2010

Interpreter-mediated neuropsychological testing of monolingual Spanish speakers: does it have an effect on test scores?

Rachel Nichole Casas
University of Iowa

INTERPRETER-MEDIATED NEUROPSYCHOLOGIAL TESTING OF MONOLINGUAL SPANISH SPEAKERS: DOES IT HAVE AN EFFECT ON TEST SCORES?

by

Rachel Nichole Casas

An Abstract

Of a thesis submitted in partial
fulfillment of the requirements for
the Doctor of Philosophy degree in Psychology
in the Graduate College of
The University of Iowa

December 2010

Thesis Supervisor: Professor Daniel Tranel

ABSTRACT

Nearly 13.8 million Hispanics in the United States speak English "less than very well." This has important implications for the field of clinical neuropsychology. Patients who do not speak English fluently are being increasingly referred for neuropsychological services, and many of these individuals are assessed with the aid of language interpreters. However, whether or how the use of an interpreter has an effect on neuropsychological test scores is not known. For lack of a better alternative, it generally is assumed that the test data obtained through an interpreter are a valid indication of the patient's cognitive functioning, but with almost no empirical support, this assumption appears tenuous at best. The effect of an interpreter, in fact, could be substantial, making this issue all the more deserving of rigorous investigation. The primary objective of the current study was to determine whether using an interpreter to conduct neuropsychological testing of monolingual Spanish speakers had an effect on the neuropsychological test scores. Participants included 40 neurologically normal Spanish-speakers with limited English proficiency, ages 18-65 years, (M= 39.65, SD =13.91) who completed a 2-hour battery of verbal and nonverbal neuropsychological tests both with and without an interpreter. The condition of test administration was counterbalanced across participants and test score differences between the two conditions were compared. Results indicated that use of an interpreter significantly affected mean scores for some neuropsychological tests from the verbal modality. Also, variability in test scores generally was higher when an interpreter was used, significantly so for one verbal test. Results of this study contribute to the extant literature concerning the use of interpreters to facilitate neuropsychological testing of individuals with limited English proficiency. Specifically, they indicate that neuropsychologists should avoid interpreter use and refer patients to bilingual clinicians whenever possible. For situations in which this may not be a viable option, neuropsychologists should limit their test batteries to measures that require minimal reliance on the interpreter. Tests that rely almost entirely on interpreter skills for

administration and scoring – such as the Vocabulary and Similarities subtests of the Wechsler Adult Intelligence Scale (WAIS) – should probably not be used. Larger confidence intervals should be used when interpreting observed scores from interpreter-mediated neuropsychological tests.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

INTERPRETER-MEDIATED NEUROPSYCHOLOGIAL TESTING OF MONOLINGUAL SPANISH SPEAKERS: DOES IT HAVE AN EFFECT ON TEST SCORES?

by

Rachel Nichole Casas

A thesis submitted in partial
fulfillment of the requirements for
the Doctor of Philosophy degree in Psychology
in the Graduate College of
The University of Iowa

December 2010

Thesis Supervisor: Professor Daniel Tranel

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D THESIS

_____

This is to certify that the Ph.D thesis of

Rachel Nichole Casas

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Psychology at the December 2010 graduation.

Thesis Committee: _____
Daniel Tranel, Thesis Supervisor

_____
Lee Anna Clark

_____
James Marchman

_____
Kristian Markon

_____
William Liu

To my parents, Ann Marie and Carlos

ACKNOWLEDGEMENTS

I am grateful to all of the individuals who volunteered their time to participate in this study. I would also like to thank the outstanding team of research assistants from Proyecto Neuro whose dedication and talent were the lifeline for this study: Edmarie Guzmán Vélez, Javier Fernando Rodriguez, Nayra Rodriguez, Gabriela Quiñones, and Adam Arnold. This study would also not have been made possible without the generosity and hospitality of Guillermo Bernal and the faculty and staff of the Istituto de Investigación Psicológica (IPsi).

In addition, I am grateful for the ongoing guidance of my graduate advisor, Daniel Tranel, whose mentorship has been a constant source of support. I am also thankful for Jill Razani, Carrie Saetermoe, and Jennifer Tushla who planted seeds of inspiration early in my education and continued to nurture them to bloom.

I would especially like to thank Peter Guilarte for his unconditional love and enduring patience. He's been my steadfast champion, partner, and forever friend.

Finally, a special thanks to all of the programs, people, and funding agencies who supported my professional development and research endeavors, including: the National Institute of Mental Health's Career Opportunities in Research Program (COR); the University of California, San Diego's Summer Training Academy for Research in the Sciences (STARS); the University of Iowa's Summer Research Opportunities Program (SROP); the American Psychological Association's Diversity Program in Neuroscience (DPN); and the National Science Foundation's Graduate Research Fellowship Program (GRFP).

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

The United States is a racially, ethnically and culturally diverse country. Data from the 2000 Census indicated that nearly 25% of the population identified themselves as being of a race other than "white" or of more than one race. Of these groups, Hispanics represent the largest and fastest growing segment of the ethnic minority population, demonstrating a 57.9% increase in their numbers from 1990 to 2000. Approximately 43.5% of the Hispanic population lived in the Western United States in 2000, with estimates as high as 42.1% in states such as New Mexico (by contrast, Mississippi had the largest percentage of African-Americans at about 36% of that state's population). A large number of Hispanics were also living in several states traditionally without Hispanic populations, such as Georgia, Iowa, and Arkansas, and in certain counties within those states they represented as much as 24.9% of the total county population. The Hispanic population is continuing to grow at a rapid rate. To wit, the most recent projections from the United States Census Bureau estimate that as many as 102.6 million Hispanics will be living in the country by 2050.

The science and practice of clinical neuropsychology have been affected by the changing United States demographics. Neuropsychologists are being increasingly called upon to conduct assessments with Hispanic individuals who suffer from cerebrovascular disease, dementia, traumatic brain injuries, epilepsy, movement disorders, brain tumors, brain infections, and various other neurological conditions. In addition, Hispanics are at greater risk for diseases such as diabetes and heart disease (Stern & Mitchell, 1995; Goff et al., 1997), which are also known to affect cognitive functioning (Pavlik et al., 2005) and increase vulnerability for neurological disease. It is probably not controversial to speculate that under ideal conditions it may be beneficial for Hispanic patients to be evaluated by Hispanic neuropsychologists, especially when the patient has limited English proficiency (LEP). However, Hispanics are grossly underrepresented within the

field of neuropsychology. According to the American Psychological Association (Bailey, 2004), only seven percent of all individuals earning their Ph.Ds in psychology in 2002 were Hispanic. The percentage drops even lower for Hispanic psychologists who complete the additional educational and training requirements to become board certified clinical neuropsychologists, and lower still for those Hispanic neuropsychologists who are bilingual speakers of Spanish and English. Furthermore, even Hispanic neuropsychologists who are bilingual speakers of Spanish and English may not be equally fluent in each of their languages. It is not uncommon for Hispanics to speak "kitchen Spanish," or enough Spanish to communicate easily with friends and family, but not in professional contexts that require specialized vocabularies and terminology (Artiola y Fortuny et al., 1998; Herman & Hermosillo, 1998).

In 1997 Echemendia and colleagues conducted a national survey to determine the extent to which American neuropsychologists provided services to Hispanic clients. The results indicated that neuropsychologists spent approximately 28% of their time providing services to ethnic minority patients, including Hispanics. However, the overwhelming majority of these neuropsychologists admitted to having had little training for working with diverse clients. Nearly 90% of the respondents indicated that their graduate training program did not provide a course on neuropsychological assessment of individuals from "culturally different" backgrounds. Regarding Hispanics in particular, results indicated that of the neuropsychologists who were surveyed, each of them received an average of 14 referrals per year for Hispanic patients. However, the modal number of evaluations conducted with Hispanic patients was zero, which prompted a re-examination of the data. With the top 5% of the distribution removed from the analysis, the number of Hispanic patients evaluated each year per neuropsychologist was reduced to 6, suggesting that most neuropsychologists had little or no experience in working with Hispanic patients. Additionally, 22% of the sample identified themselves as "not at all

competent" to work with Hispanics, and the majority identified themselves as somewhere between "somewhat competent" and "not at all competent."

Mental Health Disparities and Hispanics

The success of any profession depends upon its ability to meet the demands of its consumers, and the future of clinical neuropsychology is surely not any different. To the extent that the Hispanic population continues to grow, the field of neuropsychology needs to address the emerging needs of Hispanic clients, a demand which research suggests the field is currently failing to meet (Echemendia, 2004). In addition to the results from the Echemendia et al. (1997) survey, it is also a well-replicated research finding that Hispanic individuals underutilize mental health services (Sue et al., 1973, 1978, 2003). When they do seek mental health treatment they also tend to have high premature termination rates, at least for traditional treatments such as psychotherapy. For example, Sue (1978) found that 42% of Mexican-American clients dropped out of treatment after one session. Members of other ethnic minority groups, such as Asian-Americans, also underutilize mental health services compared to the general population (Sue, D. & Sue, S., 1987; Zhang et al., 1998). Only 34.1 % of all Asian-Americans who had a probable *DSM-IV* diagnosis during a 12-month period sought services compared with 41.1% of all individuals with a probable *DSM-IV* diagnosis during the same time period (Abe-Kim et al. 2007).

The fact that Hispanic individuals underutilize mental health services cannot be explained by lower rates of psychopathology in this population. For example, the prevalence of mental illness among Mexican-Americans tends to be similar to whites (Vega et al., 1998). It has even been suggested that Hispanics, at least those born in the United States, are at increased risk for certain psychological disorders, such as depression and anxiety (Alegria et. al., 2007). Indeed, acculturation has sometimes been identified as a mediator of mental illness among Hispanic individuals. Recent research by Hernandez,

Plant, Sachs-Ericsson, and Joiner (2004) also found that Hispanics were more likely to have met criteria for a psychiatric disorder in the past year when compared to whites.

The Surgeon General has determined that barriers in access to care have at least partially contributed to the mental health disparities between Hispanics and whites (U.S Department of Health and Human Services, 2001). Practical constraints are among the most critical factors contributing to the disparity in mental health use for Hispanic individuals. Limited hours of clinic operation, lack of accessible transportation to and from treatment centers, and the unavailability of clinicians who speak the client's language are some of the biggest barriers (Kline et al., 1980, Perez-Arce et al., 1984, Sue et al., 1974, 1981, Takeuchi et al., 1995). Traditional psychotherapy, for example, is typically conducted during business hours at a private clinic or hospital and in English. Many ethnic minority individuals live in poverty, work during the day, and do not own their own vehicles, which makes scheduling appointments difficult and dependent upon public transportation and/or family and friends. Child care also becomes an obstacle for many ethnic minority clients who do not have the financial resources to pay for a babysitter even if the psychotherapy fee has been adjusted to meet their financial needs. Moreover, brochures and advertisement materials for mental health services and centers are not uncommonly written beyond the literacy level of many ethnic minority clients, some of whom have limited educational histories and/or who may speak English as a second language if at all.

