC). Several of the subjects were unable to understand the content of this test when given in English. Therefore, the STAIC was translated informally by the evaluators into Spanish. The subject was asked in which language he or she wanted the questions and our anecdotal observation was that many of the subjects requested the questions in Spanish. Following the administration of the STAIC, the subjects were given a brief individual interview (see Appendix D). Again, the evaluators communicated in the spontaneous language of the child. Finally, all children were briefed on the purpose of the study and were presented with a summary of the content found in the Multicultural condition. Children were encouraged to talk about their feelings and experiences.

Instruments

Wechsler Intelligence Scale for Children-III
(WISC-III; Wechsler, 1989)

The WISC-III is a recent update of the Wechsler Intelligence Scale for Children-Revised (WISC-R) and the third revision of the Wechsler Intelligence Scale for Children (WISC) developed in 1949. From its inception, the Wechsler scales were intended to measure the conglomerate of mental abilities that are presumed to promote rational and purposeful thinking and behavior. For the past 4 decades,
the psychometric properties of the scales have been extensively researched (for a review, refer to Sattler, 1990). In general, indexes of reliability and validity have been found to be robust.

The WISC-III was developed (Wechsler, 1989) for the purpose of updating norms, incorporating and expanding empirically-supported factor analytical concepts, and improving the administration and presentation of the items. The test, for use with children between the ages of 6 years-0 months to 16 years-11 months, consists of thirteen sub-tests that are organized into two sets consisting of verbal tasks and perceptual-motor tasks. Such organization yields a Verbal IQ score and a Performance IQ score; the aggregate of the sub-test scores yields a Full Scale IQ score. Four factor scores can also be obtained: Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed.

In order to understand the inclusion of Latino children in the standardization sample, a close inspection follows. In general (Wechsler, 1989, p. 20), the standardization sample consisted of 2,200 cases consisting of 200 children in each of 11 age groups from 6 through 16. One hundred females and one hundred males in each age group were included. In terms of race and ethnicity, the sample consisted of the following:
For each age group in the standardization sample, the proportion of Whites, Blacks, Hispanics, and other race/ethnic groups were based on the race/ethnic-group proportions of children aged 6-16 in the US population according to the 1988 Census survey. Each child in the standardization sample was categorized by his or her parent(s) as belonging to one of the following race/ethnic groups: White, Black, Native American, Eskimo, Aleut, Asian, Pacific Islander, or Other. The parents also indicated whether the child was of Hispanic origin. Consistent with the March 1988 Census survey, if the parent(s) indicated the child was of Hispanic origin, the child was assigned to the Hispanic category regardless of how the parent responded to the other race/ethnic-group question. (p. 20)

Regarding linguistic background and linguistic status, "Children were tested only if they could speak and understand English" (emphasis added). (p. 22). No further description of the procedure to assess the child's linguistic status is provided in the manual.

Woodcock Muñoz Language Survey
(WM; Woodcock & Muñoz-Sandoval, 1993).

The WM, English and Spanish Form, is an individually administered multiple-item test which assesses a broad sampling of proficiency in oral language, reading, and
writing. Each test form offers an overall measure of language competence. The test is intended to measure cognitive-academic language proficiency in both languages. The English normative data was obtained from a sample of 6,359 subjects which participated in the normalization of the Woodcock Johnson Psychoeducational Battery- Revised. The norms for the Spanish Form was obtained by equating levels of Spanish performance to like levels of performance in English. The data utilized for the Spanish sample consisted of 2,000 Spanish-speakers from Argentina, Costa Rica, Mexico, Peru, Puerto Rico, Spain, and the United States. Four subtests are included: Picture Vocabulary, Verbal Analogies, Letter-Word Identification, and Dictation. Three composite scores are also offered: Oral Language, Reading-Writing, and Broad Ability.

State-Trait Anxiety Inventory for Children

(STAIC; Spielberger, Edwards, Montouri, & Lushene, 1970).

The STAIC is comprised of two separate 20-item self-report scales, which assess either situation or chronic anxiety symptoms on a four-point Likert scale. State anxiety is understood to represent a transitory condition of unpleasant and conscious feelings of nervousness, tension, and apprehension. Trait anxiety represents relative stable individual feelings of anxiety (Novy, Nelson, Smith, Rogers, & Rowzee, 1995): the psychometric properties of the scale
have been found adequate, including internal consistency, factor structure, item-remainder correlations, and convergent and discriminant validity (p. 210). The State-Trait Anxiety Inventory for Children was initially developed to be used as a research tool to study the occurrence of anxiety in school-age children. It must be noted that, although the State-Trait Anxiety Inventory was recently translated into Spanish (Novy, Nelson, Smith, Rogers, & Rowzee, 1995), there is no such translation for the children's version.

Testing Environment

The testing environment was defined in terms of the instructions given to children prior to the standard administration of the WISC-III. This definition follows, in part, the definitions of Berry (as cited in Quintana, 1995):

1. Multicultural. All children will be presented with a discourse of cultures and the value of all cultures in society. The value of all languages and of bilingualism would be discussed. All children will be informed all testing will be given in English and that, during testing, to concentrate in all they know in order to answer all of the questions. Bilingual children, in addition, will be given this discourse in Spanish and will be encouraged to communicate, at any time in both Spanish and English.
2. Assimilative. All children will be presented with a discourse on the importance of bringing all cultures to a common ground in the United States and that such common ground was given by the English language. They will be reminded that all of the testing will continue in English and that only English would be allowed during testing because of the importance to follow the procedures given by the test.

