# False positive rates in the neuropsychological assessment of neurologically intact bilingual Hispanic American adults

Cassandra Dayanira Gonzalez
*University of Texas-Pan American*

# FALSE POSITIVE RATES IN THE NEUROPSYCHOLOGICAL ASSESSMENT OF NEUROLOGICALLY INTACT BILINGUAL HISPANIC AMERICAN ADULTS

A Thesis

by

CASSANDRA DAYANIRA GONZALEZ

Submitted to the Graduate School of the
University of Texas-Pan American
In partial fulfillment of the requirements for the degree of

MASTER OF ARTS

December 2010

Major Subject: Psychology

# FALSE POSITIVE RATES IN THE NEUROPSYCHOLOGICAL ASSESSMENT OF NEUROLOGICALLY INTACT BILINGUAL HISPANIC AMERICAN ADULTS

A Thesis
by
CASSANDRA DAYANIRA GONZALEZ

COMMITTEE MEMBERS

Dr. Philip G. Gasquoine
Chair of Committee

Dr. James W. Aldridge
Committee Member

Dr. Kristin L. Croyle
Committee Member

December 2010

ABSTRACT

Gonzalez, Cassandra Dayanira, False Positive Rates in the Neuropsychological Assessment of Neurologically Intact Bilingual Hispanic American Adults. Master of Arts (MA), December, 2010, 30 pp., 6 tables, 3 illustrations, 40 references.

Hispanic Americans perform more poorly than White Americans on neuropsychological tests, leading to a higher misclassification rate of brain injury (false positives) among Hispanics when using the 50th percentile of published norms as the estimate of preexisting neuropsychological skill. This study aimed to determine if using an individual comparison standard to estimate preexisting skill levels would reduce the number of false positives in a sample of 20 bilingual Hispanic American adults. Two individual comparison standards were used: (a) WMLS-R Picture Vocabulary scores, and (b) WMLS-R Picture Vocabulary scores and WAIS-III Matrix Reasoning scores combined. Both individual comparison standards were found to produce less false positives than the 50th percentile of published norms. For all three methods of estimating preexisting skill levels, there were more false positives in English than Spanish.

# DEDICATION

This goes to all of you who continually encouraged me, no matter how hopeless it seemed.  To my mom and dad, who assisted in any way that they could; my sister, Stephanie, who helped me practice testing and scoring, even if it was against her will; and to Antonio, who pushed me and motivated me when I needed it most.  Thank you for your love and support.

## ACKNOWLEDGEMENTS

I would first like to acknowledge Dr. Philip Gasquoine, my committee chair, for his profound mentorship and advice throughout this project. It was his patience and guidance that provided the driving force towards the completion of this thesis, and I am forever grateful. My thanks go out to all of my committee members for their advice and input and for helping me achieve the best quality of work possible. I would also like to thank all of the participants in this study for giving up so much of their time and effort to help me complete my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

Neuropsychology is a branch of psychology that studies the relationship between the brain and behavior. It has a clinical wing that treats patients with structural brain injury, from such neurological conditions as stroke, traumatic brain injury, and seizure disorder. The main concern of clinical neuropsychology is the assessment and treatment of structural brain injury among human individuals.

Structural brain injury occurs when there is neuronal damage within the brain. Symptoms of structural brain injury are primarily manifested behaviorally, in the form of sensory, motor, and/or cognitive deficits. Secondary personality and/or emotional changes sometimes occur as well. When cognitive symptoms exist, structural brain injury is suspected.

Clinical neuropsychologists are commonly called upon to answer questions regarding the presence or absence of cognitive impairment and ecological questions such as whether or not a patient is able to drive a vehicle or return to work after structural brain injury (Chaytor & Schmitter-Edgecombe, 2003). When patients are referred to clinical neuropsychologists for assessment, they typically undergo administration of a neuropsychological test battery to assess different domains of cognitive functioning like: attention, memory, language, visual/perceptual, sensory/motor, and executive skills. Poor scores in any of these domains can point to cognitive impairment, allowing neuropsychologists to make clinical decisions regarding the presence of structural brain injury and its implications for the patient's future.

## Estimating Preexisting Skill Levels

Difficulty in clinical interpretation arises when a neuropsychologist must decide whether a poor neuropsychological test score is due to structural brain injury or a preexisting condition. This process is known as deficit measurement (Lezak, Howieson, & Loring, 2004). Deficit measurement requires a neuropsychologist to estimate preexisting neuropsychological skill levels as a basis for comparing actual post-onset scores. Preexisting neuropsychological skill levels have traditionally been estimated by using one of two approaches: the 50$^{th}$ percentile of published norms (Heaton, Miller, Taylor, & Grant, 2004) or an individual comparison standard (e.g., Lezak et al., 2004). The first approach always estimates the preexisting neuropsychological skill level at the mean (50$^{th}$ percentile) of the standardized sample, and deficit measurement compares the actual level of cognitive functioning to that estimate. An individual comparison standard provides a unique estimate of preexisting neuropsychological skill levels for each patient based on certain individual variables such as demographic data, achievement records, and actual test scores.

## Misclassifications

Results of neuropsychological assessment can lead to four possible outcomes, as shown in Figure 1.

*Results of Neuropsychological Assessment*

| | | Impaired | Not Impaired |
|---|---|---|---|
| *"True" State* | **Impaired** | Sensitivity | False negative |
| | **Not Impaired** | False positive | Specificity |

*Figure 1.* Classifications resulting from neuropsychological assessment. This figure illustrates the four possible outcomes of interpretations made from neuropsychological assessments based on results of neuropsychological assessments and patients' true state.

A neuropsychological assessment that indicates a patient is impaired when impairment truly exists is said to possess sensitivity, and one that correctly rules out the presence of impairment is said to have specificity. Because clinical neuropsychologists typically have no way of knowing a patient's actual preexisting neuropsychological functioning, however, the problem of misclassification arises. Misdiagnosis among patients is inevitable to some degree. When a neuropsychological assessment indicates the presence of impairment that is not actually present, it is called a false positive. The flip side is a false negative, where results of an assessment fail to indicate impairment that actually exists.

Deficit measurement usually determines the presence of impairment when post-onset scores fall 1 *SD* below the estimate of preexisting skill since this cutoff score has typically maximized sensitivity and specificity in discriminating between individuals who are neurologically intact and those with structural brain injury (Heaton et al., 2004). Based on this cutoff, about 16% of neurologically intact individuals would be misdiagnosed as cognitively impaired (false positives), as seen in Figure 2.