However, practical constraints are not the only barriers contributing to the mental health disparities dilemma. Some researchers have suggested that ethnic minority clients, including Hispanics, may be more apt to receive discriminatory forms of treatment. For example, Arroyo (1996) showed videotaped intake interviews of the same client to non-Hispanic white therapists. In the first condition, an actress portrayed a client who spoke perfect English and had a light skin complexion. In the second condition, an actress portrayed a client who spoke English with a Spanish accent and had a dark skin

complexion. The content of the interviews was identical; however, the client in the second condition was rated as having a poor prognosis for treatment. Indeed, therapist preferences for clients may place ethnic minority individuals at a disadvantage. Schofield (1964) argues that therapists tend to exhibit "YAVIS" syndrome, or a preference for working with young, attractive, verbal, intelligent, and successful individuals. The typical "YAVIS" patient in the United States is most likely to come from white, educated, and middle to upper class backgrounds. Other researchers have written about the impact of so-called "microaggressions," or insults (verbal, nonverbal, and/or visual) directed toward people of color, often automatically or unconsciously" (Sue et al., 2007). While a single microaggression may seem innocuous, their cumulative effects over time may lead to a number of adverse consequences for physical and mental well-being. To the extent that microaggressions are repeatedly committed and unaddressed throughout the context of psychotherapy or other types of mental health services, ethnic minority clients may receive sub-optimal treatment and care. For example Muhkerjee and colleagues (1983) found that Hispanics with Bipolar disorder were frequently misdiagnosed as having Schizophrenia. Similar findings have been reported for African-American clients presenting with depressive symptoms (Baker, 2004). In addition, African-American clients are more likely to be hospitalized or committed to state mental institutions compared to white patients (Lawson et al., 1994).

Considering the issues presented in the aforementioned discussion, some researchers and clinicians have argued that the ethnic match between a patient and their mental health provider can affect diagnostic outcomes and clinical impressions. For example, ethnic match has been associated with higher therapist-rated psychological functioning scores for African-American and Asian-American clients who were participating in psychotherapy (Russell et al., 1996). In another study, Takeuchi and colleagues (1995) compared the return rate, length of treatment, and treatment outcomes of ethnic minority clients who received services from either ethnicity-specific (more than

50% of clients were from a specific ethnic group) or mainstream (majority of the patients were white) mental health programs. They found that African-Americans, Asian-Americans, and Hispanic-Americans who received treatment from ethnicity-specific mental health programs demonstrated higher return rates and stayed in treatment longer than ethnic minority clients who were treated in mainstream treatment programs. It is therefore plausible that Hispanic clients may be especially drawn to ethnicity-specific psychological services depending upon their immigration status, English language proficiency, acculturation level, and generation status in the United States.

The Term "Hispanic"

At this point it is important to discuss the use of the term "Hispanic," which is not without controversy. For instance, use of the word "Hispanic" has been criticized by researchers, clinicians, and social activists as an umbrella term that masks both the racial and ethnic heterogeneity of the extremely diverse individuals (Bernal, 1994). The term was coined and introduced by the United States government in the 1970s as the official nomenclature for identifying individuals from Spanish-speaking countries. It was used during data collection for the census at that time and has remained ever since. In fact, if one were to venture outside of the United States, he or she would be hard pressed to find anyone who identified themselves as "Hispanic" (Artiola y Fortuny, 1998). An additional problem with the term is the fact that "Hispanics" come from a variety of countries in South and Central America as well as Europe. They can also be of any racial background – Indigenous, African, Asian, or European – and yet use of the term "Hispanic" places an artificial emphasis on European ancestry from Spain (McLaren, 1995). Given the history of oppression and subjugation of indigenous and African people by European conquistadors in the United States, use of the term is understandably offensive to many. Therefore, I will use– Latina or Latino – throughout the remainder of this proposal when referring to the diversity of peoples from Latin America currently living in the United States. Whenever possible, I will use more specific labels that clearly identify the specific

ethnicity of the Latino group being discussed, such as Mexican-American, Cuban-American, and so forth.

Race, Ethnicity, and Culture

The terms "race" and "ethnicity" are also frequently misunderstood and misused. They are often used interchangeably in the scientific literature and in conversation among the general population, yet they have distinct meanings. "Race" refers to differences in physical characteristics between groups of people, such as differences in skin color, hair texture, and body proportions. "Ethnicity," on the other hand, refers to more socially and behaviorally motivated distinctions between groups based on variables such as language dialects and customs (Pfeffer, 1998). Despite the fact that Latinos are heterogenous with respect to race and ethnicity, they tend to share some features of a common "culture," which is considered to be a more broad and overarching concept encompassing the values, beliefs, and behavioral styles shared by members of a society. Importantly, culture and ethnicity are not mutually exclusive constructs, and in fact, they are often inter-related. Essentially, culture embodies specific ways of living for a particular group.

Several researchers have highlighted the effects of culture on cognition and neuropsychological testing. Ardila (2005) for example, has argued that neuropsychological testing is a social construct, and that like all social constructs, it is embedded with value-laden assumptions based upon a particular culture which determines what is and is not relevant to the assessment. As an example, he states that cognitive evaluations assume that a special type of communication will occur between the examiner and the patient which is directly related to an intimacy-formality dimension of interpersonal behavior. There are also assumptions about what constitutes "best performance" and the perceived importance of time, particularly for speeded tests. For the most part, neuropsychological assessment procedures have been developed and standardized within industrialized, Western society. If such measures are used to assess individuals with cultural experiences and backgrounds foreign to those in which the

assessment methods were developed, then they may be ineffective and/or inappropriate. It makes sense then that both the culture in which the test was created and the cultural group of the test taker could affect performance on neuropsychological tests. Such is the case for many Latinos living in the United States whose cultural experiences are often very different than individuals from the majority culture.

Education, Illiteracy, and Latinos

Importantly, culture is not the only determinant of human behavior and cognition. It interacts with various other variables, such as educational, economic, political, environmental, and situational influences to name but a few. Some of these variables capture factors that are particularly relevant to the neuropsychological assessment of Latino patients living in the United States, as they have been shown to substantially alter the interpretation of test results as well as the psychometric properties of neuropsychological instruments (Ponton 2001.; Artiola y Fortuny et al., 1998; Perez-Arce et al., 1984). Such factors include low levels of education, illiteracy, lack of health insurance, impoverished living conditions, illegal immigration and/or refugee status, and acquiring English as a second language.

Education is one of the most important variables to consider during the neuropsychological assessment of Latinos. In the United States, access to free, public education is considered a basic right for every citizen. The laws specifying when a child must begin attending school and when they can leave are determined by each state, but in general, children begin attending Kindergarten between the ages of 4-5 and must continue their schooling until ages 16-18. As a result, US-born citizens tend to be fairly well educated. According to the findings from the 2003 Annual Social and Economic Supplement (ASEC) to the Current Population Survey (CPS), 85% of the adult population ages 25 or older reported that they had completed at least high school; 65% of the population ages 75 or older had completed high school. Notably, only 67% of persons born outside of the United States had a high school diploma, and only 57% of the Latino

population had one. Differences between the percentage of non-Hispanic whites and Latinos who held a high school diploma were significant regardless of whether or not Latinos were born in a foreign country or in the United States, however the differences were even more pronounced for foreign-born Latinos.

Many neuropsychological tests were created by and for relatively well educated individuals. The majority of the extant normative data for these tests, which are used to classify "impairment" and to help detect the presence and extent of brain damage, were collected on samples of individuals who have obtained at least an eighth grade level of education (Ponton et al., 1999). However, many Latinos, especially those who were born outside of the United States and in countries where access to education is not a basic right, but rather, a privilege of the upper classes who can afford to send their children to school, have not obtained an eighth grade education. For many Latinos living in the United States, especially those who work in agriculture or unskilled labor factories, obtaining a high level of education also may not be a priority because it is not considered essential for a successful life (Artiola y Fortuny et al., 1998). Research investigating the effects of education level and quality on neuropsychological test performance suggests that meaningful differences exist. Individuals with extremely low levels of education tend to have substantially lower scores compared to individuals with at least an eighth grade level of education (Wiederholt et al., 1993; Ardila, 1995; Ostrosky-Solis, 2004)

Closely related to the issue of low education is the issue of illiteracy in the Latino population. The literacy rate in the United States is extremely high overall – 99% according to the United States Census Bureau (2000), but the literacy rate for Latinos is substantially lower. What's more, the literacy rate for Latinos has actually been declining since 1992. Considering the fact that literacy and level of education are highly correlated, and Latinos overall have had less access to formal education, this is not surprising. Manly and colleagues (1999) have found that literacy rates can affect neuropsychological test performance. Normal individuals with lower literacy levels obtain lower scores on tests

of naming, comprehension, verbal abstraction, orientation, figure matching, and recognition, even after controlling for language of test administration for individuals who speak English as a second language.

Limited English Proficiency Speakers

Arguably the most difficult and complex barrier to neuropsychological assessment of Latinos is the linguistic factor. Although the language use and linguistic patterns of Latinos vary considerably, many of them speak Spanish as their dominant or only language. In fact, the most recent data from the United States Census Bureau (2000) estimated that 28.1 million people speak Spanish at home. Thus, although Spanish speakers have traditionally been considered linguistic minorities, a substantial portion of the United States population now speaks Spanish. Moreover, nearly half of these people - 13.8 million – reported that they speak English "less than very well."

The fact that there simply are not enough bilingual Spanish-speaking neuropsychologists to provide services to Latino clients with LEP has already been highlighted. Thus, many monolingual English-speaking neuropsychologists are being called upon to evaluate Spanish-speaking clients with various levels of English language proficiency, often with assistance from a language interpreter.

In fact, there is evidence that language interpreter services are being increasingly utilized in healthcare settings. A survey conducted in 1995 by the National Public Health and Hospital Institute (NPHHI) revealed that 11% of all patients seen at 83 public and private hospitals required an interpreter to communicate with medical providers. Strikingly, one third of these hospitals indicated that at least 27% of their patients required the use of an interpreter to facilitate communication with providers (Ginsberg et al., 1995). Several years ago LaCalle (1987) reported that in California 25% of psychological and psychiatric forensic evaluations were conducted by a clinician who was not fluent in the patient's primary language. One could hypothesize that these numbers are even higher in 2010 given the substantial increase in the Latino population

in California over the last two decades. Regarding the local situation at the University of Iowa Hospitals and Clinics, it is estimated that a neuropsychological evaluation requiring the use of a Spanish-English interpreter is conducted some 20 to 30 times per year.

There is also evidence that patients believe that interpreters can facilitate communication between themselves and their health care providers. Baker and colleagues (Baker, Hayes, & Fortier, 1998) conducted a survey of 457 patients being treated in an emergency department of a public hospital. Twenty-two percent of these patients did not use an interpreter to communicate with their health care providers but indicated that they believed they should have.

Controversy of Interpreter Use

The effect of using a bilingual interpreter to facilitate the neuropsychological assessment of linguistic minorities has not been empirically investigated. However, use of an interpreter in neuropsychological evaluations of non-English speakers has been challenged on ethical grounds. Artiola y Fortuny and colleagues (1998), for example, have argued that using interpreters can bias the assessment process for a number of reasons. One of their primary concerns surrounds the fact that a monolingual neuropsychologist could not evaluate the language proficiency of the interpreter. The neuropsychologist in this situation would therefore be unable to verify the accuracy of the translations made during the assessment process, and yet important clinical diagnoses and treatment decisions will be based upon the information obtained from this process. One could argue that such ambiguity challenges the competency of the neuropsychologist. Standard 2.01 of the APA Ethical Principles and Code of Conduct (2002) asserts that "Psychologists provide services, teach, and conduct research with populations and in areas only within the boundaries of their competence, based on their education, training, supervised experience, consultation, study, or professional experience." It surely seems possible, if not even likely, that monolingual English-speaking clinicians who attempt to conduct neuropsychological evaluations of Spanish-speaking patients who do not speak

English fluently fail to meet the spirit of these guidelines, even if the assessment is conducted through an interpreter. However, the issue is not that clear cut, and there are other aspects of the APA ethics code which suggest that use of an interpreter may be justified under certain situations and circumstances. For example, Standard 2.02 states that psychologists can "… provide services to individuals for whom other mental health services are not available and for which psychologists have not obtained the necessary training. Psychologists may provide such services in order to ensure that services are not denied." Such is the case for many Spanish-speaking clients, especially those clients seeking services in rural areas or in parts of the South or Midwest, where the nearest Spanish-speaking clinician may be many hours away and for all intents and purposes, inaccessible.