3. Neutral. No instructions regarding use of languages will be provided.
Chapter IV

RESULTS

The subjects consisted of 30 females and 30 males, ranging from ages 8 years to 12 years 4 months. The mean age was 9.83, with a standard deviation of .88. Thirty students were entering the 4th grade, 29 the 5th grade, and 1 student the 6th grade. Thirty subjects were born in the United States and 25 were born in Puerto Rico. There were also 2 subjects from Peru, 1 from Honduras, 1 from Mexico, and 1 from Nicaragua. Nine subjects had been retained at least one grade. The average number of years of bilingual education for the sample was 3.8 years (SD=1.8). Twenty subjects were randomly assigned to each of the experimental condition.

The level of education of the parents in the sample was as follows: 16% had less than a 9th grade education; 25% had between 9th grade and 12th grade; 39% had graduated from High School; 5% were college graduates; and 12% had completed some form of technical training.

The means, standard deviation and sample sizes for the total sample on the Woodcock-Muñoz Language Survey and on the WISC-III are reported (see Table 1) to describe the sample in terms of the variables of language proficiency and
intellectual test scores. The average score for the W-M English Broad score was significantly lower than for the W-M Spanish Broad score, with $t(59) = -9.22, p < .001$. The subjects, therefore, were more proficient in Spanish than in English. Likewise, the average Performance IQ score was statistically greater than the average Verbal IQ score, with $t(57) = 10.16, p < .001$. In addition, the average discrepancy of -16.55 between Verbal IQ and Performance IQ was -16.55, with a standard deviation of 12.40, pointed to higher Performance IQ scores.

Table 1

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>W-M ENGLISH ORAL</td>
<td>87.823</td>
<td>18.3000</td>
</tr>
<tr>
<td>W-M ENGLISH READING</td>
<td>82.123</td>
<td>18.3000</td>
</tr>
<tr>
<td>W-M ENGLISH BROAD</td>
<td>77.5167</td>
<td>18.2716</td>
</tr>
<tr>
<td>W-M SPANISH ORAL</td>
<td>85.8167</td>
<td>12.8000</td>
</tr>
<tr>
<td>W-M SPANISH READ</td>
<td>103.4200</td>
<td>18.6516</td>
</tr>
<tr>
<td>W-M SPANISH BROAD</td>
<td>96.8967</td>
<td>12.8128</td>
</tr>
<tr>
<td>VERBAL IQ</td>
<td>72.3763</td>
<td>12.2018</td>
</tr>
<tr>
<td>PERFORMANCE IQ</td>
<td>96.7167</td>
<td>10.2004</td>
</tr>
<tr>
<td>FULL IQ</td>
<td>77.8661</td>
<td>11.0500</td>
</tr>
</tbody>
</table>

The means and standard deviations for the total group on the WISC-III factor scores were as follows: Verbal
Comprehension, Mean = 70.10, SD = 12.66; Perceptual Organization, Mean = 88.87, SD = 11.67; Freedom from Distractibility, Mean = 81.85, SD = 11.15; and Processing Speed, Mean = 93.74, SD = 9.77.

To examine the variation within the Woodcock-Muñoz Language Survey, the sub-test scores are summarized in Table 2. From a population Mean of 100 and a Standard Deviation of 15, the sample obtained a significantly low score on English Vocabulary and a significantly high score on Spanish Letter-Word Identification, when compared to the population, that is standardization, norms.

Table 2

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>W-M English Vocabulary</td>
<td>54.91</td>
<td>10.00</td>
</tr>
<tr>
<td>W-M English Analogies</td>
<td>85.23</td>
<td>10.12</td>
</tr>
<tr>
<td>W-M English Letter-Word</td>
<td>90.81</td>
<td>16.04</td>
</tr>
<tr>
<td>W-M English Dictation</td>
<td>73.36</td>
<td>17.31</td>
</tr>
<tr>
<td>W-M Spanish Vocabulary</td>
<td>88.02</td>
<td>10.48</td>
</tr>
<tr>
<td>W-M Spanish Analogies</td>
<td>86.58</td>
<td>7.72</td>
</tr>
<tr>
<td>W-M Spanish Letter</td>
<td>126.00</td>
<td>20.35</td>
</tr>
<tr>
<td>W-M Spanish Dictation</td>
<td>83.13</td>
<td>8.04</td>
</tr>
</tbody>
</table>

* Significant Strength
** Significant Weakness
Table 3 summarizes the sub-test scores on the WISC-III for the entire sample. When compared to the population mean of 10 and a Standard Deviation score of 3 for the WISC-III sub-test scores, the sample obtained significantly low verbal scores on Similarities, Comprehension, and Vocabulary.