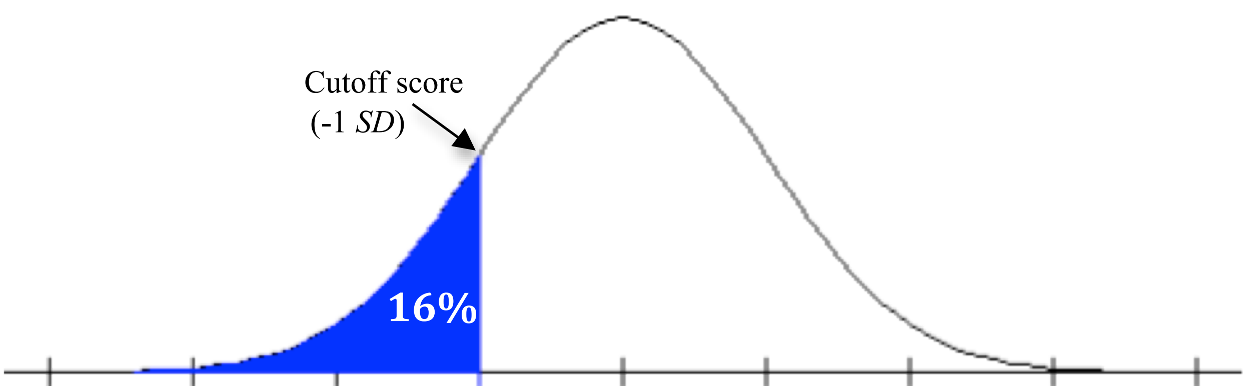


*Figure 2*. False positive percentage using the 50th percentile of published norms as the estimate of preexisting neuropsychological skill level. When the 50th percentile of published norms is used as the estimate of preexisting skill level, approximately 16% of individuals are expected to be misdiagnosed as cognitively impaired assuming neuropsychological skills are normally distributed.

**Neuropsychological Evaluation of Hispanic Americans**

Neurologically intact Hispanic Americans tend to score more poorly that White Americans on intelligence tests and neuropsychological tests (e.g. Gurland, Wilder, Cross, Teresi, & Barrett, 1992; Arnold, Montgomery, Castaneda, & Longoria, 1994). This difference is of approximately .5 *SD* on language measures (e.g., Puente & Salazar, 1998). This places Hispanic Americans at greater risk of false positive misdiagnosis. With a normal distribution of scores approximately .5 *SD* lower than White Americans, approximately 31% of Hispanic Americans would be misdiagnosed as cognitively impaired using the 50$^{th}$ percentile of published norms as the estimate of preexisting skill, as shown in Figure 3.

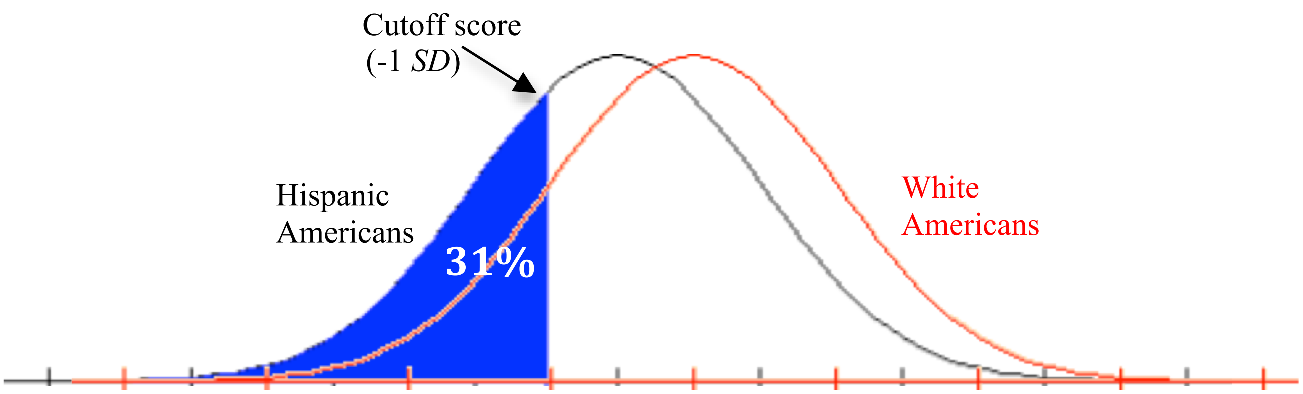


*Figure 3.* False positive percentage for Hispanic Americans versus White Americans using the 50$^{th}$ percentile of published norms as the estimate of preexisting neuropsychological skill level. This figure shows the expected percentage of a normal distribution of Hispanic Americans falling below the cutoff for cognitive impairment compared to non-Hispanic White Americans when using the 50$^{th}$ percentile of published norms as the estimate of preexisting skill.

Race-norming (creating different stratifications of neuropsychological test norms by race/ethnicity) is one method that has been suggested to reduce false positive rates of cognitive impairment among minority groupings. As illustration, Gladsjo et al., (1999) showed that race-norming decreased false positive misclassification rates for neurologically intact African Americans on the Verbal Fluency Test; sensitivity and specificity could not be determined. Although there may be some advantages to race-norming, there are numerous problems with this approach, including: possible justification for differential treatment of different race/ethnic groupings; the overlooking of factors possibly responsible for group racial differences, such as acculturation; unknown effects on false negative rates; non-scientific definitions of race/ethnic groupings; and an impossibly large number of potential race/ethnic groupings for which to generate race-norms (Gasquoine, 2009).

Individual comparison standards are an alternative approach to estimating preexisting neuropsychological skill, which have been found to be successful in reducing false positive rates among neurologically intact Hispanic Americans (Gasquoine, Croyle, Cavazos-Gonzalez, & Sandoval, 2007). However, there is no consensus regarding the best method for determining individual comparison standards for clinical use. The purpose of this study is to try to find the best method of estimating preexisting neuropsychological skill for bilingual Hispanic Americans among the 50$^{th}$ percentile of published norms and two individual comparison standards based on current verbal and visual-perceptual skill.

# CHAPTER II

## LITERATURE REVIEW

English language fluency, a correlate of race/ethnicity, is known to affect neuropsychological test scores (Gasquoine, 1999). Consequently, much research has focused on creating Spanish language tests with separate norms for Spanish speakers. This reduces the risk of false positives among Spanish speakers. The problem with this approach is that these Spanish-language norms are usually foreign-generated using monolingual Spanish-speaking subjects from Latin countries. As such, they may not generalize to bilingual populations (Gasquoine, 2008).