The APA currently does not have an official and legally-binding policy addressing the issue of language competency with respect to clinicians. However, the APA Board of Ethnic Minority Affairs (BEMA) established a Task Force on the Delivery of Services to Ethnic Minority Populations in 1988, which henceforth developed a set of *Guidelines for Providers of Psychological Services to Ethnic, Linguistic, and Culturally Diverse Populations*. The preamble for the guidelines asserts that they are to be "aspirational" in nature and to provide "suggestions" for psychologists working with diverse populations. Among their suggestions is that "Psychologists interact in the language requested by the client and, if this is not feasible, make an appropriate referral." Preferably, the psychologist would be sufficiently fluent in the client's preferred language to conduct the assessment directly; however when this is not possible the guidelines specifically recommend that "psychologists offer the client a translator with cultural knowledge and an appropriate professional background." Furthermore, the guidelines state that, "when no translator is available, then a trained paraprofessional from the client's culture is used as a translator/culture broker." The guidelines do not specify how

to determine the competency of the linguistic abilities of either the psychologist or the interpreter.

The U.S. Government and Interpreter Services

The United States government has also implemented a set of guidelines and recommendations for non-English speaking individuals. The Office of Minority Health of the United States Department of Health and Human Services has published 14 National Standards on Culturally and Linguistically Appropriate Services (CLAS) (2001). Four of these standards are now federal mandates. Thus, psychologists may be held accountable to laws established by the United States government in some situations regarding the assessment of individuals with LEP. Current legislation mandates that interpreter services be made available to any patient who receives treatment from a health care provider who receives federal funding, should the need for such services arise. It further states that language assistance services are not contingent upon the amount of federal funding the provider receives and that the provider is responsible for making the patient aware of his or her right to these services. Thus, psychologists or neuropsychologists who work in clinics or hospitals that accept federal funding -which includes government subsidized insurance such as Medicare and Medicaid- are also bound by these regulations.

The federal laws requiring language assistance services for individuals with limited English proficiency (LEP) who seek services from government-funded health care providers stem as far back as Title VI of the Civil Rights Act of 1964. This acts states that "No person in the United States shall, on grounds of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving federal financial assistance." Non-English speakers or LEP speakers, who because of their national origin are at a significant disadvantage for communicating, would certainly be denied the benefits of quality interactions with their health care providers without the assistance of an interpreter. Some states have enacted additional laws to protect LEP speakers. California,

for example, implemented the Dymally-Alatorre Bilingual Services Act (Snowden et al., 2007), to ensure that its residents are able to secure language assistance services whenever necessary when communicating with public agencies. The act extends to both oral interpretation and written translation and states that "…the effective maintenance and development of a free and democratic society depends on the right and ability of its citizens and residents to communicate with their government and the right and ability of the government to communicate with them."

Potential Problems with Interpreter Use

The goal of language interpretation, strictly speaking, is to bridge a gap in verbal communication, the so-called "language gap." An interpreter is expected to translate verbal communication in one language into another using as close to the same words as possible, but without compromising the meaning of the utterance (Acevedo et al., 2003). It is often assumed at least tacitly that an interpreter will be able to do this effectively and accurately in all contexts, regardless of the subject matter. However, practically speaking, this is an almost impossible task. Although an astute interpreter is continuously working to enhance his or her vocabulary, ultimately, his/her vocabulary is limited by his/her various educational and life experiences, and even by underlying neurobiology. The brain, for example, is constantly being bombarded with stimuli and making decisions about which information to attend to. Information that is left unattended, including new words, does not have the opportunity to be stored into memory or recalled for later use. Furthermore, it is virtually impossible for a single individual to be exposed to and adept in every subject area or field. An interpreter cannot accurately bridge a communication gap when he or she does not understand the context or substance of the information to be translated, let alone the specialized jargon (Hwa-Froelich et al., 2003). Interpreting psychological terms and interpreting financial terms, for example, involve largely non-overlapping lexicons.

It has also been argued that some aspects of communication are lost in translation while using an interpreter. Sabin (1975) has illustrated in striking terms the difficulty in determining a patient's affective status when an interpreter is used to facilitate a psychiatric interview. He reported two cases in which interpreters were used during the diagnostic interviews of two psychiatric patients. Both patients committed suicide after the interview, yet the two psychiatrists who had worked with these patients had failed to detect their suicidal ideation. In another study, Baker et al (1998) reported that patients who used an interpreter rated their providers as both less friendly and less respectful, less concerned for them as a people, and less likely to make them feel comfortable compared to patients who did not use an interpreter. Interestingly, patients who did not use an interpreter but felt that one should have been used gave their providers even lower satisfaction ratings for aspects of interpersonal care. On the other hand, there is evidence that at least some important information is conveyed through interpreters. Indeed, Bernstein (2002) found that patients in an emergency room department of a university hospital who received language assistance services through trained hospital interpreters were more likely to attend follow-up appointments than patients who did not use trained hospital interpreters. These patients were also less likely to return to the emergency department.

The medical literature is replete with examples of how non-English speakers receive poor quality care. Bernstein (2002) reported that non-English speaking patients spent less time with physicians in a hospital emergency department than English-speaking patients. Results were significant even after controlling for severity of illness at the time of the visit. Non-English speakers also received fewer tests and procedures while in the emergency department, and were given fewer medications during their stay when compared to English-speaking patients. Use of a trained hospital interpreter appeared to moderate these effects somewhat, as patients who received language assistance services from trained interpreters received more services during their visit than non-English

speaking patients who did not use a trained interpreter. Perhaps one of the reasons why non-English speaking patients receive fewer services is because providers either consciously or unconsciously bias the evaluation (Flores, 2005). For example, they may use more structured questions and not be as thorough during their examinations. They might ask more closed-ended questions, thereby limiting the quality and quantity of information that they can use to formulate diagnostic conclusions and treatment recommendations. The provider could also misunderstand the patient and fail to ask for clarification or to follow up with critical questions that could help inform diagnostic and treatment decisions

The relationship between clinicians and their patients plays an important role in psychological assessment and therapy. In fact, research on the efficacy and effectiveness of psychotherapy indicates that the therapeutic alliance is one of the best predictors of positive treatment outcomes (Horvath & Luborsky, 1993). The clinician's primary mechanism for building this relationship is through direct, verbal communication. Although nonverbal gestures are also important, the dynamic exchange of verbal interaction helps the clinician engage the client in the therapy from the basics such as obtaining a personal history to more complex techniques such as challenging automatic thoughts, which are themselves internal verbalizations. Use of an interpreter is a barrier to direct engagement. Additionally, variations in verbalizations convey emotional information that may not be translated well through the use of an interpreter. Psychiatrists who had experience in working with an interpreter expressed concern that affective symptoms were likely to be distorted while using an interpreter during psychiatric interviews (Marcos, 1979). Other studies have shown that health care providers are also less satisfied with the quality of communication with their patients when an interpreter is used, especially when the interpreter is untrained (Bischoff et al., 2003). Interestingly, physician satisfaction ratings of the quality of patient-provider communication tend to be lower than those of the patients even when a trained interpreter is used.

There are several reasons why using an interpreter might impair the patient-provider relationship. For example, the communication process is delayed because both patients and clinicians have to wait until the interpreter is finished speaking before they can respond, thereby limiting the spontaneity and natural flow of the interaction (Acevedo, 1998). Both patients and providers may also forget what they originally wanted to say or limit the nature and number of their responses because they take twice as long to explain. Providers may also forget to maintain eye contact with their patients or fail to address them directly, choosing instead to direct their attention to the interpreter as if the patient is a third-party observer in the interaction (Wallin, 2003). It may be akin in some ways to a conversation in which two people are having a conversation but the third person is primarily listening rather than participating. The three individuals involved are receiving the same verbal input and content of the conversation, but their level of engagement in the interaction is different, which can affect their emotional reaction to the situation as well. One might imagine that use of an interpreter could have even greater influential effects because the individuals involved in this case would not even share the same language or have equal access to the verbal content of the conversation. However, it is important to note that the opposite effect may also be true in some cases and there may be significant variability in the level of engagement between clinicians and patients when an interpreter is used.

Interpreter Burden

Most of the literature focusing on the use of interpreters in medical and mental health care settings has focused on either patients or providers. However, an increasing number of studies are beginning to highlight the myriad effects that the interpreter may experience while fulfilling their role and responsibilities. First of all, research has shown that most interpreters are not trained or professional interpreters, but rather, they tend to be friends or family members of patients or hospital or clinic staff who happen to also be bilingual. Such individuals are commonly referred to as “ad hoc” interpreters. Ad hoc

interpreters bear a difficult responsibility because they have not been trained to fulfill this role, which is more often than not thrust upon them whether they want it or not. For example, they may feel as though interpreting takes time away from their usual daily activities or employment responsibilities. They may also feel that their language skills are insufficient to serve as an interpreter, but that they cannot refuse to participate without repercussions. With respect to interpreters in mental health settings, interpreters may be ill–prepared to deal with the emotional and sometimes traumatic content of the dialogue that emerges during the assessment or therapy (LaCalle, 1987). For example, individuals who need language assistance services in mental health settings may also be refugees from foreign countries seeking political asylum, they may be torture survivors, or they may have faced horrific conditions during their immigration to the United States (Miller et al., 2003). Interpreters in such settings bear the brunt of the emotional burden because they experience it first hand, directly from the patient. Also, interpreters may find it distressing to ask difficult questions of clients, especially when the content of the questions is considered to be a cultural taboo.

Professional interpreters may also experience dilemmas with dual relationships that are similar to psychologists'. In fact, the issue of dual relationships may be an even more important issue for the professional interpreter because they are often part of the patient's cultural community and these communities can be quite small. Issues of confidentiality can therefore become difficult to manage. Also, if the patient and interpreter share similar backgrounds and experiences that are painful to talk about, the interpreter may experience difficulty remaining objective or emotionally uninvolved. Sometimes an interpreter may face a backlash in his or her community for working with certain providers or in certain settings that are negatively viewed within the community. Community members may also expect the interpreter to serve as their advocate or to work on their behalf to help them obtain certain services, a role that is not traditionally

considered within the scope of a professional interpreter's duties from the perspective of a health care provider.

Neuropsychology and Third Party Observers

Issues surrounding the use of interpreters in neuropsychological assessment are inherently complex. Further complicating the issue is the literature regarding third party observer effects on neuropsychological test scores. The majority of literature in this area has emerged from forensic neuropsychology. Neuropsychologists are sometimes asked to assist in various legal proceedings, such as personal injury claims, homicide cases, and criminal investigations. It is not uncommon for attorneys to request to accompany their clients into the evaluation during such situations in order to gain insight into the nature and types of testing, which may in turn influence a number of important decisions affecting their clients. Research by Bush and colleagues (2004) suggests that the presence of such an observer could jeopardize the validity of the neuropsychological evaluation. Several studies have reported that the presence of a third party observer lowers scores on tests of memory, complex attention, processing speed, and verbal fluency, but not tests of motor functioning (Gavett, 2005). These findings have been robust even when the "third party observer" is an audio or video recording of the evaluation (Constantinou, 2005).