Table 3

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION</td>
<td>5.0750</td>
<td>2.0602</td>
</tr>
<tr>
<td>SIMILARITIES</td>
<td>4.7500</td>
<td>3.0090</td>
</tr>
<tr>
<td>ARITHMETIC</td>
<td>8.5187</td>
<td>2.5478</td>
</tr>
<tr>
<td>VOCABULARY</td>
<td>4.4500</td>
<td>2.7083</td>
</tr>
<tr>
<td>COMPREHENSION</td>
<td>4.0920</td>
<td>2.7958</td>
</tr>
<tr>
<td>DIGIT SPAN</td>
<td>6.6441</td>
<td>2.4208</td>
</tr>
<tr>
<td>PICTURE COMPLETION</td>
<td>8.2097</td>
<td>3.6558</td>
</tr>
<tr>
<td>CODING</td>
<td>8.1709</td>
<td>2.2799</td>
</tr>
<tr>
<td>PICTURE ARRANGEMENT</td>
<td>7.2833</td>
<td>2.7687</td>
</tr>
<tr>
<td>BLOCK DESIGN</td>
<td>6.5333</td>
<td>3.1091</td>
</tr>
<tr>
<td>OBJECT ASSEMBLY</td>
<td>7.8850</td>
<td>3.1298</td>
</tr>
<tr>
<td>SYMBOL SEARCH</td>
<td>8.4005</td>
<td>2.8449</td>
</tr>
</tbody>
</table>

* Significant Strength

** Significant Weakness

A discussion of the results related to each hypothesis is presented next. Hypothesis 1 and 2, which were based on the same stepwise regression analysis, are presented one after the other. Then further discussion related to Hypothesis 1
is presented, with its respective statistical analyses. Finally, Hypothesis 3 and 4 are discussed.

**Hypothesis 1**

The discrepancy between measured verbal and performance abilities will vary according to measured relative language proficiency. Such discrepancy will be greater the greater the proficiency in Spanish versus English.

In order to explore the effect of language proficiency and the testing environment on the Verbal-Performance discrepancy, a stepwise regression was conducted with language proficiency and the Testing Environment (that is Multi-cultural, Assimilative, or Neutral), as well as Years of Bilingual Education and Gender, as predictors on the Verbal-Performance discrepancy. Both Broad English Ability and Gender were found to be significant on the regression equation ($F = 9.43, df = 2/51, p < .001$). Twenty-seven percent of the variance was explained by these factors, with Broad English Ability explaining 14.1% of the variance and Gender, 12.9% of the variance. Broad English Ability, in this manner, was the variable first entered in the regression equation, indicating its higher priority over the variable of Gender. Gender was entered second and the change is demonstrated by the change in $R^2$ ($\Delta R^2$). Table 4 shows the Stepwise Regression Analysis and Table 5 shows the Pearson's Correlation Coefficients.
Table 4
Summary of Stepwise Regression Analysis for Variables Predicting Verbal-Performance Discrepancy

<table>
<thead>
<tr>
<th>Model</th>
<th>Standardized Coefficients</th>
<th>Unstandardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Constant)</td>
<td>-38.555</td>
<td>1.824</td>
<td>-5.125</td>
<td>.000</td>
</tr>
<tr>
<td>WM</td>
<td>.291</td>
<td>.100</td>
<td>.375</td>
<td>2.278</td>
</tr>
<tr>
<td>ENGLISH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BROAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Constant)</td>
<td>-51.908</td>
<td>8.282</td>
<td>-6.256</td>
<td>.000</td>
</tr>
<tr>
<td>WM</td>
<td>.290</td>
<td>.080</td>
<td>.375</td>
<td>3.131</td>
</tr>
<tr>
<td>ENGLISH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BROAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STUDENT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENDER</td>
<td>8.045</td>
<td>3.016</td>
<td>.259</td>
<td>3.002</td>
</tr>
</tbody>
</table>

Note. R² = .141 for Step 1; ΔR² = .129 for Step 2

In summary, the results supported Hypothesis 1 in that the V-P discrepancy varies according to language proficiency. However, rather than relative language proficiency between English and Spanish, English vocabulary alone predicted the discrepancy. Further discussion will be provided in the Discussion session.

Hypothesis 2

The testing environment will influence the overall intellectual test scores of bilingual children: an empowering (supportive and bilingual) testing environment will lead to higher scores than those obtained in an assimilative (restrictive and monolingual) environment.
Table 5
Pearson's Correlation Coefficient for Variables Predicting Verbal-Performance Discrepancy

<table>
<thead>
<tr>
<th></th>
<th>M-H English Broad</th>
<th>M-H Spanish Broad</th>
<th>Experimental Condition</th>
<th>Years</th>
<th>Student</th>
<th>Gender</th>
<th>V-P Discrepancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-H English Broad</td>
<td>1.000</td>
<td>.282*</td>
<td>- .158</td>
<td>.099</td>
<td>.008</td>
<td>.387**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(p &lt; .024)</td>
<td></td>
<td></td>
<td></td>
<td>(p &lt; .003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-H Spanish Broad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>- .267*</td>
<td>-.241</td>
<td>-.202</td>
<td>.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(p &lt; .033)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental Condition</td>
<td>1.000</td>
<td>-.004</td>
<td>.000</td>
<td>-.097</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Ed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Gender</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V-P Discrepancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
</tbody>
</table>

* p < 0.05 (2-tailed);  ** p < 0.01 (2-tailed).

Based on the stepwise regression analysis and on the Regression Analysis of Table 4 and on the Pearson's Correlation Coefficients of Table 5, the Experimental Condition factor was excluded from the regression solution.

In summary, Hypothesis 2 was not supported by the results as the testing environment had no influence on the intellectual test scores of the subjects.