Monolingual, Spanish-speaking, Hispanic Americans are fairly common among the elderly (Gasquoine, 2001), and many neuropsychological tests have been normed for this age grouping within the United States (e.g., Acevedo et al., 2000; Stricks, Pittman, Jacobs, Sano, & Stern, 1998). However, monolingual, Spanish-speaking, Hispanic Americans between the ages of 18 to 65 are less common (Gasquoine, 2008). Most Hispanic Americans within this age group that are living in the United States are bilingual. As such, foreign-generated, monolingual, Spanish language norms may not be suitable for bilingual Hispanic American individuals.

Bilingualism is another variable that has the potential to negatively impact neuropsychological test scores, but this relationship has not been well studied (Gasquoine, Cavazos, Cantu, & Weimer, 2010). Research has shown that Hispanic American, Spanish/English bilinguals perform significantly more poorly than either Spanish or English monolinguals on neuropsychological tests (Gollan, Montoya, & Wener, 2002; Gasquoine et al., 2007). The reason for this deficit is still unclear, although it is believed that interlanguage interference is caused by parallel activation of both languages, which slows processing time and increases the possibility of errors (Bialystok, Craik, & Luk, 2008).

Rosselli et al. (2002) suggested that when studying the relationship between bilingualism and cognition it is best to use two languages of administration. Gasquoine et al. (2007) employed this strategy in studying the effects of bilingualism on neuropsychological test scores. Participants were tested in both English and Spanish, and false positives (calculated by using the 50$^{th}$ percentile of published norms) were compared across bilingual groupings. Higher false positive rates in English than Spanish were found across all bilingual groupings. Unexpectedly, it was noted that the use of an individual comparison standard, as compared to the 50$^{th}$ percentile of published norms, reduced these false positive rates in both languages.

## Estimating Individual Comparison Standards

There is little research on the topic of individual comparison standards probably because there is no uniform agreement regarding the best method of making these estimates (Gasquoine, 2009). At least six different variants have been used as estimates of preexisting skill level: (a) demographic variables, (b) school achievement test records, (c) post-onset word reading levels, (d) post-onset vocabulary test scores, (e) other single post-onset IQ subtest scores, and (f) the highest score obtained during post-onset neuropsychological testing.

### Demographic Variables

Various demographic characteristics have been combined into regression equations to estimate patients' Full Scale IQ (FSIQ; e.g., Barona & Chastain, 1986; Barona, Reynolds, & Chastain, 1984; Crawford & Allen, 1997). Coding values are given for each demographic variable. For the demographic variable of race, for instance, *White* might be assigned the value of 3, *Black* a value of 2, and *Other* a value of 1. The formula then gives differential weighting to each value to produce an estimated FSIQ score, such as Barona et al.'s (1984): Estimated FSIQ = 54.96 + 0.47 (age) + 1.76 (sex) + 4.71 (race) + 5.02 (education) + 1.89 (occupation) + 0.59 (region).

**School Achievement Test Records**

Scores on achievement tests administered in schools, such as the Iowa Test of Basic Skills, and those used in college admissions, such as the American College Test, are highly correlated with IQ. Baade and Schoenberg (2004) recommended the use of the *predicted difference method* to estimate FSIQ from achievement test data. In order to apply this approach, the achievement test national mean, standard deviation, and correlation with the Wechsler FSIQ must be known for the time of testing.

**Post-onset Word Reading Levels**

Tests that require the correct pronunciation of irregularly spelled words are also considered to be relatively resistant to brain injury. Those used to estimate preexisting cognitive levels have included variants of the Hopkins Adult Reading Test (HART: Schretlen et al., 2009); North American Adult Reading Test (NAART: Blair & Spreen, 1989); and the Wide Range Achievement Test, Reading subtest (WRAT: Jastak & Wilkinson, 1984). The *hold* nature of word pronunciation measures has been verified in longitudinal studies of patients with chronic schizophrenia (Harvey et al., 2006) and dementia (McGurn, et al., 2004). *Hold* refers to measures of crystallized intelligence that are considered relatively resistant to brain injury.

**Post-onset Vocabulary Test Scores**

Historically, the oldest method of estimating preexisting skill involves using a vocabulary score. Babcock (1930) measured "mental deterioration" in psychotic patients at Manhattan State Hospital by calculating the score discrepancy between measures of vocabulary and memory/timed tasks. Vocabulary was chosen because it is thought to be a *hold* skill, often preserved relative to cognitive changes from brain insult and disease. There are various ways of measuring vocabulary, including matching words to pictures and providing verbal definitions. Another common way of measuring vocabulary involves naming pictures, such as in the Picture Vocabulary subtest of the Woodcock-Munoz Language Survey-Revised (Woodcock, Munoz-Sandoval, Ruef, & Alvarado, 2005). Thus, Picture Vocabulary subtest scores were used as an individual comparison standard in the current study.

**Other Single Post-onset IQ Subtest Scores**

The Matrix Reasoning subtest of the Wechsler Adult Intelligence Scales is another single IQ subtest score that has been used to measure preexisting cognitive levels (e.g., Green et al., 2008). Like vocabulary, Matrix Reasoning skills are thought to be resistant to structural brain injury. For this reason, Matrix Reasoning subtest scores were used as an individual comparison standard for estimating preexisting visual-perceptual skill in this study.

**Highest Score Obtained During Post-onset Neuropsychological Testing**

Another approach for estimating preexisting skill levels involves using the highest post-onset score (on any test) or demographic data point as the estimate of preexisting cognitive function (Lezak et al., 2004). This approach, also known as the best performance method, has long been criticized as routinely producing overestimation of preexisting skill levels (e.g., Baade & Schoenberg, 2004).

## Applying Individual Comparison Standards

Deficit measurement proceeds by comparing post-onset neuropsychological test scores to the selected individual comparison standard and inferring cognitive impairment whenever the difference between the two scores is greater than some amount, such as 1*SD* (Heaton et al., 2004). The best method of estimating preexisting neuropsychological skill levels would be one that increases sensitivity and specificity while reducing false positives and false negatives (see Figure 1). It is difficult to make comparisons among the different estimates, however, because there is no way of knowing the actual preexisting cognitive skill levels as validation. One method of testing the effectiveness of individual comparison standards involves using a neurologically intact sample and calculating the instances of false positive misclassification (e.g., Gasquoine et al., 2007). This was the approach taken in the present study.