The Standards for Educational and Psychological Testing (AERA, APA, NCME, 2002) identify the guidelines for test use that psychologists are expected to follow. Standard 15.1 states that tests should be administered according to the instructions and procedures set forth by the publisher, which typically do not include (and sometimes specifically prohibit) the presence of a third party observer. Most of the major neuropsychological societies have also taken a policy stance that strongly discourages psychologists from allowing observers into the testing session. Observers increase the likelihood of visual or auditory distractions during the examination due to the mere presence of an additional person in the room. They may also bias the internal cognitive and affective processes of the patient, especially when the observer is an "involved"

party, or someone who has a vested interest in the outcome of the evaluation. (Bush et al., 2005).

Provider and Interpreter Alliance

The professional alliance between a provider and an interpreter may be just as important as the alliance between the provider and the patient. Like all relationships, this takes time and experience to build. Providers and interpreters who have established a professional alliance share a common knowledge base, have confidence in each other's abilities, and understand their unique contributions to the assessment process. They are also more comfortable working with each other and will likely have developed a structure and format for the evaluation that has become routine. Their familiarity with the process will likely help the patient feel more confident and at ease as well. Additionally, the development of a professional alliance assumes that the provider and the interpreter have had an opportunity to become comfortable with each other's natural working style. Every individual brings their unique personal dynamic into their interactions with others. Thus, some interpreters may be naturally more nurturing, while others may be more direct and advice-giving. The same is true for providers. Sometimes the natural working style of the provider and the interpreter do not automatically mesh, and it takes time to learn to accommodate to each other's approach.

The Current Study

Given the paucity of empirical research regarding the use of an interpreter in mental health settings, and for neuropsychological evaluations in particular, the primary objective of this study was to determine whether using a bilingual interpreter to conduct neuropsychological testing of monolingual Spanish speakers affects the neuropsychological test scores. The population of LEP speakers who speak Spanish as their native language is growing rapidly in the United States, which provided a natural rationale for choosing monolingual Spanish-speakers as participants for the study.

The emphasis was on neuropsychological testing, rather than neuropsychological assessment, the latter being a more dynamic and complex process of which testing is merely one component. Neuropsychological assessment involves integrating information from interviews with the patient, personal history, medical records, neuroimaging data, observations of the patient, and test results to develop diagnoses and treatment decisions. Additionally, whereas neuropsychological testing per se can be conducted by a trained technician, neuropsychological assessment should be completed by a Ph.D-level psychologist with extensive training in neuropsychology (the so-called "Houston model" training, with board certification being the clearest evidence of such competence). Because the study of interpreter use in neuropsychology is in its infancy, it was considered a logical and appropriate "first step" to begin this area of inquiry from one of the most essential tools neuropsychologists have available in their assessment toolbox, namely, neuropsychological tests.

This study was conducted using a neurologically healthy sample of participants. The decision to use neurologically intact participants was made based on the rationale that a normal sample would provide the most stringent test of the hypotheses. Specifically, if the effect of interpreter use was significant using a neurologically healthy sample, the belief was that the effect would be even stronger in a neurologically impaired sample, where factors such as inattention, impaired comprehension, impaired expression, and impaired memory are frequently part of the equation. Future studies are needed to address the question of how an interpreter affects test scores in a neurological population, but regardless, the results of the current study provide a necessary foundation for this future work.

Specific Aims:

1.) To contribute to the neuropsychological literature regarding the effects of using an interpreter to facilitate neuropsychological testing of LEP patients.

2.) To contribute to the neuropsychological literature regarding best practices and care for Latino patients.

Hypotheses:

Although there have not been any studies to date regarding the issue of interpreter use and neuropsychological testing, several hypotheses were generated regarding this issue based on work conducted in related fields (e.g., psychiatry and medicine). A number of neuropsychologists have also expressed strong opinions regarding the use of interpreters in neuropsychological journals based upon their personal experiences (Artiola y Fortuny et al., 1998; Bush & Lees-Haley, 2004). The following hypotheses were tested in this study:

1.) Use of an interpreter would lower neuropsychological test scores. Specifically, compared to the no-interpreter group, the following variables were predicted to be significantly lower in the interpreter group: memory, attention, language skills, visuo-spatial abilities, visuo-constructional abilities, and executive functioning skills.

The expectation that neuropsychological test scores would lower scores in the interpreter ("I") versus the no-interpreter ("No-I") condition of the study was based upon a number of factors that were previously discussed. For example, it was assumed that errors in interpretation and factors related to information being "lost in translation" would adversely affect the process of test administration and result in lower test scores. Interpersonal differences were also thought to affect tester-interpreter alliance in a way that would adversely affect test performance and result in lower test scores. In addition, research on the effects of third party observers supported the idea that the presence of an additional person during neuropsychological test administration would result in poorer test performance.

It was expected that data supporting this hypothesis would come from a finding that neuropsychological test scores in the "I" condition of the study were significantly

lower than neuropsychological test scores in the "No-I" condition. If neuropsychological test scores in the "I" condition were the same or higher than the neuropsychological test scores in the "No-I" condition, then the hypothesis would have been falsified. A MANOVA was planned to test this hypothesis, which would minimize the possibility of producing a Type 1 error due to multiple comparisons of the dependent variables. Because this test is fairly stringent, the assumption was that even if only a single neuropsychological test produced significantly lower scores in the "I" versus the "No-I" condition, it should be considered at least partially supportive of the hypothesis, and potentially clinically meaningful. However, it was predicted that all neuropsychological subtests (variables) in the battery – including tests of memory (CVLT-II Long Delay, CFT 30-minute Delay), attention (Digit Span, Trails A), language (Vocabulary, BNT), visuo-spatial abilities, visuo-constructional abilities (Block Design, CFT) and executive functioning (Matrix Reasoning, COWAT, Similarities, Trails B)- would be significantly lower in the "I" condition compared to the "No-I" condition.

2.) Use of an interpreter would lower scores on verbal tests to a greater extent than nonverbal tests.

In the "I" condition, a pattern of lower verbal scores compared to non-verbal scores was predicted based on previous findings. For example, interpreters can produce several types of errors such as omissions, elaborations, or distortions of utterances. Interpreter errors have contributed to medical misdiagnosis and miscommunication (Sabin, 1975, Marcos, 1979). Because verbal tests required more dialogue and more translations, it seemed possible that more interpretation errors could be made that would have adversely affected test scores. In addition, nonverbal tests allow testers and participants to communicate more directly, which was expected to temper (but not eliminate) the effect of interpreter presence.

Follow up analyses were planned to compare group means and effect sizes for the verbal and nonverbal subtest scores of the "I" condition. It was predicted that effect sizes

for verbal tests would be moderate (Cohen's d=.5 to .7), whereas effect sizes for nonverbal tests would be small (Cohen's d= .2 to .5) (Cohen, 1998). This pattern of effect size differentiation was expected to support hypothesis 2.

3.) There would be a positive association between amount of interpreter experience and neuropsychological test scores.

Interpreters were not given any explicit or a priori training on the purposes of the neuropsychological tests that were used in this study or how they should have been administered. However, it was expected that interpreters would acquire at least some vicarious knowledge about the purposes of each test over repeated administrations. In addition, as interpreters became familiar with test instructions, content, and response choices over time, they were expected to develop well-rehearsed, memorized language to communicate information more accurately and efficiently. This was predicted to result in a reduction in interpreter errors, which would have minimized adverse effects on test scores (although it was not expected to eliminate it).It was also assumed that interpreters would become more confident in their abilities through experience and increased comfortability in working with the tester, which was expected to temper the effect of interpreter presence on test scores.

Bivariate correlation analyses (for each subtest) were planned to test the hypothesis that interpreter experience and neuropsychological test scores would be positively associated. A significant, positive correlation between interpreter experience and neuropsychological test scores was expected to support this third hypothesis.

METHODS

Participants

A total of 40 participants completed all test conditions and requirements for this study. Participants had a mean age of 39.7 years (*SD* =13.91) and half were female (*N*=20). Most were natives of Puerto Rico (*N*=37) but three participants were born in the Dominican Republic. The mean level of education among the sample was 14.00 years (*SD*=1.99). All but one of the participants completed their education in Puerto Rico.

Participation in this study was restricted to individuals with limited English Proficiency (LEP) whose native and dominant language was Spanish. Language proficiency was operationalized based on a modified definition of an LEP speaker according to the United States Census Bureau. By their definition, an LEP speaker refers to anyone who identifies themselves as only able to speak English "less than very well." However, to further homogenize the sample with respect to level of English language proficiency only individuals who identified themselves as being able to speak English "not well" or "not at all" were eligible to participate in this study. This was operationalized based upon the individual's answer to the question listed in Appendix A, which was asked as a screening question over the telephone when individuals called to inquire about participation in the study. In addition, participants who met initial eligibility criteria also completed the Picture Vocabulary and Verbal Analogies subtests of the English version of the Woodcock-Munoz Language Survey - Revised (known simply as the "WMLS-R") during their first visit to the laboratory in order to document their level of English language proficiency objectively (Woodcock & Munoz-Sandoval, 1995). The WMLS-R was designed to measure cognitive academic language proficiency (CALP) (arguably the type of language used for formal neuropsychological testing). Additionally, the WMLS-R manual indicates that one of its intended uses is to describe the language characteristics of participants in research studies. It also has excellent psychometric properties including a large, nationally representative normative sample, high reliability

coefficients, and strong validity data. The Picture Vocabulary and Verbal Analogies subtests form an "Oral Language" cluster which can be used to assess an individual's level of expressed oral language proficiency. The WMLS-R yields five language proficiency classifications: Level 5: Advanced English CALP; Level 4: Fluent English CALP; Level 3: Limited English CALP; Level 2: Very Limited English CALP; Level 1: Negligible English CALP. Participation in this study was limited to individuals who were classified at Level 3: Limited English CALP, or lower on the Oral Language cluster.

Participants in this study were within the age range of 18-65 and neurologically healthy. Individuals who were over the age of 65 were excluded due to their increased risk for development of neurological disorders and/or cognitive dysfunction associated with advanced age (i.e. mild cognitive impairment) which would have introduced a potential confound into the study. Individuals who self-reported a history of head trauma, stroke, seizures, dementia, or other neurological condition, or psychiatric history were excluded for similar reasons (Appendix B).

A broad-based sampling strategy was used to recruit participants for this study. Participants were recruited through newspaper advertisements both on the University of Puerto Rico campus and in the community of Rio Piedras, which is located within the municipality of San Juan, Puerto Rico. Additionally, participants were recruited through flyers posted on the university campus and in community spaces, such as local churches, restaurants, libraries, and community centers. Participants responded to such advertisements by contacting a research assistant of the study via telephone using a local Puerto Rico phone number. At that time, the research assistant described the research as well as the expectations for participation, which included two visits to the laboratory, which was located in the Department of Psychology of the University of Puerto Rico at Rio Piedras. If callers were still interested in participating, the research assistant asked a series of questions to determine whether or not the individual met the eligibility requirements for participation in the study. All recruitment materials and screening

criteria were translated from English into Spanish by at least three bilingual research assistants using a "consensus translation" strategy.

A number of strategies were undertaken to assist with recruitment and retention efforts in this study. Participants were each provided with $5.00 in cash on each day of testing to assist with transportation costs. Checks in the amount of $20.00 were also mailed to each participant after completion of each session. Snacks (cookies and water) were also provided during each testing session. Finally, reminder phone calls and thank you cards were mailed to participants. Despite these efforts, there were several challenges to recruitment and retention in this study. A total of 193 individuals inquired about participation and completed telephone screening. Only 102 met all of the eligibility criteria based on the telephone screening and were enrolled in the study. The majority of individuals were excluded because their English language proficiency was judged to be too advanced, they reported that they were taking medication for a psychiatric condition, or they reported a history of head injury. Of the 102 individuals who met the screening criteria and were enrolled in the study only 52 completed at least one condition. A total of 40 participants completed both conditions of the study.