Related issues to Hypothesis 1 are presented next.
Language Proficiency and the Verbal-Performance Discrepancy

In order to further explore the impact of language proficiency on the verbal and performance discrepancy, a stepwise multiple regression analysis was performed. In this analysis, the various sub-tests of the Woodcock-Muñoz, namely, Vocabulary, Analogies, Letter-Word Identification, and Dictation, both in English and Spanish, were used as predictors of the Verbal-Performance Discrepancy. From the predictors, only Woodcock-Muñoz English Vocabulary was significant on the Verbal-Performance discrepancy scores (F = 29.26, df = 1/56, p = .00). English Vocabulary was found to predict 26.6% of the variance on the Verbal-Performance discrepancy. Excluded from the solution were Spanish Vocabulary as well as Analogies, Letter-Word Identification, and Dictation, in both English and Spanish. Table 6 represents the Regression Analysis and Table 7 represents the Pearson’s correlation coefficient matrix.

Table 6
Summary of Stepwise Regression Analysis for Variables Predicting Verbal-Performance Discrepancy

<table>
<thead>
<tr>
<th>Model</th>
<th>Standardized Coefficients</th>
<th>Unstandardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>SE</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>English Vocabulary</td>
<td>-33.203</td>
<td>5979</td>
<td>-5.570</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>WM English Vocabulary</td>
<td>298</td>
<td>.068</td>
<td>.515</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.502</td>
</tr>
</tbody>
</table>


Table 7

Pearson's Correlation Coefficient Between
Woodcock-Muñoz-English (WME), Woodcock-Muñoz-Spanish (WMS)
and Verbal-Performance Discrepancy

<table>
<thead>
<tr>
<th></th>
<th>WME</th>
<th>WME</th>
<th>WME</th>
<th>WME</th>
<th>WMS</th>
<th>WMS</th>
<th>WMS</th>
<th>V-P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vocab</td>
<td>Anal</td>
<td>Letter</td>
<td>Dicta</td>
<td>Vocab</td>
<td>Anal</td>
<td>Letter</td>
<td>Dicta</td>
</tr>
<tr>
<td>Vocab</td>
<td>.727*</td>
<td>.641*</td>
<td>.659**</td>
<td>.322**</td>
<td>.135</td>
<td>.081</td>
<td>.515**</td>
<td></td>
</tr>
<tr>
<td>Anal</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
</tr>
<tr>
<td>Letter</td>
<td>1.000</td>
<td></td>
<td>.649**</td>
<td>.764**</td>
<td>.135</td>
<td>.500**</td>
<td>.269**</td>
<td>.353**</td>
</tr>
<tr>
<td>Dicta</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
</tr>
<tr>
<td>Discrep</td>
<td>1.000</td>
<td></td>
<td>.645**</td>
<td>.764**</td>
<td>.135</td>
<td>.500**</td>
<td>.269**</td>
<td>.353**</td>
</tr>
<tr>
<td></td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
</tr>
<tr>
<td>Vocab</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anal</td>
<td>.852**</td>
<td>.527**</td>
<td>.288**</td>
<td>.391**</td>
<td>.427**</td>
<td>.266**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letter</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
</tr>
<tr>
<td>Dicta</td>
<td>1.000</td>
<td>.138</td>
<td>.320**</td>
<td>.354**</td>
<td>.414**</td>
<td>.235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrep</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
<td>(p&lt;.000)</td>
</tr>
<tr>
<td>Vocab</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dicta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrep</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p = 0.05 (2-tailed). **p = 0.01 (2-tailed).

In summary, the V-P discrepancy was explained by one specific aspect of overall proficiency in English, namely, English Vocabulary. Earlier, in examining Hypothesis 1 (refer to page 76), Broad English Ability was found to be one of the predictors of the Verbal-Performance Discrepancy.
However, it is only English Vocabulary which contributes to the variance in the Verbal-Performance discrepancy.

**Gender and the Verbal-Performance Discrepancy**

To explore the effect of gender on the VIQ-PIQ discrepancy, with a mean discrepancy score of -21.14 for females and -11.97 for Males, follow-up tests were conducted. A simple factorial Anova with Gender as the independent variable and V-P Discrepancy as the dependent variable, along with English and Spanish Broad Ability as covariates, was conducted. The main effect of Gender was still statistically significant, with an $F = 9.16$, $df = 1,54$, $p = .004$. English Broad Ability was also statistically significant, with an $F = 10.92$, $df = 1,54$, $p = .002$. However, Spanish Broad Ability was not statistically significant, with an $F = .315$, $df = 1,54$, $p = .57$.

In summary, Gender also had an effect on the V-P discrepancy, with females demonstrating larger discrepancies, even when the contribution of Language Proficiency was removed. Gender was introduced in this study to account for possible confounding effects of demographic factors and a brief discussion is provided in Chapter V.
Language Proficiency and the WISC-III Factor Scores

In order to further explore the impact of language proficiency on each of the WISC-III factor scores, separate stepwise multiple regression analyses were performed, with the scores for the Woodcock-Muñoz Language Survey, both in English and in Spanish, as predictors. From the predictors, Woodcock-Muñoz Broad English Ability was significant on Verbal Comprehension ($F = 69.437$, $df = 1/57$, $p < .001$), Perceptual Organization ($F = 13.941$, $df = 1/58$, $p < .001$), and Freedom from Distractibility ($F = 41.899$, $df = 1/57$, $p < .001$). Broad English Ability predicted 54.9% of the variance on Verbal Comprehension, 19.4% on Perceptual Organization, and 42.4% on Freedom from Distractibility.