Research comparing the efficacy of individual comparison standards to the 50$^{th}$ percentile of published norms for estimating preexisting neuropsychological skill levels among Hispanic American participants is very limited, hence the importance of the current study. This study aims at determining the best method of estimating preexisting neuropsychological skill levels among a sample of neurologically intact bilingual Hispanic Americans. The three methods of estimating preexisting neuropsychological skill levels are (a) the 50$^{th}$ percentile of published norms, and two individual comparison standards: (b) WMLS-R Picture Vocabulary scores, and (c) WMLS-R Picture Vocabulary scores for subtests requiring predominantly language processing combined with WAIS-III Matrix Reasoning scores for subtests requiring predominantly visual-perceptual processing.

# CHAPTER III

# METHODS

## Participants

Twenty Spanish/English bilingual adults were recruited by word-of-mouth from different areas of the Rio Grande Valley region of South Texas. The culture of this region is a unique mixture of Mexican and American customs and language. Over 90% of its residents are Hispanic Americans, many of whom speak both Spanish and English fluently. Inclusion criteria included: (a) age range of 18 to 65 years; (b) Mexican heritage; (c) conversational fluency in both English and Spanish, as judged subjectively by their ability to carry on a conversation in both languages; and (d) no history of brain damage, psychiatric disorder, or drug or alcohol abuse.

The sample consisted of eleven (55%) males and nine (45%) females ranging from 19 to 62 ($M$ = 34.15; $SD$ = 11.33) years of age. Education level ranged from 9 to 19 ($M$ = 12.95, $SD$ = 2.35) years. Half of the participants were born in the United States, and the other half were born in Mexico. Those who were born in Mexico had lived in the United States from 4 to 37 ($M$ = 18.40; $SD$ = 8.26) years. All US-born participants were of Mexican heritage, as determined by ancestry.

The majority of the participants (13) were educated in the United States, 3 were educated in Mexico, and 4 in both countries. Spanish was the first language for most (16) of the participants. Eleven (55%) of the participants reported speaking mostly Spanish at home, while nine (45%) reported living in homes with English as the dominant language.

**Measures**

There were three measures of Spanish/English fluency: (a) self-report triad measures; (b) self-report Likert ratings of language competency; and (c) an objective achievement summary test score estimate of Broad Spanish and English Ability. Self-report triad measures of Spanish/English fluency were obtained by asking each participant to indicate their preferred language for communication among the triad: “English”, “Spanish”, or “Both.” Self-report Likert ratings were obtained by asking participants to rate their level of competency to speak and understand English and Spanish on a five-point Likert scale (from “minimal” to “high”). Self-report Likert ratings were collapsed into a single difference score by subtracting the sum of the English fluency ratings from the sum of the Spanish fluency ratings. Positive scores indicated greater Spanish proficiency. Four subtests (Picture Vocabulary, Verbal Analogies, Letter-Word Identification, and Dictation) of the Woodcock-Munoz Language Survey-Revised (WMLS-R; Woodcock et al., 2005) English Form A and Spanish version were administered to produce an objective achievement summary test score estimate of Broad Spanish and English Ability that have age corrected norms ($M$ = 100; $SD$ = 15). Broad English Ability scores were subtracted from Broad Spanish Ability scores to produce an achievement test difference score measure of language dominance. Positive scores indicated greater Spanish proficiency.

The Matrix Reasoning subtest of the Wechsler Adult Intelligence Scale-III (WAIS-III; Wechsler, 1997a) was used to obtain a measure of visual-perceptual intelligence in both Spanish and English. This subtest requires participants to make visual-perceptual matches of sameness and symmetry and solve visual-perceptual analogy problems. Scale scores from age-corrected norms (*M* = 10; *SD* = 3) were used. Instructions were translated into Spanish using standard back-translation techniques.

A neuropsychological test battery was administered in both Spanish and English. The Spanish battery (*La batería neuropsicológica en Español*: Artiola i Fortuni, Hermosillo, Heaton, & Pardee, 1999) consisted of eight subtests: (a) *Figure Memory*, adapted from Heaton, Grant, and Matthews (1991), using the three geometric figures from the Wechsler Memory Scale (WMS; Wechsler, 1945); (b) *Story Memory*, adapted from Heaton et al. (1991); (c) *Wisconsin Card Sorting Test (WCST)*, with the same format as Heaton, Chelune, Talley, Kay, & Curtis (1993); (d) *Spanish Letter Fluency*, which has the same format as the Word Fluency subtest of the Neurosensory Center Comprehensive Examination for Aphasia (Spreen & Benton, 1969) but uses the letters *P*, *M*, and *R*; (e) *Spanish Verbal Learning Test*, the format of which parallels the California Verbal Learning Test-II (Delis, Kramer, Kaplan, & Ober, 2000); (f) *Digit Span*, which has the same format as the WAIS-III subtest; (g) *Visual Span*, with the same format as the Spatial Span subtest of the WMS-III (Wechsler, 1997b); and (h) *Stroop Color and Word Test*, which has the same format as Golden and Freshwater (1998). Norms used were age- and education-corrected from the United States-Mexico border region as published in the manual, except for Digit Span, where WMS-III norms were used in both English and Spanish as the Batería Neuropsicológica norms have no English equivalent in either the WAIS-III or the WMS-III.

The English neuropsychological test battery consisted of the original eight tests upon which the *batería neuropsicológica* subtests were based. Verbal Letter Fluency used the letters *F*, *A*, and *S*. Digit Span and Visual Span sequences were from the WMS-III. Originally published norms were used for the California Verbal Learning Test-II, Digit Span, Wisconsin

Card Sorting Test, and Visual Span. Age-corrected meta-analytical norms (Mitrushina, Boone, Razani, & D'Elia, 2005) were used for Verbal Letter Fluency and the Stroop Color and Word Test. Story Memory and Figure Memory used the gender, age, and education corrected Caucasian norms from Heaton, Miller, Taylor, & Grant (2004), as there is no separate table for Hispanics.

A total of 19 measures were obtained from the eight subtests in each language: 3 from *Visual Memory* (trial 1 recall, learning, and delayed recall); 3 from *Verbal Memory* (trial 1 recall, learning, and delayed recall); 1 from *WSCT* (total errors); 1 from *Verbal Fluency* (FAS/PMR total); 7 from the *California Verbal Learning Test* (trial 1 recall, total words, list B recall, short delay free-recall, short delay cued-recall, long delay free-recall, and long delay cued-recall); 1 from *Verbal Attention* (digit span total); 2 from *Visual Attention* (number correct forward and number correct backward); and 1 from the *Stroop Color and Word Test* (number correct in the interference condition).