It is possible that individuals who decided not to participate in the study or dropped out before its completion differed in some way from participants who did complete the study. For example, when research assistants called participants who skipped their scheduled appointments to try to reschedule for a later date, many of them indicated that they had recently become employed and no longer had daytime availability to participate. Some also expressed difficulty with obtaining transportation. However, college students, ethnic minorities, and individuals with limited economic resources have also been shown to have lower participation rates in scientific research.

Procedure

A within-subjects, repeated measures design was employed to examine the study hypotheses. Each participant completed a two-hour battery of neuropsychological tests, including both verbal and nonverbal tests, twice. In one condition, a monolingual English-speaking tester administered the neuropsychological tests via a bilingual Spanish/English speaking interpreter (i.e. Interpreter-Mediated condition, abbreviated "I"). In the other condition, a bilingual Spanish/English speaking tester administered the neuropsychological tests directly in Spanish (No-Interpreter condition, abbreviated "No-I"). The condition of test administration was counterbalanced across participants such that 24 participants received the No-I condition first and 16 received the "I" condition first. Testing sessions were video and audio recorded in both study conditions. The length of time between the first and second testing sessions was between 3 weeks and 5 months.

This study was conducted through collaboration with the Department of Psychology at the University of Puerto Rico at Rio Piedras. All subject recruitment and data collection took place in Rio Piedras, which has a population of about 515,600, and is the largest district of the city of San Juan. The rationale for conducting the study in Puerto Rico was multifaceted. First, the heterogeneity of the Latino population was discussed previously. Conducting the study in Puerto Rico provided the opportunity to recruit a more culturally homogenous sample of participants to help isolate the most important variable of interest, interpreter use. Second, in the same way that Latino culture is diverse, variations in dialects make the Spanish language somewhat diverse as well. Using a Puerto-Rican-only sample helped to minimize the heterogeneity of dialectical variations in the Spanish language among participants. Third, Spanish is the primary language of most Puerto Ricans, and although English is taught in the country's schools, 71.9% of Puerto Ricans indicate that they speak English "less than very well" (United States Census Brief, 2000).

Upon arrival in the laboratory for the first testing session, participants were provided with a copy of a Spanish version of the *Informed Consent* document and a Spanish-speaking research assistant reviewed the document with the participant. Participants had the opportunity to ask any questions they might have had prior to beginning testing, and if they agreed to participate by signing the *Informed Consent* document, they completed testing in either the "I" or "No-I" condition the same day depending on which initial condition they were assigned to based upon the counterbalancing order.

Research Assistants

There were three different and mutually exclusive roles for research assistants (RAs) in this study.

1.) Monolingual English-speaking tester – administered the neuropsychological tests to participants in English via an interpreter.

2.) Bilingual Spanish/English-speaking testers – directly administered the neuropsychological tests in Spanish without an interpreter.

3.) Bilingual Spanish/English speaking interpreters – facilitated communication between the monolingual English-speaking tester and participants.

Training for Testers

Both the monolingual and bilingual testers were trained in accordance with training guidelines of the Benton Neuropsychology Laboratory, Department of Neurology, University of Iowa Hospitals and Clinics. These RAs were taught how to administer all standardized measures during laboratory meetings devoted to training of their administration. Training was conducted in English, and testers were trained on each individual test included in the study. First, the purpose of each test was explained, and testers were shown how to administer each test with specific and detailed instructions in a step-wise fashion. Testers were then given the opportunity to practice each step involved in the administration of the tests. Testers observed two full assessments performed on

volunteers and practiced on their own until they felt they were ready to be observed administering the tests in English. They then administered the tests to at least two volunteers under the supervision of the principal investigator. Feedback was provided and discussed during this administration and subsequent administrations until the amount of feedback needed was minimal to none. Testers were then required to administer the tests error-free twice in succession on volunteers prior to confirmation of competence at this stage. Volunteers were required to be healthy, normal adults between the ages of 18-65.

Specific Training for Bilingual Testers

Instructions for each neuropsychological test used in this study were translated from English into Spanish using a consensus translation strategy. Bilingual testers had the opportunity to familiarize themselves with the standardized Spanish instructions for each test after first learning how to administer the tests in English. When they felt they were ready, bilingual testers were then required to administer the tests in Spanish on two volunteers and without errors before being allowed to administer the tests to study participants.

Training for Bilingual Interpreters

To enhance ecological validity and model the reality of typical interpreter experience, research assistants who were selected to serve as bilingual interpreters for this project were not given explicit a priori training on how to administer the neuropsychological tests used in this study. In real life, many interpreters do not receive any training for how to conduct their work (i.e. ad hoc interpreters), let alone training for administering neuropsychological tests. Neuropsychological testing is a highly specialized service, and it is unlikely and currently unrealistic to expect that employers will pay for interpreters to receive specific training in neuropsychological test use and administration. However, there has been a recent push toward providing interpreters, at least those who work in health care contexts, with more general training for how to

interpret in mental health settings. Thus, although interpreters in this study were not given detailed information regarding how to administer neuropsychological tests, they underwent intensive and systematic training on effective strategies for mental health interpreting.

The Deaf Wellness Center at the University of Rochester has developed a number of resources (in addition to producing a plethora of research) on training mental health interpreters to work with the deaf. Much of their work can be extended and/or adapted for foreign language interpreters. For example, their *Mental Health Interpreting: A Mentored Curriculum* (Pollard, 2004) was designed for use with both foreign language and American Sign Language interpreters. The curriculum includes a textbook with nine chapters as well as an accompanying DVD. Each chapter outlines a number of learning objectives and also includes a brief examination at the end for interpreters-in-training to demonstrate mastery of each objective. The DVD includes a series of eleven vignettes that are typically related to a learning objective and can be used to highlight the dominant themes of each chapter. A copy of the table of contents for the text as well as summaries of each of the videotaped vignettes can be found in Appendix C.

Research assistants who served as bilingual interpreters for this study were provided with a copy of the text and worked individually with the principal investigator to review and discuss each chapter and its accompanying video vignette in detail. Interpreters were required to pass the examination at the end of each chapter before being able to participate in the study. Training for interpreters in this study was intentionally made more intensive and systematic for several reasons. First, it is has been recommended in the literature that interpreters who work in mental health settings receive specialized training for working with psychologists (Marcos, 1979). Second, although in real life many interpreters are “ad hoc” interpreters with little or no training, the use of ad hoc interpreters in this study would have jeopardize the validity of the results. Using trained interpreters provided for a more stringent test of the hypotheses and limited the

possibility of interpreter training being used as a methodological critique. Finally, the interpreter training procedures needed to be stringent and clearly delineated so that future researchers could replicate the study.

Language Proficiency Issue for Interpreters

An important consideration in the design of this study was the way in which Spanish and English language proficiency was operationalized for the bilingual interpreters. Level of bilingualism is on a continuum, and not every bilingual is equally proficient in each of their languages. However, by virtue of their job responsibilities, language interpreters should be able to communicate very well in both of their languages. For this study, it was important that the tester in the tester-only condition also be able to speak Spanish and English very well. Ideally, their skills should be as closely matched as possible to the bilingual interpreters to eliminate the possibility that proficiency-related differences in Spanish or English could account for test score differences between the interpreter-mediated and tester-only conditions of the experiment.

The American Council on the Teaching of Foreign Languages (ACTFL) is one of the foremost organizations that is dedicated to the education and training of foreign language instructors and students at various levels of education around the world. Founded in 1967, the organization now has more than 9,000 members and has been a pioneer and prominent leader in working toward the creation of Proficiency Guidelines to identify levels of competency in speaking, listening, reading, and writing a foreign language. The guidelines articulate five levels of increasing proficiency and were initially utilized by the United States Foreign Service Institute. They are intended for "global assessment" of language function and proficiency, and are not based upon any particular linguistic theory. In addition, the guidelines are intended to identify an individual's ability to use the language in various capacities independent of how the language was acquired (e.g. classroom instruction, language immersion program, etc.).

In 1992 the ACTFL began using Language Testing International (LTI) as its official licensee for conducting language proficiency assessments based upon their Proficiency Guidelines. They now offer an Oral Proficiency Interview (OPI), which is a test of oral language proficiency, in more than 65 languages (including Spanish and English). The OPI consists of either a face to face interview or a telephone interview with an ACTFL certified tester, and includes a series of questions, probes, and level checks that are designed to produce a ratable speech sample from the interviewee that is audio recorded. A computerized version of the exam is also available. The speech sample is assigned a rating of "Novice, Intermediate, Advanced, or Superior" with three possible sub-levels within each ranking (except Superior) – "Low, Mid, or High" – by the interviewer and an additional independent rater. If the two raters do not agree on the same proficiency ranking for the interviewee, then a third rater is assigned to review the speech sample and make an independent rating.

Because this study was conducted in Puerto Rico, and the roles of both the bilingual interpreters and the bilingual testers were filled by Puerto Ricans who were native Spanish speakers, it was considered reasonable to assume that their Spanish language abilities were entirely proficient. However, because these students acquired English as a second language, we utilized the ACTFL OPI as an objective measure of their English speaking proficiency. Bilingual interpreters and bilingual testers in this study were required to pass the OPI at the Advanced-Low level of proficiency at a minimum. The Advanced-Low ranking has been used as the yardstick for licensure for some state foreign language instructors at the PK-12 level, such as the state of Virginia. According to the ACTFL Proficiency Guidelines for the Advanced-Low level, speakers at this level can "…contribute to the conversation with sufficient accuracy, clarity, and precision to convey their intended message without misrepresentation or confusion, and it can be understood by native speakers unaccustomed to dealing with non-natives, even though this may be achieved through repetition and restatement."

Materials

Test Selection Considerations

Determining which tests to include in this study involved a number of important considerations. Among the factors considered were the reliability and validity of the instruments, their utility, and their intended purpose. In addition, variables such as administration time and cost also were important. The impact of the study's design was also taken into account, as certain factors may be more or less salient when testing the same group of people twice. Furthermore, both verbal and nonverbal tests were selected in order to test the second hypothesis.

Reliability often is viewed as the extent to which test scores for a particular instrument are consistent over some interval of time. This type of reliability is called test-retest reliability, and in fact, is one of several types of reliability, which more broadly defined refers to the consistency of test scores (rather than a property of the test itself; Urbina, 2004). Test-retest reliability was one of the most important types of reliability to consider when selecting neuropsychological tests for this study, because of its within subjects, repeated measures design, in which the same tests were administered twice to the same individuals. To minimize the possibility that practice effects, memory effects, and other sources of measurement error could account for potential differences between the test scores in the two conditions of this study, it was essential to select tests with high test-retest reliability coefficients. Otherwise, fluctuations in test scores between the two groups could be completely unrelated to the presence of an interpreter in the test administration process, which would introduce a confounding variable into the study.

Another particularly important type of reliability considered in this study was inter-rater reliability. Inter-rater reliability refers to the consistency of test scores across different raters or scorers of the test. There were different raters in this study - because there were two different conditions. Thus, it was important to select tests with high inter-

rater reliability coefficients so as to minimize the likelihood that test score differences between the two conditions were the result of errors in the scoring process.