For Processing Speed, both Broad Spanish Ability and Broad English Ability were significant ($F = 10.299$, $df = 2/54$, $p < .001$), contributing to 27.6% of the variance. Broad Spanish Ability, in turn, contributed to 20.9% of the variance while Broad English Ability predicted 6.7% of the variance. The Stepwise Regression Analyses for Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed, are presented on Tables 8, 9, 10, and 11, respectively. The Pearson's Correlation Matrix is presented on Table 12. Please refer to the Discussion section.
### Table 8
**Summary of Stepwise Regression Analysis for Variables Predicting Verbal Comprehension**

<table>
<thead>
<tr>
<th>Model</th>
<th>Standardized</th>
<th>Unstandardized Coefficients</th>
<th>Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>30.618</td>
<td>4.906</td>
<td>6.078</td>
<td>.000</td>
</tr>
<tr>
<td>W-1I</td>
<td>ENGLISH</td>
<td>552</td>
<td>0.066</td>
<td>.741</td>
<td>.333</td>
</tr>
<tr>
<td></td>
<td>BROAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 9
**Summary of Stepwise Regression Analysis for Variables Predicting Perceptual Organization**

<table>
<thead>
<tr>
<th>Model</th>
<th>Standardized</th>
<th>Unstandardized Coefficients</th>
<th>Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>87.274</td>
<td>5.708</td>
<td>15.702</td>
<td>.000</td>
</tr>
<tr>
<td>W-1I</td>
<td>ENGLISH</td>
<td>254</td>
<td>0.175</td>
<td>.440</td>
<td>.732</td>
</tr>
<tr>
<td></td>
<td>BROAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 10
**Summary of Stepwise Regression Analysis for Variables Predicting Freedom from Distractibility**

<table>
<thead>
<tr>
<th>Model</th>
<th>Standardized</th>
<th>Unstandardized Coefficients</th>
<th>Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>52.583</td>
<td>4.893</td>
<td>11.388</td>
<td>.000</td>
</tr>
<tr>
<td>W-1I</td>
<td>ENGLISH</td>
<td>388</td>
<td>0.080</td>
<td>.591</td>
<td>.473</td>
</tr>
<tr>
<td></td>
<td>BROAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table II

Summary of Stepwise Regression Analysis for Variables Predicting Processing Speed

<table>
<thead>
<tr>
<th>Model</th>
<th>Standardized Coefficients</th>
<th>Unstandardized Coefficients</th>
<th>Std. Error</th>
<th>Beta</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>60.948</td>
<td>2.760</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WAM</td>
<td>.347</td>
<td>.091</td>
<td>.457</td>
<td>3.611</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>SPANISH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BROAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(Constant)</td>
<td>55.418</td>
<td>8.773</td>
<td></td>
<td>8.174</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>WAM</td>
<td>.281</td>
<td>.091</td>
<td>.383</td>
<td>3.162</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>SPANISH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BROAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ENGLISH</td>
<td>.145</td>
<td>.055</td>
<td>.270</td>
<td>2.23a</td>
<td>.029</td>
</tr>
</tbody>
</table>

Note: $R^2 = .209$ for Step 1; $R^2 = .067$ for Step 2

Hypothesis 3

The testing environment will influence measures of anxiety for bilingual children: an empowering (supportive and bilingual) testing environment will lead to lower scores of anxiety than those obtained in an assimilative (restrictive and monolingual) environment.

In order to explore the effect of the testing environment, a discrete variable, on the measures of anxiety, a continuous variable, an Analysis of Variance (ANOVA) was conducted, with scores on the State scale of the State-Trait Anxiety Inventory as the dependent variable, and with scores of the Woodcock-Muñoz Language Survey.
Table 12
Pearson's Correlation Coefficient Between Woodcock-Muñoz-English (WME), Woodcock-Muñoz-Spanish (WMS) and WISC-III Factor Scores

<table>
<thead>
<tr>
<th></th>
<th>WME</th>
<th>WMS</th>
<th>VERBAL</th>
<th>PERCEPTUAL</th>
<th>FREEDOM</th>
<th>PROCESSING</th>
<th>DISTRACTIBILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>WME</td>
<td>1.000</td>
<td></td>
<td>.292**</td>
<td>.741**</td>
<td>.440**</td>
<td>.685**</td>
<td>.373**</td>
</tr>
<tr>
<td>WMS</td>
<td></td>
<td>1.000</td>
<td>.197</td>
<td>.175</td>
<td>.133</td>
<td>.457**</td>
<td></td>
</tr>
<tr>
<td>VERBAL</td>
<td>1.000</td>
<td></td>
<td>.366**</td>
<td>.411**</td>
<td>.264**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPREHENSION</td>
<td></td>
<td></td>
<td>(p=.001)</td>
<td>(p=.001)</td>
<td>(p=.019)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERCEPTUAL</td>
<td></td>
<td></td>
<td>.517**</td>
<td>.347**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORGANIZATION</td>
<td></td>
<td></td>
<td>(p=.000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FREEDOM</td>
<td></td>
<td></td>
<td>.567</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FROM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DISTRACTIBILITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROCESSING</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPEED</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p = .05 (2-tailed), **p = .01 (2-tailed).

English and Spanish Broad Ability, as covariates. The Mean score of STAIC-STATE was 51.933, SD = 10.23. With an alpha level of .05, the effect of the experimental condition was not statistically significant, F(4, 55) = 1.519, p = .209. In summary, Hypothesis 3 was not supported by the results.
Hypothesis 4

The testing environment will influence language use for bilingual children. Code switching (alternating use of Spanish and English) will be greater in an empowering (supportive and bilingual) versus an assimilative (restrictive and monolingual) environment.