**Procedure**

Participants were tested over two sessions 27 to 49 ($M$ = 32.35, $SD$ =5.58) days apart. At each session, both the examiner and participant spoke only English or Spanish. Participants were randomly assigned to a language condition where half of the participants were tested in English during the first session and Spanish during the second and the other half was tested first in Spanish then English. Self-report measures of Spanish/English fluency and demographic data were only assessed during the first session in the assigned language. The WMLS-R, Matrix Reasoning, and neuropsychological test battery were administered in both sessions in the assigned language. The length of each session was approximately two hours.

False positive rates were calculated using published norms (Heaton et al., 2004) and individual comparison standards as estimates of preexisting skill levels. A false positive normative score was defined as a score ≤ 1*SD* below the mean of published norms. There were two individual comparison standards: (a) Picture Vocabulary scores from the WMLS-R, and (b) Picture Vocabulary for subtests requiring predominantly language processing (Verbal Memory, Verbal Fluency, California Verbal Learning Test, and Verbal Attention) and WAIS-III Matrix Reasoning scores for subtests requiring predominantly visual-perceptual processing (Visual Memory, Wisconsin Card Sorting Test, Visual Attention, and Stroop Color and Word Test). All scores were transformed to standard scores ($M$ = 0, $SD$ = 15) in order to make equal comparisons. False positive individual comparison scores were defined as scores ≤ 1*SD* below the Picture Vocabulary score for tests requiring primarily language processing and scores ≤ 1*SD* below the Matrix Reasoning score for tests requiring primarily visual-perceptual processing.

# CHAPTER IV

## RESULTS

Table 1 shows the number (percentage) of participants in the Spanish Dominant, Balanced Bilingual, and English Dominant groups for the three measures of Spanish/English fluency. For the self-report triad measure where participants were asked to indicate their preferred language of communication, those who answered "Spanish" were considered part of the Spanish Dominant group, those who answered "English" were considered part of the English Dominant group, and those who answered "Both" were considered part of the Balanced Bilingual group. For the self-report triad measure, there was no significant difference between the percentage of participants falling in the Spanish Dominant group (25%) and the English Dominant group (30%), $\chi^2(1, N = 20) = .45$, n.s. Self-report difference scores from the self-report Likert scale ranged from -3 to +6 (maximum = -20 to +20). Difference scores $\geq +3$ were considered Spanish Dominant, $\leq -3$ were considered English Dominant, and those between -2 and +2 were considered Balanced Bilingual. For the self-report Likert ratings of language competency, there was also no significant difference between the percentage of participants within the Spanish Dominant group (20%) and the English Dominant group (25%), $\chi^2(1, N = 20) = .56$, n.s. Broad Spanish and English Ability difference scores from the WMLS-R ranged from -28 to +27. Difference scores $\geq +11$ were considered Spanish Dominant, $\leq -11$ were considered English Dominant, and those from -10 to +10 were considered Balanced Bilingual. The percentages of participants in the Spanish Dominant and English Dominant groups for this third measure were exactly the same (15%). There were no significant differences among the percentages of participants in the Spanish Dominant [$\chi^2(2, N = 20) = 2.50$, n.s.], Balanced Bilingual [$\chi^2(2, N = 20) = 5.59$, n.s.], and English Dominant [$\chi^2(2, N = 20) = 5.01$, n.s.] groups for the three measures. This indicates a balance between both self-report and objective measures of language dominance. In fact, 13 participants (65%) correctly identified their objective WMLS-R group based on self-report of "English," "Spanish," or "Both" alone, and 17 participants (85%) correctly identified their objective WMLS-R group based on self-report Likert ratings.

Table 1

*Number and Percentage of Participants in Spanish Dominant, English Dominant, and Balanced Bilingual Groups as a Result of Three Measures of Spanish/English Fluency*

| | Self-report triad measures | Self-report Likert ratings (Difference score) | Broad Spanish and English Ability (WMLS-R) (Difference Score) |
|---|---|---|---|
| Spanish Dominant | 5 (25%) | 4 (20%) | 3 (15%) |
| English Dominant | 6 (30%) | 5 (25%) | 3 (15%) |
| Balanced Bilingual | 9 (45%) | 11 (55%) | 14 (70%) |

Table 2 shows the number of false positives obtained using the 50$^{th}$ percentile of published norms as the estimate of preexisting neuropsychological skill levels. Even though all participants were neurologically intact, test results showed 126 false positives on the 19 English measures and 68 false positives on the corresponding 19 Spanish measures. Both Spanish and English measures produced a larger number of false positives than the number expected for both languages (60.8) based on the 16$^{th}$ percentile of the normal distribution. The expected number is the percentage falling below 1*SD* of any normal distribution and was determined in this case by calculating 16 percent of the 19 measures for all 20 participants (i.e., Expected number = $19 \times 20 \times 16\% = 60.8$). The number of false positives (126) from the English measures was significantly higher than the number of false positives (68) from the Spanish measures, $\chi^2(1, N = 38) = 17.34, p < .01$.

Table 2

*False Positive Normative Scores Obtained from 19 Measures of the 8 Subtests of the Neuropsychological Test Battery in English and Spanish Using the 50th Percentile of Published Norms as the Estimate of Preexisting Skill Level*

| Subtest | Measure | False Positives (English) | False Positives (Spanish) |
|---|---|---|---|
| Visual Memory | Trial 1 recall | 2 | 1 |
| | Learning | 1 | 0 |
| | Delayed recall | 0 | 0 |
| Verbal Memory | Trial 1 recall | 16 | 4 |
| | Learning | 16 | 4 |
| | Delayed recall | 10 | 3 |
| WCST | Total errors | 9 | 2 |
| Verbal Fluency | FAS/PMR total | 9 | 12 |
| California Verbal Learning Test | Trial 1 recall | 8 | 4 |
| | Total Words | 4 | 6 |
| | List B recall | 10 | 4 |
| | Short delay free-recall | 5 | 4 |
| | Short delay cued-recall | 7 | 3 |
| | Long delay free-recall | 6 | 2 |
| | Long delay cued-recall | 6 | 3 |
| Verbal Attention | Digit Span Total | 7 | 12 |
| Visual Attention | Forward | 5 | 2 |
| | Backward | 0 | 1 |
| Stroop | Correct in interference | 5 | 1 |
| Total | | 126 | 68 |

Table 3 shows the mean Picture Vocabulary of the WMLS-R and WAIS-III Matrix Reasoning scores in both languages. For the Picture Vocabulary subtest, the average for all participants was approximately 1*SD* below the mean in both English and Spanish. The average participant score was above the mean (by approximately .5*SD*) on Matrix Reasoning in both

English and Spanish. One-way analysis of variance (ANOVA) found no significant difference between means of Picture Vocabulary scores in English and Spanish, $F(1, 38) = 2.15$, n.s. One-way ANOVA also found no significant difference between means of Matrix Reasoning scores in English and Spanish, $F(1, 38) < 1$, n.s.