Internal consistency reliability is another type of reliability that was important to consider as it is critical for any standardized test. It provides an estimate for how well the items of a test correlate with each other and helps to contribute to construct validity by ensuring that the test measures what it purports to measure. However, construct validity is much more than internal consistency reliability alone; it also includes content validity, concurrent and discriminant validities, criterion validity, and external validity. Tests that were included in this study needed to demonstrate excellent construct validity in order to help ensure that they were a legitimate measure of the target construct.

In addition to selecting reliable and valid tests, it was also important to select instruments based upon their clinical utility. Neuropsychologists tend to favor the use of certain instruments during their assessments (Rabin et al., 2005), and to maximize the study's ecological validity, tests that reflected current neuropsychological usage trends were included.

Neuropsychologists also select specific tests based upon their intended use or purpose, in order to probe the major domains of cognitive functioning. For example, they may include an assortment of tests designed to assess different cognitive domains such as verbal and visual memory, language, naming ability, visuospatial skills, executive functioning, or information processing speed to name but a few. It was important, therefore, not only to include widely used neuropsychological tests in this study, but tests that were representative of different cognitive abilities as well. It is possible, for example, that some tests may be more sensitive to the use of an interpreter during the test administration process than others. Thus, this study explicitly incorporated a verbal versus nonverbal distinction as one hypothesis.

Time constraints and cost were two additional variables that needed to be considered during the test selection process for this study. Although the emphasis was on

designing an empirically sound experiment, it also was important to evaluate the cost-effectiveness ratio of implementing the project, as well as to work within a reasonable budget. Thus, a battery of tests that could be administered within approximately 2 hours, but that also retained the above-mentioned features, was desirable. Limiting the test battery to 2 hours helped the economy of the study by reducing the amount of participant –reimbursement and research-assistant costs. In addition, it helped to minimize research-assistant fatigue and facilitated participant recruitment.

Bearing the aforementioned considerations in mind, the following tests were selected to be included in this study:

1. Weschler Adult Intelligence Scale – 3rd Edition (WAIS-III) – Vocabulary, Similarities, Block

Design, Matrix Reasoning

The Vocabulary, Similarities, Block Design, and Matrix Reasoning subtests of the WAIS-III provide parallel items to the Weschler Abbreviated Scale of Intelligence (WASI) and provide an estimate of intellectual ability in a brief period of time. The subtests demonstrate excellent test-retest reliability as well as inter-rater reliability. Administration of these four subtests typically takes no longer than 30 minutes, and in fact, an estimate of FSIQ can be obtained in as little as 15 minutes using only the Vocabulary and Matrix Reasoning subtests. For an additional 15 minutes, an estimate of Verbal IQ (VIQ) can be obtained with the Vocabulary and Similarities subtests, and an estimate of the Performance IQ (PIQ) can be obtained with the Block Design and Matrix Reasoning subtest. Together, these four subtests include two verbal and two nonverbal tasks, which provided one means of examining the hypothesis that verbal subtests would be more sensitive to interpreter-related effects when compared to nonverbal tests. Third, these verbal and nonverbal subtests are actually used to examine specific types of abilities in brain damaged patients, such as acquired verbal knowledge (Vocabulary), abstract verbal reasoning (Similarities), visuo-spatial processing (Block Design) and

abstract nonverbal reasoning (Matrix Reasoning). The Vocabulary and Similarities subtests were administered without regard to discontinuation rules in this study.

2. *Digit Span*

Digit Span is a subtest of the Weschler Adult Intelligence Scale. It is widely used by neuropsychologists and has excellent test-retest and inter-rater reliabilities. With respect to validity, however, it has some limitations. Digit Span actually consists of two separate tests – Digits Forward – which is a test of simple attention, and Digits Backward, which measures verbal working memory, a more complex skill. The WAIS-III combines the scores on Digits Forward and Digits Backward into a single total score, which can be misleading because the scores are indicative of different types of cognitive abilities. Nevertheless, Digit Span is extremely useful as a clinical instrument when the scores on each subtest are reviewed independently by neuropsychologists. For this study, we compared the *Digits Forward* and *Digits Backward* scores across the two study conditions to determine whether or not interpreter-related effects differentially affected either of those scores.

3. Controlled Oral Word Association Test (COWAT)

A commonly used test of executive functions, COWAT requires patients to produce as many words as possible within 1 minute after being given a particular letter of the alphabet. COWAT is highly correlated with verbal IQ and is useful in helping to detect the presence of dementia, aphasia, and head injury, to name a few conditions. Test-retest reliability coefficients range tend to be high (above .70) at 1-week intervals as well as intervals as long as 5 years in healthy individuals. Inter-rater reliability is very high (.99) for healthy participants under normal testing conditions. In order to administer this test in the "I" condition, the monolingual tester read the instructions verbatim in English. The interpreter then translated the instructions into Spanish for the participant. Then, the interpreter recorded the responses of the participant in writing and later translated them back into English for the examiner.

4. *California Verbal Learning Test – Second Edition*

The CVLT-II is a test of verbal learning and memory ability that often is administered in neuropsychological assessments. It was standardized on a large sample of 1087 individuals, ages 16–89, from 47 states, and includes stratification data on a number of important variables such as level of education, gender, and ethnicity, designed to mirror the US Census data. The test also includes both a standard and an alternate form, with alternate form reliability coefficients ranging from 72-.79 for all score indices. With respect to validity, the instrument is highly correlated with the first edition of the CVLT (coefficients ranged from .72-.80), which has accumulated validity data from more than 200 independent studies. In addition, it demonstrated the predicted pattern of association with a number of demographic variables, including age, gender, and level of education. The predicted association between the CVLT-II and IQ was also supported. Like COWAT, the CVLT-II required the interpreter to record participant responses in writing and later translate them back into English for the examiner to score.

5. Complex Figure Test (CFT)

The Complex Figure Test is commonly used as a test of visuo-constructional ability as well as nonverbal memory. It is one of the top 10 tests used by neuropsychologists (Rabin et al., 2005) and, like *Digit Span*, the CFT actually consists of separate components that can be analyzed both quantitatively and qualitatively. For the purposes of this study, the copy score was used as a measure of visuo-constructional ability, as it requires individuals to look at a copy of a figure and make as accurate a copy as he/she can (a 5-minute time limit was used in this study). Thirty minutes later the participant is asked to reproduce the figure as accurately as possible from memory. This 30-minute delayed recall score was used as an index of long term visual memory. The CFT has excellent inter-rater reliability with estimates ranging from .88 to .96 for each of the total score indices (Copy and Delayed Recall).

6. Boston Naming Test

Naming defects are a hallmark of many types of brain impairments. It makes sense, then, that naming tests are among the most frequently administered neuropsychological tests. The Boston Naming Test (BNT), in particular, was reported as one of the top ten tests used by neuropsychologists (Rabin, 2005). Test-retest reliability is high over short intervals of a few weeks in healthy individuals. In this study all 60 BNT items were administered beginning with the first item and discontinuation rules were not applied.

7. Trailmaking Test

The TMT is another measure of executive functioning, and more specifically, a test of attention, speed, and mental flexibility. It is a widely used instrument by neuropsychologists, and is arguably one of the most sensitive tests of cognitive defects (although not the most specific). The TMT can be administered in less than 10 minutes.

RESULTS

Data Screening and Preliminary Analysis

Prior to conducting formal statistical analyses the data were screened for errors, outliers, and violations of assumptions. Examination of frequency distributions, descriptive statistics, and histograms for each dependent variable revealed that there were several outliers for the Trail Making Test, Parts A and B, primarily in the No-Interpreter condition of the study. Review of the test protocols revealed that 22 of the 40 participants received poor quality photocopies of the Trail Making Test, Parts A and B in the No-Interpreter Condition of the study. Consequently, the Trail Making Test, Parts A and B were removed from subsequent statistical analyses.

Test of Hypothesis One

The first hypothesis stated that neuropsychological test scores would be significantly lower in the "I" condition of the study compared to the "No-I" condition. To test this hypothesis a repeated measures multivariate analysis of variance (MANOVA) was performed with one independent variable (test condition) and 10 dependent variables (neuropsychological test scores), which are identified in Table 1. Descriptive statistics for all dependent variables are included in Table 2. Order of test condition administration was included as a between-subjects factor to determine the effects of counterbalancing the study conditions. Using Pillai's trace, results indicated that there was a significant main effect of test condition on neuropsychological test scores, F (10, 29) =8.39, p<.000. This finding indicates that there were differences in neuropsychological test scores depending upon whether or not an interpreter was utilized. The main effect of order was not significant, F(10, 29) = 1.35, p>.05, indicating that neuropsychological test scores did not differ significantly as a function of whether participants received the "I" or "No-I" conditions of the study first. The interaction effect of test condition by order was not significant, F(10, 29)=.84, p>.05. Table 3 includes results from the MANOVA source table.

Follow Up Analyses

Because the overall MANOVA statistic was significant for a main effect of test condition, univariate ANOVAs were conducted to ascertain the nature of the effects on each of the dependent variables. Table 4 includes the source table with results from the follow-up ANOVAs. Significant differences in mean scores between the "I" and "No-I" conditions of the study were observed on four neuropsychological tests: CVLT-II Long Delay, Digit Span-Forward, Similarities, and Vocabulary. There was also a trend toward significance for Digits Backward.

Consistent with the overall prediction, mean scores were lower for CVLT-Long Delay in the "I" condition (M=9.77, SD=3.21) as compared to the "No-I" condition (M=10.53, SD=2.91) F (1, 38) =5.05, p<.05. Although not significant, there was a trend toward lower scores on Digit Span-Backward when an interpreter was used (M=4.77, SD=2.02) versus when a bilingual examiner administered the tests (M=5.35, SD=1.99), F(1, 38)= 3.09, p<.10.

Several findings, however, went in the direction opposite to prediction. Mean scores were higher for the Vocabulary subtest in the "I" condition (M= 40.65, SD=8.29) compared to the "No-I" condition (M=31.98, SD=6.50), F= (1, 38) = 48.52, p<.01. Scores were also higher on the Similarities subtest when an interpreter was present (M= 20.60, SD= 5.32) relative to the when an interpreter was not present (M=18.53, SD=4.39), F (1, 38) = 9.92, p<.01. Finally, scores on Digits Forward were higher for interpreter-mediated (M=8.63, SD= 2.35) relative to standard (M= 7.83, SD=1.92) test administration, F (1, 38) = 8.48, p<.01.

There were no significant differences between the "I" and "No-I" conditions of the study for Block Design, BNT, CFT Long Delay, COWAT, or Matrix Reasoning. Table 2 includes descriptive data, including mean scores and variability estimates for each of the dependent variables by test condition.

Practice Effects:

Because participants in this study completed the same battery of neuropsychological tests twice, it is important to understand the potential influence of practice effects on the dependent variables. Thus, data were collapsed across the "I" and "No-I" conditions of the study to create a new variable, "testing session," with two levels (i.e. Time 1 vs Time 2). A repeated measures MANOVA with testing session as the independent variable was then conducted on the 10 dependent variables. Results are listed in Table 5. Using Pillai's Trace, the results indicate that the effect of practice on neuropsychological test scores was not significant, $F(10, 29) = 1.054$, $p>.05$.

Test of Hypothesis Two

The second hypothesis stated that use of an interpreter would lower scores on verbal tests to a greater extent than nonverbal tests. Effect size comparisons between verbal and nonverbal neuropsychological tests in the "I" condition of the study were planned to test this hypothesis; however, results from the first hypothesis indicated that there were no significant differences between the "I" and "No-I" conditions of the study on any nonverbal measures. As previously reported, significant or near-significant differences were observed on 5 of the 7 verbal tests, although interpreter use did not consistently lower scores across all 5 measures.