Preliminary observations of the factor of Total Language Shifts revealed a severe positive skewness, with most of the scores in the low range. The significance of this skewness was analyzed by use of the ratio of Kurtosis, as suggested by Tabachnick and Fidell (1983). The ratio between the mean score of Language Shifts (20.979) to its Standard Error (.608) led to a rejection of normality for this factor. Therefore, a logarithmic transformation was applied to the data. Such transformed scores were then analyzed via an ANOVA procedure in order to assess whether the testing environment influenced code switching (alternating use of English and Spanish). The effect was not statistically significant, $F(2, 57) = .970$, $p = .385$.

In summary, Hypothesis 4 was not supported by the results.

In terms of the base language, this is, the language chosen by the subjects after the experimental condition and as they first responded to the WISC-III, only 10 subjects (16.7%) used Spanish first; 50 subjects utilized English first.
Regarding the Individual Interview, the Mean Score was 5.84, with an \( \text{SD} = 1.53 \). It must be recalled that the Individual Interview consisted of a 7-point scale. In this manner, the mean response to the question of "How did you feel during this activity?" was "Somewhat Comfortable". In terms of responses to whether the subjects were aware of cognitive strategies utilized during the testing, anecdotal observation indicated that the subjects were unable to verbalize such strategies.
Chapter V

DISCUSSION

Corroborating prior research studies (Cohen, Parmalee, Beckwith, & Sigman, 1986; Sandoval, 1979; Taylor, Ziegler, & Partenio, 1984), the present findings documented the pervasive VIQ-PIQ discrepancy found in Latino samples. The obtained discrepancy of -16.55 is greater than the 15 point discrepancy found to be significant at the alpha level of .01 found for all ages (Kaufman, 1994). Similarly, the discrepancy of -18.76 between the factor scores of Verbal Comprehension - Perceptual Organization is greater than the reported discrepancy of 16 points found significant at the alpha level of .01 (Kaufman, 1994).

The regression analyses confirmed, moreover, the prediction of the present study, that is, that such discrepancy varies according to language proficiency. At low levels of English proficiency, the VIQ-PIQ discrepancy was marked. But it seems that it is English Vocabulary alone which contributes to such discrepancy rather than other aspects of language: regression analysis corroborated that Vocabulary contributed to 26.6% of the variance in the VIQ-PIQ discrepancy. This finding challenges current views of the level of proficiency in a second language needed to
process decontextualized linguistic tasks (Cummins, 1980) and, in turn, the theoretical foundation of the Woodcock-Muñoz Language Survey (1993). According to Cummins, not all aspects of L2 language proficiency are involved in the mastering of cognitive and academic skills. While basic communicative skills (referred to as Basic Interpersonal Communicative Skills, or BICS) are involved in day-to-day interpersonal communication, higher cognitive and academic learning requires abstract language skills (also referred to as Cognitive/Academic Language Proficiency, or CALP). CALP presumably consists of skills related to literacy, such as reasoning, abstraction, and inferential skills. As such, the Woodcock-Muñoz Language Survey’s sub-tests (namely, Vocabulary, Analogies, Letter-Word Identification, and Dictation) were developed as an overall measure of CALP. In the present study, however, it was only Vocabulary, a task which requires expressive responses, that is, labeling pictures, what contributed to the Verbal-Performance discrepancy. Seemingly more abstract tasks, such as verbal analogies, reading decoding, and spelling, were found not significant in the regression solution for the Verbal-Performance discrepancy, in the present study.

Similarly, language proficiency in Spanish was not found to be significant as a predictor of the Verbal-Performance discrepancy in the present study. In the
context of the average scores obtained by the sample in Broad Spanish ability (Mean = 95.97, SD = 12.81) it could be argued that these children had attained solid mastery of their first language. In this way, based on Bickerton’s (1995) views, the subjects had developed the ability to communicate both with others and with themselves, internally. Given that these subjects, moreover, had no history of referral to special education, their obtained average Full IQ score of 77.96 should be questioned. On the other hand, could it be that the present sample had indeed low intelligence and thus low attainment of English? The answer to that question is beyond the data available in this study especially because the face value of the intellectual test scores obtained is in question. Moreover, intelligence is only one of the complex and intertwined factors associated with the development of a second language (Hakuta, 1986; Krashen, 1992) and simplistic explanations of the association between intelligence and second-language acquisition are not likely. Moreover, there seems to be no support for the view that a sign of intelligence is the ability to master English as a second language.

The findings of the present study also raised concerns regarding current practices in the assessment of Latino children. English proficiency explained a significant amount of variance not only in the VIQ-PIQ discrepancy, but also in the factor scores. English proficiency predicted
54.9% of the variance in Verbal Comprehension and predicted
19.4% of variance in Perceptual Organization. Again,
Perceptual Organization scores were lower at low levels of
English proficiency. The present results regarding language
proficiency and factor scores must be interpreted with
cautions: the separate regression analyses may have had
increased the chances of spurious effects. Nevertheless,
current recommendations on the use of the Perceptual
Organization Index as the best estimate of intellectual
functioning for bilingual children (Kaufman, 1994) must be
revisited.