Table 3

*Mean Picture Vocabulary and Matrix Reasoning Scores for all 20 Participants*

| Measure | Mean Scaled Score | Mean Standard Score |
|---|---|---|
| Picture Vocabulary (WMLS-R) | | |
| English | 86.6 | -0.9 |
| Spanish | 81.4 | -1.2 |
| Matrix Reasoning (WAIS-III) | | |
| English | 11.7 | 0.6 |
| Spanish | 11.4 | 0.5 |

Table 4 shows the number of false positives obtained using the participants' WMLS-R Picture Vocabulary score as the estimate of preexisting skill levels. Using the WMLS-R Picture Vocabulary score as the estimate of preexisting skill level decreased the number of false positives from 126 to 59 (53% less) on the English measures and from 68 to 8 (88% less) on the Spanish measures. Again, there were significantly more false positives in English (59) than in Spanish (8), $\chi^2(1, N = 38) = 38.82, p < .01$.

Table 5 shows the number of false positives obtained using the participants' WMLS-R Picture Vocabulary scores as the estimate of preexisting skill for subtests that required predominantly language processing skills (Verbal Memory, Verbal Fluency, California Verbal Learning Test, and Verbal Attention) and the WAIS-III Matrix Reasoning subtest scores as the estimate of preexisting skill for subtests that required predominantly visual-perceptual processing (Visual Memory, Wisconsin Card Sorting Test, Visual Attention, and Stroop Color and Word Test). Using both Picture Vocabulary scores and Matrix Reasoning scores decreased the number

Table 4

*False Positive Individual Comparison Scores Obtained from 19 Measures of the 8 Subtests of the Neuropsychological Test Battery in English and Spanish Using the Picture Vocabulary Score as the Estimate of Preexisting Skill Level*

| Subtest | Measure | False Positives (English) | False Positives (Spanish) |
|---|---|---|---|
| Visual Memory | Trial 1 recall | 2 | 0 |
| | Learning | 0 | 0 |
| | Delayed recall | 0 | 0 |
| Verbal Memory | Trial 1 recall | 13 | 0 |
| | Learning | 11 | 1 |
| | Delayed recall | 3 | 0 |
| WCST | Total errors | 5 | 0 |
| Verbal Fluency | FAS/PMR total | 3 | 1 |
| California Verbal Learning Test | Trial 1 recall | 4 | 1 |
| | Total Words | 3 | 1 |
| | List B recall | 2 | 0 |
| | Short delay free-recall | 1 | 0 |
| | Short delay cued-recall | 2 | 1 |
| | Long delay free-recall | 1 | 0 |
| | Long delay cued-recall | 1 | 0 |
| Verbal Attention | Digit Span Total | 2 | 2 |
| Visual Attention | Forward | 2 | 1 |
| | Backward | 0 | 0 |
| Stroop | Correct in interference | 4 | 0 |
| Total | | 59 | 8 |

of false positives from 126 to 83 (34% less) on the English subtests and from 68 to 25 (63% less) on the Spanish subtests compared to the use of the 50th percentile of published norms as the estimate of preexisting skill. The difference between the number of false positives obtained using Picture Vocabulary and Matrix Reasoning combined (83) versus Picture Vocabulary scores alone (59) did not reach statistical significance for the English measures [$\chi^2(1, N = 38) = 4.06$, n.s.] but did for the Spanish measures, $\chi^2(1, N = 38) = 11.11$, $p < .01$. Using Picture Vocabulary and Matrix Reasoning scores combined as the individual comparison standard also produced a statistically significant difference between the number of false positives in English (83) and Spanish (25), $\chi^2(1, N = 38) = 31.15, p < .01$.

Table 5

*False Positive Individual Comparison Scores Obtained from 19 Measures of the 8 Subtests of the Neuropsychological Test Battery in English and Spanish Using the WMLS-R Picture Vocabulary Score as the Estimate of Preexisting Skill Level for Verbal Subtests and the WAIS-III Matrix Reasoning Score as the Estimate of Preexisting Skill Level for Visual-Perceptual*

| Subtest | Measure | False Positives (English) | False Positives (Spanish) |
|---|---|---|---|
| Visual Memory^ | Trial 1 recall | 4 | 1 |
| | Learning | 2 | 0 |
| | Delayed recall | 0 | 0 |
| Verbal Memory* | Trial 1 recall | 13 | 0 |
| | Learning | 11 | 1 |
| | Delayed recall | 3 | 0 |
| WCST^ | Total errors | 9 | 4 |
| Verbal Fluency* | FAS/PMR total | 3 | 1 |
| California Verbal Learning Test* | Trial 1 recall | 4 | 1 |
| | Total Words | 3 | 1 |
| | List B recall | 2 | 0 |
| | Short delay free-recall | 1 | 0 |
| | Short delay cued-recall | 2 | 1 |
| | Long delay free-recall | 1 | 0 |
| | Long delay cued-recall | 1 | 0 |
| Verbal Attention* | Digit Span Total | 2 | 2 |
| Visual Attention^ | Forward | 8 | 6 |
| | Backward | 3 | 2 |
| Stroop^ | Correct in interference | 11 | 5 |
| Total | | 83 | 25 |

*Note.* * verbal tasks, ^ visual-perceptual tasks

Table 6 summarizes the overall findings. Chi-square analysis indicated statistically significant differences between the numbers of false positives in English and Spanish for all three estimates of preexisting skill: (a) 50th percentile of published norms, $\chi^2(1, N = 38) = 17.34$, $p < .01$; (b) Picture Vocabulary, $\chi^2(1, N = 38) = 38.82$, $p < .01$; and (c) Picture Vocabulary and Matrix Reasoning combined, $\chi^2(1, N = 38) = 31.15$, $p < .01$. In English, the number of false positives was significantly reduced by using Picture Vocabulary scores alone, $\chi^2(1, N = 38) = 24.26$, $p < .01$, and Picture Vocabulary and Matrix Reasoning scores combined, $\chi^2(1, N = 38) = 8.84$, $p < .01$, as the estimate of preexisting skill compared to the 50th percentile of published norms. The same was true in Spanish, where using Picture Vocabulary scores, $\chi^2(1, N = 38) = 47.37$, $p < .01$, and Picture Vocabulary and Matrix Reasoning combined, $\chi^2(1, N = 38) = 16.68$, $p < .01$, significantly reduced the number of false positives compared to the 50th percentile of published norms.