Careful examination of Table 2 revealed an interesting and unexpected finding relevant to hypothesis two. Standard deviations appeared to be higher when an interpreter was used to administer 8 of the 10 neuropsychological tests. Variability differences appeared to be most notable for neuropsychological tests from the verbal domain, particularly Vocabulary and BNT. Thus, F-tests were used to test objectively the equality of variances between the "I" and "No-I" conditions of the study for each dependent measure. To compute the *F* statistic, a ratio was obtained by dividing the variance in the "I" condition by the variance in the "No-I" condition for each neuropsychological test.

This ratio was then compared to a critical value of F (39, 39) = 1.7 with a p-value of less than .05. Results are listed in Table 6. Variability was significantly higher in the "I" condition of the study for the BNT (F=1.84, p=.030) and approached significance for Vocabulary (F = 1.62, p=.068).

Test of Hypothesis 3

The third hypothesis asserted that there would be a positive association between amount of interpreter experience and neuropsychological test scores. Bivariate correlations were conducted to examine the relationship between interpreter experience and each of the dependent variables. Interpreter experience was operationalized based upon the number of test administrations each interpreter facilitated. For example, an interpreter's first administration was coded as "1," their second administration as "2" and so forth. A total of six interpreters were employed as a part of this study. Each of the six interpreters participated in a minimum of three test administrations but the number of test administrations was not equally distributed amongst interpreters. Thus, the total number of test administrations ranged from 3 to 13, depending upon the interpreter (Table 7). Because the data collected in this study were ordinal (vs. interval) in nature, one-tailed Kendall's correlation coefficients were used to test this hypothesis. Kendall's correlation coefficients are also recommended for use for data sets with small sample sizes. Results are included in Table 8. Interpreter experience was not significantly related to any of the 10 dependent variables.

Table 1

Scoring Method for Dependent Variables

Dependent Variable	Scoring Method
Block Design Subtest	Maximum of 68 possible points
Boston Naming Test (BNT)	Total correct out of 60 items
California Verbal Learning Test, 2nd Edition, Long Delay (CVLT-II, Long Delay)	Total correct out of 16 items
Controlled Oral Word Association Test (COWAT)	Total correct responses for letters C, F, and L combined (no maximum)
Digit Span Subtest, Forward condition (Digits Forward)	Maximum of 16 possible points
Digit Span Subtest, Backward condition (Digits Backward)	Maximum of 14 possible points
Matrix Reasoning Subtest	Total correct out of 26 items
Similarities Subtest	Maximum out of 33 possible points
Vocabulary Subtest	Maximum out of 66 possible points
Rey Osterreith Complex Figure Test, 30-minute Delay (CFT, Long Delay)	Maximum out of 36 possible points

Table 2

Descriptive Statistics for Dependent Variables

	Interpreter Condition		No- Interpreter Condition	
Measure	Mean	Standard Deviation	Mean	Standard Deviation
Block Design	31.65	9.90	31.70	10.15
BNT	43.10	7.09	42.08	5.22
CFT Long Delay	14.24	7.50	14.56	6.54
COWAT	35.73	8.03	35.67	8.12
CVLT-II Long Delay	9.77	3.21	10.53*	2.91
Digits Forward	8.63*	2.35	7.83	1.92
Digits Backward	4.77	2.02	5.35†	1.99
Matrix Reasoning	13.65	5.28	14.18	4.83
Similarities	20.60*	5.32	18.53	4.39
Vocabulary	40.65*	8.29	31.98	6.50

Notes. N=40

*higher mean, $p<.05$; † higher mean, $p<.10$

BNT = Boston Naming Test

CFT Long Delay = Complex Figure Test, Long Delay Trial

COWAT = Controlled Oral Word Association Test

Digits Forward = Digit Span, Forward Trial

Digits Backward = Digit Span, Backward Trial

Table 3

MANOVA Results for Effects of Test Condition and Order

Effect		Value	*F*	Hypothesis df	Error df	Sig.
Test Condition	Pillai's Trace	.74	8.39	10	29	.000*
	Wilk's Lambda	.26	8.39	10	29	.000*
	Hotelling's Trace	2.89	8.39	10	29	.000*
	Roy's Largest Root	2.89	8.39	10	29	.000*
Order	Pillai's Trace	.32	1.35	10	29	.549
	Wilk's Lambda	.68	1.35	10	29	.549
	Hotelling's Trace	.47	1.35	10	29	.549
	Roy's Largest Root	.47	1.35	10	29	.549
Test Condition * Order	Pillai's Trace	.23	.840	10	29	.342
	Wilk's Lambda	.76	.840	10	29	.342
	Hotelling's Trace	.29	.840	10	29	.342
	Roy's Largest Root	.29	.840	10	29	.342

Note. N=40

$p<.05$

Table 4

ANOVA Results for Effects of Test Condition

Source	Measure	SS	df	MS	*F*	*p*
Test Condition	Block Design (NV)	1.30	1	1.30	.053	.819
	BNT (V)	26.60	1	26.60	1.31	.200
	CFT Long Delay (NV)	2.85	1	2.85	.143	.707
	COWAT (V)	1.30	1	1.30	.063	.804
	CVLT-II Long Delay (V)	16.88	1	16.88	5.05	.030*
	Digits Forward (V)	14.01	1	14.08	8.48	.006*
	Digits Backward (V)	6.08	1	6.08	3.09	.087
	Matrix Reasoning (NV)	5.21	1	5.21	.880	.354
	Similarities (V)	75.21	1	75.21	9.92	.003*
	Vocabulary (V)	1383.80	1	1383.80	48.52	.000*
Error						
(Test Condition)	Block Design	933.70	38	24.57		
	BNT	771.39	38	20.30		
	CFT Long Delay	757.64	38	19.94		
	COWAT	790.70	38	20.81		
	CVLT-II Long Delay	126.88	38	3.34		
	Digits Forward	62.79	38	1.65		

Table 4 (Continued)

Source	Measure	SS	df	MS	*F*	*p*
	Digits Backward	74.81	38	1.97		
	Matrix Reasoning	224.98	38	5.92		
	Similarities	287.98	38	7.58		
	Vocabulary	1083.89	38	28.52		

Notes. N=40

* $p<.05$

BNT = Boston Naming Test

CFT Long Delay = Complex Figure Test, Long Delay Trial

COWAT = Controlled Oral Word Association Test

Digits Forward = Digit Span, Forward Trial

Digits Backward = Digit Span, Backward Trial

V= Verbal Test

NV = Nonverbal Test

Table 5

MANOVA Results for Effects of Practice

Effect		Value	F	Hypothesis df	Error df	Sig
Practice	Pillai's Trace	.23	1.05	10	29	.42
	Wilk's Lambda	.77	1.05	10	29	.42
	Hotelling's Trace	.31	1.05	10	29	.42
	Roy's Largest Root	.31	1.05	10	29	.42

Note. N=40

* p<.05

BNT = Boston Naming Test

CFT Long Delay = Complex Figure Test, Long Delay Trial

COWAT = Controlled Oral Word Association Test

Digits Forward = Digit Span, Forward Trial

Digits Backward = Digit Span, Backward Trial

Table 6

F-Test Comparisons of Variability Differences

Measure	*F*-statistic	*p*	
Block Design	1.03	.463	
BNT	1.84	.030	*
CFT Long Delay	1.32	.195	
COWAT	1.02	.476	
CVLT-II Long Delay	1.22	.269	
Digits Forward	1.50	.105	
Digits Backward	1.03	.464	
Matrix Reasoning	1.19	.295	
Similarities	1.47	.117	
Vocabulary	1.62	.068	

Note. N=40

* $p<.05$

BNT = Boston Naming Test

CFT Long Delay = Complex Figure Test, Long Delay Trial

COWAT = Controlled Oral Word Association Test

Digits Forward = Digit Span, Forward Trial

Digits Backward = Digit Span, Backward Trial

Table 7

Number of Sessions Administered by Each Interpreter

Interpreter	Number of Sessions
Anabellie	3
Alberto	7
Agustin	6
Gisele	5
Lauraliz	13
Madeleine	6

Table 8

Correlations Between Interpreter Experiences and Dependent Variables

Measure	Interpreter Experience	p
Block Design	-.031	.393
BNT	.014	.453
CFT Long Delay	-.120	.148
COWAT	.012	.458
CVLT-II Long Delay	.121	.152
Digits Forward	-.006	.495
Digits Backward	-.001	.495
Matrix Reasoning	.048	.340
Similarities	.073	.538
Vocabulary	-.037	.375

Note: N=40

BNT = Boston Naming Test

CFT Long Delay = Complex Figure Test, Long Delay Trial

COWAT = Controlled Oral Word Association Test

Digits Forward = Digit Span, Forward Trial

Digits Backward = Digit Span, Backward Trial

DISCUSSION

Of the 45 million Spanish speakers currently living in the United States, slightly fewer than half of them speak English "very well" (United States Census Bureau, 2007). As the Latino community continues to increase over the next several decades, the population of Spanish speakers is predicted to grow as well, including the number of limited-English-proficiency speakers. This shift in United States demographics almost certainly will have important consequences for the mental health profession, particularly for specialized services such as clinical neuropsychology, where there is already a critical shortage in the number of bilingual and bicultural professionals available to meet the demand for services for Spanish speakers with limited English proficiency. Consequently, language interpreters may be used to facilitate neuropsychological testing between English speaking neuropsychologists and Spanish-speaking patients. However, the effects of interpreter use on neuropsychological testing are largely unknown. This study was the first of its kind to investigate empirically the effects of interpreter use on neuropsychological tests in a sample of 40 healthy, monolingual Spanish-speaking adults in Puerto Rico.

The primary objective of this study was to determine whether or not use of an interpreter significantly affected scores on neuropsychological tests. The first hypothesis predicted that scores would be lower across all neuropsychological measures when an interpreter was used to facilitate testing compared to direct administration by a bilingual examiner. Results were not entirely consistent with this hypothesis. Contrary to prediction interpreter use did not significantly affect tests scores for 6 of the 10 measures: Block Design, Matrix Reasoning, COWAT, BNT, Digits Backward, or the CFT Long Delay. Furthermore, test scores were higher (opposite to the predicted direction of lower scores) when an interpreter was used to administer Vocabulary, Similarities, and Digit Span, Forward. Consistent with expectations, scores were lower when an interpreter was used on CVLT, Long Delay and there was a similar but non-significant trend for lower

scores on Digit Span, Backward. To summarize, there was a significant main effect of interpreter use for a minority of dependent variables, with opposite effects from prediction on three of the four measures with significant results.

The second hypothesis predicted that interpreter use would lower scores on verbal measures of neuropsychological functioning to a greater extent than nonverbal measures. Given that interpreter use did not significantly affect scores on any nonverbal test and was associated with higher scores on three versus lower scores on one verbal test, the results clearly do not support this hypothesis. Rather, results indicated that verbal tests are more sensitive to interpreter effects than nonverbal tests, but generally resulted in higher, not lower scores.

Use of the term "verbal" in this case applies more generally to the modality in which test stimuli are presented and responses are gathered. For example, the CVLT is a measure of learning and memory in which a list of 16 words is presented verbally across five trials. Participants verbally recall as many of the 16 words as possible after each trial as well as after short and long delays. Significant effects were observed when an interpreter was used to facilitate tests of language (Vocabulary), executive functioning (Similarities), long term memory (CVLT, Long Delay) and simple attention (Digit Span, Forward) when they were presented in the verbal modality. There were no significant differences in mean scores for nonverbal tests depending upon whether or not an interpreter was used across any cognitive domain.