The effect of language proficiency on the factor score
of Processing Speed merits close examination, as both
English and Spanish Broad Ability predicted PS scores. This
factor is considered a measure of psychomotor speed as well
as of mental speed. According to Kaufman (1994), PS also
reflects good visual memory, planning ability and, possibly,
the capacity to deal with simultaneous information. Paired
with Bialystok and Hakuta's (1994) contention that a
bilingual speaker has two linguistic codes for one
conceptual representation, then it is speculated that the
subjects tapped into their knowledge in either language.
This is, the subjects had mastered the concept of visual
memory and planning and were thus able to complete PS tasks
(Coding and Symbol Search). And, although both languages
were included in the regression equation, Spanish
proficiency explained 20.9% of the variance while English proficiency explained only 6.7% of the variance for Processing Speed. Further explanation of the influence of both languages into PS awaits further research.

Gender effects were an unexpected finding in the present study: 12.9% of the variance in the VIQ-PIQ was attributed to gender differences, and greater discrepancies were obtained by females. English proficiency seemed to modulate this interaction to a small degree but the effect of gender remained even after language proficiency was factored out. Given that the standardization sample of the WISC-III accounted for the variability due to gender, explanations for the present findings are to be found somewhere else. One possibility was the presence of an interaction effect between the gender of the subjects and the gender of the examiners, in this case, females. If this were the case, the interaction could be potentially complex. Glutting, Oakland, and McDermott (as cited in Kaufman 1994), have documented the relationship between IQ test scores and certain observed behaviors. At this time, explanations of gender effects on the VIQ-PIQ discrepancy for Latino children needs future research. Although the present results raised more questions than it answered regarding gender effects, this variable proved to be a fruitful area of research.
In the present study, the Experimental Condition, namely Neutral, Assimilative, and Multi-cultural, had no bearing on the VIQ-PIQ discrepancy, on measures of anxiety, nor on language shift. One factor may explain the lack of experimental effect of the Experimental Condition. Only 10 subjects began the WISC-III in Spanish. Both in the Neutral Condition and in the Multi-cultural Condition, where subjects were instructed to use either English and Spanish, the subjects preferred not to utilize their native language. Thus, this may suggest that the linguistic environment of the test, that is English, and not the experimental instructions, established the response set. This follows along the lines of Reichman (1997), who notes the tendency of Hispanic subjects to use English to respond to impersonal questions whereas Spanish seems to be the language of choice when expressing physical or emotional distress.

Linguistic factors might have also explained the lack of influence of the testing environment in measures of anxiety. Although most subjects preferred English during the structured administration of the WISC-III, our anecdotal observation was that many subjects requested the STAIC to be given in Spanish. On the one hand, such a request then highlights that the subjects indeed sought their first language when requested to process emotional information. On the other hand, however, it may have contaminated their responses because the acknowledgment and reassurance from
the examiners could have diffused any anxiety experienced during the WISC-III. In fact, the response from the examiners might have inadvertently led to the cultural sensitivity believed to enhance the adjustment of Latinos in treatment (Malgady, Rogler, & Constantino, 1994). In addition, given that the STAIC was indeed informally translated by the examiners, then the test might have lost its validity. Furthermore, it is here acknowledged that there may be further statistical analyses to study the effect of the testing environment on anxiety, such as a regression analysis. This analysis would require a larger number of categories to approximate the continuous variables required in regression equations. This was beyond the possibilities of the present study, which defined the testing environment as a discrete variable with only three levels (that is, Multicultural, Assimilative, and Neutral).

Finally, the experimental condition failed to influence language shifting during the WISC-III. In fact, this variable was found to be skewed toward low scores, that is, low instances of code switching. Again, it is speculated that the linguistic environment of the WISC-III established the response set and thus most subjects utilized English only.
Conclusions

The above discussion has highlighted the reservations in the accurate interpretation of intellectual test scores for Latino children. The above data questioned the validity of interpreting both verbal and non-verbal IQ scores as these scores seem to reflect proficiency in the second language of the subjects. But Latinos are also known to obtain lower scores on other standardized tests, such as neuropsychological tests (Adams, Boake, & Crain, 1982) and reading tests (Argulewicz & Sanchez, 1992). Moreover, research on white-Hispanic differences on psychological tests continue (Kaufman, McLean, Kaufman, & Kaufman, 1994) and the lower scores of the Hispanics continue to be attributed to the fact that bilingualism "might impair crystallized abilities" (p 1286). For the 32 million individuals, in the United States, who are speakers of English as a second language, the understanding of the impact of bilingualism on intelligence is far from academic. Bilingualism is not solely the speaking of two languages: bilingualism is a force by which people organize their internal experiences as well as their social interactions. It is tightly related to self-identity and self-esteem. And once one becomes bilingual, there is no such a thing as becoming monolingual. As such, present assessment practices must be challenged to ensure that language proficiency and ethnic identification are accounted for in intellectual test
research. Future research is needed to expand on the understanding of how language proficiency impact test scores. It has been previously discussed that, one solution is to statistically remove the contribution of language proficiency from test scores. However, such solution is unnecessarily artificial: language proficiency, in any language, is inherent in intelligent behavior.

Beyond the individual, however, the understanding of bilingualism has an impact at the society level. Ethnic minorities continue to be poorly understood in the United States and pervasive problems remain in mental health care, such as paucity of research and scarcity of ethnic minority professional (Vargas & Willis, 1994), barriers to mental health care (Ruiz, 1995), and proper conceptual models to study minority children (Garcia Coll et al., 1996). Given the ongoing attention to acculturation and psychosocial adaptation between cultural groups (Chávez, Morán, Reid, & López, 1997), the issue of bilingualism must be in the forefront.