Table 6

*Summary of Total False Positive Scores Obtained Using Published Norms, Picture Vocabulary Scores, and Picture Vocabulary & Matrix Reasoning Scores Combined as the Estimates of Preexisting Skill Level*

| Method of Comparison | False Positives (English) | False Positives (Spanish) |
|---|---|---|
| 50th Percentile of Published Norms | 126 | 68 |
| Picture Vocabulary | 59 | 8 |
| Picture Vocabulary & Matrix Reasoning Combined | 83 | 25 |

# CHAPTER V

# DISCUSSION

In the field of clinical neuropsychology, the $50^{th}$ percentile of published norms is commonly used as an estimate of preexisting neuropsychological skill level, and scores below the $16^{th}$ percentile indicate cognitive impairment (Heaton et al., 2004). As Hispanic Americans perform below White Americans on neuropsychological tests (Gollan et al., 2002; Gasquoine et al., 2007), there is a higher false positive rate of cognitive impairment among Hispanic groupings. Another method of estimating preexisting neuropsychological skill level is the use of an individual comparison standard (e.g., Barona, Reynolds, & Chastain, 1984). This study used two individual comparison standards based on current verbal and visual-perceptual skill and aimed to determine which of these methods—the $16^{th}$ percentile of published norms or the individual comparison standard—was optimal for bilingual Hispanic American adults. This was done by comparing the number of false positives in a neurologically intact sample.

The sample included 20 bilingual Hispanic Americans of Mexican heritage who were all adults between the ages of 19 and 62 from the Rio Grande Valley region of South Texas, a region along the Mexican border where both Spanish and English are used in conversation. Participants had no history of brain damage, psychiatric disorder, or drug or alcohol abuse.

Three measures of Spanish/English fluency were used: (a) self-report triad; (b) self-report Likert ratings of language competency; and (c) an objective achievement summary test score. Self-report triad and Likert ratings of language competency corresponded with the achievement test difference score measure of language dominance for 65% and 85% of participants, respectively. As such, a self-report Likert scale may be sufficient in determining Spanish-English fluency, eviscerating the need for lengthy achievement test administrations.

Three methods of estimating preexisting neuropsychological skill levels were compared: (a) the 50th percentile of published norms; (b) WMLS-R Picture Vocabulary scores; and (c) WMLS-R Picture Vocabulary scores for subtests requiring predominantly language processing (Verbal Memory, Verbal Fluency, California Verbal Learning Test, and Verbal Attention) and WAIS-III Matrix Reasoning scores for subtests requiring predominantly visual-perceptual processing (Visual Memory, Wisconsin Card Sorting Test, Visual Attention, and Stroop Color and Word Test). False positives were determined from a neuropsychological test battery, administered in both Spanish and English, consisting of eight subtests of *La batería neuropsicológica en Español* and the original English tests upon which they were based. Scores on any of the 19 measures from the eight subtests falling at or below 1*SD* of each of the methods of estimating preexisting skill levels were considered false positives. As the sample consisted of neurologically intact adults, the optimum method would produce the minimum number of false positive identification of cognitive impairment (i.e., determining the presence of impairment where none exists).

Both methods using the individual comparison standards were found to be more effective than the commonly used 50th percentile of published norms as the number of false positives was significantly lower with these estimation methods in both English and Spanish. This is consistent with previous research on neurologically intact Hispanic Americans (Gasquoine et al., 2007). Of the two individual comparison standards, WMLS-R Picture Vocabulary scores alone as the estimate of preexisting skill were more effective at reducing false positive rates, as participants generally scored higher on Matrix Reasoning than Picture Vocabulary (see Table 3).

For all three methods of estimating preexisting skills in this study, there were more false positives in English than Spanish. This is important given the fact that most participants classified as Balanced Bilinguals for all three measures of Spanish/English fluency: (a) self-report triad measures (45%), (b) self-report Likert ratings of language competency (55%), and (c) WMLS-R Broad Spanish and English Ability (70%). There was also no significant difference between the percentages of Spanish Dominant versus English Dominant participants for any of the three measures. Additionally, there were no significant differences between mean Picture Vocabulary and Matrix Reasoning scores in English versus Spanish. As the distribution of Spanish language norms falls below English language norms, the use of Spanish measures significantly reduces the number of false positives in this sample of neurologically intact bilingual Hispanic American participants.

Reducing the number of false positives typically increases the chance of false negatives (true cognitive impairment not detected by neuropsychological assessment). This study used only neurologically intact participants; therefore, false negative rates could not be calculated. In order to test for false negatives, a brain-injured sample must be used. This is an important consideration for future research in this topic.

## REFERENCES

Acevedo, A., Loewenstein, D. A., Barker, W. W., Harwood, D. G., Luis, C., Bravo, M., et al. (2000). Category Fluency Test: Normative data for English- and Spanish-speaking elderly. *Journal of the International Neuropsychological Society*, *6*, 760-769.

Arnold, B. R., Montgomery, G. T., Castaneda, I. & Longoria, R. (1994). Acculturation and performance of Hispanics on selected Halstead-Reitan neuropsychological tests. *Assessment, 1,* 239-248.

Artiola i Fortuni, L., Hermosillo, D., Heaton, R. K., & Pardee, R. E. (1999). *Manual de normas y procedimientos para la batería neuropsicológica en español*. Tucson, AZ: m Press.

Baade, L. E., & Schoenberg, M. R. (2004). A proposed method to estimate premorbid intelligence utilizing group achievement measures from school records. *Archives of Clinical Neuropsychology, 19,* 227-243.

Babcock, H. (1930). An experiment in the measure of mental deterioration. *Archives of Psychology, 117,* 1-105.

Barona, A., & Chastain, R. L. (1986). An improved estimate of premorbid IQ for Blacks and Whites on the WAIS-R. *International Journal of Clinical Neuropsychology, 8,* 167-173.

Barona, A., Reynolds, C. R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology, 52,* 885-887.

Bialystok, E., Craif, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, ad Cognition*, *34*, 859-873.

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *Clinical Neuropsychologist, 3,* 129-136.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review, 13,* 181-197.

Crawford, J. R., & Allen, K. M. (1997). Estimating premorbid WAIS-R IQ with demographic variables: regression equation derived from a UK sample. *Clinical Neuropsychologist, 11,* 192-197.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test-II: Manual*. San Antonio, TX: The Psychological Corporation.

Gasquoine, P. G. (1999). Variables moderating cultural and ethnic differences in neuropsychological assessment: The case of Hispanic Americans. *Clinical Neuropsychologist, 13,* 376-383.

Gasquoine, P. G. (2001). Research in clinical neuropsychology with Hispanic American participants: A review. *Clinical Neuropsychologist*, *15*, 2-12.

Gasquoine, P. G. (2008). *Psychology of Multicultural Competence: Critical Issues.* Deer Park NY: Linus Publications.

Gasquoine, P. G. (2009). Race-norming of neuropsychological tests. *Neuropsychology Review, 19*, 250-262.

Gasquoine, P. G., Cavazos, A., Cantu, J., & Weimer, A. A. (2010). Bilingualism and Hispanic American intelligence test scores. In E. F. Caldwell (Ed.), *Bilinguals: Cognition, Education, and Language Processing*. NY: Nova Science Publishers, pp.181-199.

Gasquoine, P. G., Croyle, K. L., Cavazos-Gonzalez, C, & Sandoval, O. (2007). Language of administration and neuropsychological test performance in neurologically intact Hispanic American bilingual adults. *Archives of Clinical Neuropsychology, 22,* 991-1001.

Gladsjo, J.A., Schuman, C. C., Evans, J. D., Peavy, G. M., Miller, S. W., & Heaton, R. K. (1999). Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment*, *6*, 147-178.

Golden, C. J., & Freshwater, S. M. (1998). *The Stroop Color and Word Test: A manual for clinical and experimental uses.* Wood Dale, IL: Stoelting.

Gollan, T. H., Montoya, R. I., & Wener, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology, 16,* 562-576.

Green, R. E. A., Melo, B., Christensen, B., Ngo, L.-A., Monette, G., & Bradbury, C. (2008). Measuring premorbid IQ in traumatic brain injury: An examination of the validity of the Wechsler Test of Adult Reading (WTAR). *Journal of Clinical and Experimental Neuropsychology, 30,* 163-172.

Gurland, B. J., Wilder, D. E., Cross, P., Teresi, J., & Barrett, V. W. (1992). Screening scales for dementia: Toward reconciliation of conflicting cross-cultural findings. *International Journal of Geriatric Psychiatry*, *7*, 105–113.

Harvey, P. D., Friedman, J. I., Bowie, C., Reichenberg, A., McGurk, S. R., Parrella, M., et al. (2006). Validity and stability of performance-based estimates of premorbid educational functioning in older patients with schizophrenia. *Journal of Clinical and Experimental Neuropsychology, 28,* 178-192.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an expanded Halstead-Reitan battery*. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan battery: Demographically adjusted neuropsychological norms for African American and Caucasian Adults.* Odessa, FL: Psychological Assessment Resources.

Jastak S., & Wilkinson G. S. (1984). *The Wide Range Achievement Test–Revised Administration Manual.* Wilmington, DE: Jastak Associates.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). NY: Oxford University Press.

McGurn, B., Starr, J. M., Topfer, J. A., Pattie, A., Whiteman, M. C., Lemmon, H. A., et al. (2004). Pronunciation of irregular words is preserved in dementia, validating pre-morbid IQ estimation. *Neurology*, *62,* 1184-1186.

Mitrushina, M., Boone K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (2nd ed.). NY: Oxford University Press.

Puente, A. E., & Salazar, G. D. (1998). Assessment of minority and culturally diverse children. In A. Prifitera, & D. H. Saklofske (Eds.), *WISC-III: Clinical use and interpretation* (pp. 227–248). San Diego, CA: Academic.

Rosselli, M. A., Ardila, A., Santisi, M. N., Del Rosario Arecco, M., Salvatierra, J., Conde, A., et al. (2002). Stroop effect in Spanish-English bilinguals. *Journal of the International Neuropsychological Society*, *8*, 819-827.

Schretlen D. J., Winicki J. M., Meyer S. M., Testa S. M., Pearlson G. D., & Gordon B. (2009). Development, Psychometric Properties, and Validity of the Hopkins Adult Reading Test (HART). *The Clinical Neuropsychologist, 23*(6), 926–943.

Spreen, O., & Benton, A. L. (1969). *Neurosensory Center Comprehensive Examination for Aphasia*. Victoria, BC: Neuropsychology Laboratory, University of Victoria.

Stricks, L., Pittman, J., Jacobs, D. M., Sano, M., & Stern, Y. (1998). Normative data for a brief neuropsychological battery administered to English- and Spanish-speaking community-dwelling elders. *Journal of the International Neuropsychological Society*, *4*, 311-318.

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19*, 87-95.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale-III.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale-III*. San Antonio, TX: Psychological Corporation.

Woodcock, R. W., Munoz-Sandoval, A. F., Ruef, M. L., & Alvarado, C. G. (2005). *Woodcock-Munoz Language Survey-Revised.* Itasca, IL: Riverside Publishing.

## BIOGRAPHICAL SKETCH

Cassandra Dayanira Gonzalez is a native of Tamaulipas, Mexico, who obtained most of her schooling in the Rio Grande Valley region of South Texas. In 2006, she graduated as valedictorian of her high school class from South Texas Business, Education, & Technology Academy in Edinburg, Texas. Three years later, she obtained a Bachelor of Science in Psychology from The University of Texas-Pan American (UTPA), earning top honors (*Summa Cum Laude*). Cassandra continued her graduate education at UTPA and earned a Master of Arts in Experimental Psychology in December of 2010. Ms. Gonzalez worked as a graduate assistant for UTPA throughout her graduate career, and she hopes to continue working with students in a university setting. Her permanent mailing address is as follows:

Cassandra Dayanira Gonzalez

4305 Canadiana Lane

Mission, TX 78572-9449