The question of whether emotional factors also impact the scores obtained in standardized tests given in a second language also remains to be further explored. The present study suggested, however, that the subjects preferred their first language when requested to handle emotional material. However, their preference for English during standardized
testing, although less developed than their overall proficiency in Spanish, suggests that the linguistic environment of the test itself sets the stage for the use of their less dominant language. Finally, future research can further explain the effect of gender in the VIQ-PIQ discrepancy for such groups.

Although the present study attempted to shed light on such issues, further research is needed. It remains to be empirically explored whether the present results can be duplicated for other standardized measures and for other linguistic minorities of various SES background. Specially critical is to expand the definition and the measurement of the variable of language proficiency: the operational definition in the present study was narrow.

The need for further research on the intellectual assessment of Latino children must be underscored. As such, this author agrees with Costantino (1992) in that the controversy regarding test bias for Latinos remains unresolved despite fifty years of debate. The "status quo hypothesis" (p. 94), that there is no test bias, must be challenged. And test bias permeates areas beyond intellectual testing. Tests in education, industry, research, and clinical fields, are also in dire need of reform. These issues are discussed at length by Geisinger (1992).
As a final word, although the present study examined testing issues with Latino children, this area of research is likely to benefit all linguistic groups as well. In challenging the status quo described by Costantino (1992), the linguistic majority groups, in this case, monolingual English speakers in the United States, also benefit. One immediate benefit is the eradication of wasted potential in all citizens. By eliminating test bias, all citizens can reach their full potential. A second benefit is more subtle: the more we know about the bilingual mind, the more we will discover about the monolingual mind. In other words, the understanding of bilingualism and of bilinguals will also enhance our understanding of "language, mind, and society" (Hakuta, 1996, p. 240).
References


Gonzalez, V. (1994). A model of cognitive, cultural, 
and linguistic variables affecting bilingual Hispanic 
children's development of concepts and language. Hispanic 

social context of testing. American Psychologist, 36, 
1167-1171.

assessment of racial and ethnic minority groups: 
Implications. Journal of Counseling and Development, 64, 
635-637.

similarity of the WISC-R for Anglos and Chicanos referred 
for psychological services. Journal of School Psychology, 
18, 34-39.

Gutthbreund, D. G. (1990). Effects of language usage on 
the emotional experience of Spanish-English and 
English-Spanish bilinguals. Journal of Consulting and 
Clinical Psychology, 58, 604-607.

Hakuta, K. (1986). Mirror of language: The debate on 

Hakuta, K. (1987). Degree of bilingualism and cognitive 
ability in mainland Puerto Rican children. Child 
Development, 58, 1372-1388.

acquisition: A book of readings (pp. 23-32). Rowley, MA: 
Newbury House Publishers, Inc.

In C. A. Ferguson and S. B. Heath (Eds.). Language in the 
USA. Cambridge: Cambridge University Press.

Herrnstein, R. J., & Murray, C. (1994). The bell curve: 
Intelligence and class structure in American life. New York, 
NY: The Free Press.

Holland, W. R. (1960). Language barrier as an 
educational problem of Spanish-speaking children. 
Exceptional Children, September, 42-50.

Individuals with Disabilities Education Act, Pub. L. 
No. 105-17 (1997).


Prawitt Diaz, J. O., & Rivera, R. (1989). Correlations among scores on Woodcock-Johnson Achievement subtest (Spanish), WISC-R (Spanish) and Columbia Mental Maturity Scale. Psychological Reports, 64, 987-990.


Appendix A

Neutral Condition
Appendix A
Neutral Condition

The United States is the Country where we live. In the United States, many cultures live together and people are of different colors, backgrounds, and languages. People of different cultures like to eat different foods and might even dress in different ways. Some people came to the United States many, many years ago, while some people came to the United States recently.
Appendix B
Assimilative Condition
Appendix B
Assimilative Condition

The United States is the Country where we live. In the United States, many cultures live together and people are of different colors, backgrounds, and languages. People of different cultures like to eat different foods and might even dress in different ways. Some people came to the United States many, many years ago, while some people came to the United States recently.

Because there are so many people in the United States, it is important that all people speak the same language, English. Everybody should learn English and speak English only. English only makes the United States a great country!

In the next task you will do, the questions will be given in English. You should think and speak English only. Remember, use English only!
Appendix C

Multi-cultural Condition
Appendix C

Multi-cultural Condition

The United States is the Country where we live. In the United States, many cultures live together and people are of different colors, backgrounds, and languages. People of different cultures like to eat different foods and might even dress in different ways. Some people came to the United States many, many years ago, while some people came to the United States recently.

Because there are so many people in the United States, all cultures and languages are important. Everybody should feel free to speak their own language and should also speak English. When people speak only English, they should learn another language. All languages and all cultures make the United States a great nation!

In the next task you will do, all questions will be given in English. However, you should think and speak in either Spanish and English or in both. Remember, you can use Spanish or English!
Appendix D

Individual Interview
Appendix D

Individual Interview

1) In a scale of 1 to 7, with 1 being Extremely Uncomfortable to 7 Extremely Comfortable, how did you feel during this activity?
   1 - Extremely Uncomfortable
   2 - Very Uncomfortable
   3 - Somewhat Uncomfortable
   4 - Comfortable
   5 - Somewhat Comfortable
   6 - Very Comfortable
   7 - Extremely Comfortable

2) When you answered the questions and solved the problems, how did you do it? How did you figure it out in your mind, especially when the questions were hard?