The pattern of results in this study is interesting. It suggests that there may be more than one factor or process by which interpreter use affects neuropsychological test scores (i.e. by either increasing or decreasing scores on specific types of tests). Although this study was not designed to identify or elucidate those factors, the pattern of observed results helps to provide some potential explanations. Significant results were observed on two types of tests: those that rely heavily on participant verbal expression (i.e.

Vocabulary, Similarities) and those that require auditory attention, working memory, and long term memory (i.e., Digit Span, Forward, CVLT-II, Long Delay).

The finding that interpreter use affected scores on neuropsychological tests that rely heavily on verbal expression is most likely is related to the increased demand on interpreter abilities and skills that are inherent in these types of tasks. Items for the Vocabulary and Similarities subtests, for example, are given a score of 0, 1, or 2 points depending upon the accuracy of verbal responses produced by the participant. Differences between 0, 1, or 2 point responses can sometimes be subtle and even slight changes in word choice can affect item level scores. Participant responses also can be vague and difficult to understand or differentiate without specialized expertise and familiarity with administering the exam. Although examiners are trained explicitly to detect nuanced differences in participant responses, interpreters typically are not even familiar with the purposes of the test. Thus, interpreters may translate participant responses in a way that, unbeknownst to the examiner, enhances their true scores. Although it was predicted that interpreter errors such as omissions, distortions, and elaborations would lower neuropsychological tests scores, it is possible that these types of errors also could result in significantly higher scores as well. As an example, participants may have provided responses that were best categorized as "0" or "1"point answers. However, if these responses were "edited" by interpreters such that information was omitted, distorted, or otherwise elaborated upon in a way that made the responses appear better, then scores would be inflated artificially.

Interpreter effects on neuropsychological tests of attention, working memory, and long term memory also may be related to interpreter errors. One possibility is that interpreters could have distorted the administration of the Digit Span and CVLT-II tests by reading the digit sequences or word lists too quickly, too slowly, or with variable emphasis on specific digits or words. Studies have shown that recall of digits on the Digit Span subtest are affected by both the rate at which the digits are presented and the pitch

of the voice on the last digit of a series (Newton, 1950; Hagen et al. 2006, Taub, 1972). For example, Newton (1950) found that psychiatric patients recalled more digits on Digit Span, Forward (but not Digit Span, Backward) when the test examiner dropped the pitch of their voice on the last digit in the series. Children also have been shown to increase their performance on Digit Span, Forward when the rate of digit presentation is faster (Engle et al., 2004). In addition, the rate in which word lists are presented on list-learning tasks such as the CVLT-II has been shown to affect long term recall of words (Heun et al., 1998; Weible et al., 2002). Weible and colleagues (2002) discovered that older adults recall more words when the presentation rate of the word list is slowed down. In the context of the existing literature, one possible explanation for the observed findings from this study is that interpreters may have read the digit series and word lists too quickly when administering the Digit Span and CVLT-II tests. Consequently, scores were higher on Digit Span, Forward, a test of simple attention, but lower CVLT-II, Long Delay, a more cognitively demanding task of long-term memory.

Another potential explanation for the observed findings for Digit Span and CVLT-II, Long Delay relates to third-party observer effects. It is well documented that individuals tend to perform better on relatively easy or over-learned tasks when they are aware that they are being observed by others, the so-called "social facilitation effect" (Bond et al., 1983). On the other hand, people generally perform more poorly on difficult tasks when they are aware of being watched. Taken together, this information may provide a parsimonious explanation for why interpreter use significantly lowered scores on the CVLT-II, Long Delay, but increased scores on Digit Span, Forward. CVLT-II, Long Delay requires relatively high cognitive demands on long term memory and working memory, respectively, which could be compromised by the presence of a third party, in this case an interpreter. However, Digit Span, Forward is a fairly simple measure of auditory attention that is relatively unchallenging (Green, 2000), so it is the type of test for which social facilitation likely would improve scores. Research on the

effects of third-party observers during neuropsychological testing generally is consistent with these findings. Third-party observers have been shown to affect scores adversely on complex cognitive tests of long term verbal memory and working memory, but to improve scores on less demanding neuropsychological tests (Gavett, 2005).

Although not consistent with prediction, the finding that interpreter use did not uniformly affect neuropsychological test scores in the same direction is nevertheless clinically relevant. For example, these findings suggest that neuropsychologists who administer the Vocabulary subtest via an interpreter could misinterpret the observed scores as "normal" when true impairment may exist. Because interpreter use appears to increase scores on Vocabulary by several points, there is a risk that patients with clinically significant impairments will not be identified due to inflated test scores. The clinical implications of this particular finding are especially important because Vocabulary is used not only as a measure of receptive and expressive vocabulary but as a neuropsychological "hold" test. The so-called "hold" tests consist of a small number of measures that are believed to be robust to a variety of factors, including age-related cognitive decline and various types of brain injury. They are used commonly as estimates of premorbid intelligence (Axelrod et al., 1999) and as a basis for comparing scores from other tests in a neuropsychological test battery. The finding that interpreter use inflates scores on the Vocabulary subtest therefore could have broad and significant effects on the interpretation of scores in the entire test battery.

It is noteworthy that not every verbal measure of neuropsychological functioning was affected by interpreter use in this study. Mean scores for two tests, COWAT and BNT, were unaffected regardless of whether an interpreter or a bilingual examiner administered the measures. The finding that these two tests were unaffected by interpreter use is particularly interesting in light of the fact that cultural factors have been shown to bias performance on both of these instruments. For example, COWAT requires participants to name as many objects as possible within 60 seconds for each of the letters

C, F, and L. However, research has demonstrated that the frequency of those letters is not equivalent for English and Spanish (Roselli & Ardila, 2002). Similar problems have been identified with the BNT in that the rank order of items, which is supposed to become progressively more difficult for native, monolingual English speakers, is not the same for bilingual individuals who speak English as a second language (Kohnert et al., 1998). It is therefore possible that these factors may have obscured any effects of interpreter use on test scores for COWAT and the BNT.

For example, if the more culturally appropriate letters of "P, M, and R" were used on the COWAT test, participants may have been able to produce several more verbal responses, resulting in increased interpreter demands during administration of the test, and a higher probability of errors that could affect test scores. Similarly, participants may have skipped items on the BNT because they did not recognize many of the items, thereby minimizing the amount of translation required by the interpreter.

Although not explicitly predicted at the outset of the study there was also an interesting pattern of increased variability in the interpreter condition that emerged for most neuropsychologist tests. Variability was higher when an interpreter was used to administer 8 of the 10 dependent measures, although only one comparison was statistically significant. This finding is particularly noteworthy in light of the fact that mean scores were unaffected by interpreter use on the BNT. It underscores the previous point that interpreter effects may not necessarily manifest as differences in mean scores but could nevertheless could have important consequences at an individual case level.

Higher variability is associated with reduced precision in the measurement of observed scores. According to classical test theory, an individual's observed score on a given test reflects a combination of their true score plus measurement error. A true score is defined as the score an individual would obtain over an infinite number of administrations of the test, which of course, is not possible. Measurement error includes both random and systematic errors that are introduced during testing. Although it is

impossible to control for random error in observed scores, psychologists try to minimize systematic error so that observed scores are more precise and reliable. However, results from this study suggest that interpreter use may introduce systematic error into the measurement of neuropsychological test performance which manifests as increased variability in scores. In other words, interpreter use increases random error.

The finding that interpreter use increases variability for some neuropsychological tests, specifically BNT and possibly Vocabulary, has important clinical implications. Specifically, it means that these scores are less trustworthy when an interpreter is used. This is a major source of concern given that neuropsychological tests are used routinely used to help make diagnostic and treatment decisions. Neuropsychologists who work with interpreters may be at risk for erroneously interpreting an individual's performance as "normal" (false negative) or impaired (false positive) on a particular test, especially if a narrow confidence interval is used to inform clinical interpretation. On the other hand, larger confidence intervals adversely affect test sensitivity. What is clear from this finding is that the psychometric properties of interpreter-mediated tests remain unknown and should be investigated more rigorously in future studies. If precision and reliability of measurement are compromised, the validity of the test is also jeopardized.

In many ways it is not surprising that interpreter use increased variability in neuropsychological test scores. Neuropsychologists are trained explicitly to administer tests in a standardized manner based upon manualized instructions. The goal of standardization is intentionally to reduce variability and error in measurement so that test scores can be interpreted precisely and reliably and compared across multiple administrations. Interpreter use marks a significant departure from standardized test procedures and therefore increases the likelihood of inconsistency in observed scores. Few, if any, neuropsychologists receive explicit education and training regarding how to work with an interpreter in graduate school. Manuals rarely, if ever, provide precise and replicable instructions for how to administer neuropsychological tests with the assistance

of an interpreter. Patients are generally unaccustomed to communicating with their health service providers via a third party. Consequently, it is relatively easy to understand how interpreter use could increase variability in neuropsychological test scores. There are simply too many unknown and uncontrolled variables that could contribute to errors in measurement.

Considering the many aforementioned factors, perhaps it is surprising is that variability was not even higher for most tests in the interpreter conditions. Interpreters in this study were relatively balanced bilinguals with advanced linguistic proficiency in both English and Spanish. They also completed a program of training regarding general strategies for interpreting in mental health settings. It is possible, perhaps even likely, that use of interpreters with lower levels of linguistic proficiency and/or without any formal training would have lead to even greater variability in test neuropsychological scores. Similarly, patients with brain damage and neurological injury might be even more prone to variable performance when an interpreter is used to facilitate testing. Such patients already show more intra-individual variability in neuropsychological test performance under standardized test conditions (Holtzer et.al, 2008).

A third hypothesis of the study was that interpreter experience would be positively associated with neuropsychological test scores for all dependent measures. The idea was that as interpreters gained more experience and familiarity with the instruments through multiple test administrations they would make fewer errors and become more confident in their abilities. Test scores, therefore, were expected to be higher as interpreter experience increased. However, results did not support this hypothesis. Interpreter experience was not associated with increased test scores for any of the dependent variables even after controlling for the effects of age or education. These results suggest that the effects of interpreter use on neuropsychological test scores are persistent and that familiarity with the instruments and testing process alone does not minimize adverse interpreter effects on test scores. Perhaps if interpreters were provided

with more dynamic feedback about their performance through ongoing supervision, evaluation, and training, significant effects of interpreter "experience" would have emerged. This study was not designed to examine more dynamic effects of interpreter experience. It was also not designed to account for potential effects of tester experience in working with interpreters, which also could have affected the current findings.

Strengths and Weaknesses

A particular strength of this study is the fact that it is the first of its kind to use an empirical design to help ascertain the effects of interpreter use. To date, the question of interpreter use has not been investigated rigorously or objectively within the neuropsychological literature. Existing research within the broader health professions including medicine, counseling, and social work are replete with qualitative examples, case studies, and anecdotal reports regarding interpreter mishaps and their clinical consequences but none of this work has been undertaken using an experimental design. On the other hand, the quantitative emphasis on data analysis in this study arguably could be considered a limitation as well. Some of the findings from this experiment suggest that interpreter effects may be difficult to detect using group-level mean score differences as the primary criterion. The most comprehensive approach to investigating and understanding the effects of interpreter use may therefore be a combination of both empirical and qualitative research techniques.

This study used a within-subjects, repeated-measures design which has a number of strengths. In particular, this type of design has the benefits of increased efficiency, reduced variability, and increased power with fewer participants. Nevertheless the sample size was still relatively small, and consequently, the power to detect meaningful differences between treatment conditions was compromised. Additionally, although counterbalancing the order of test administration helped to minimize the confounding effects of these variables by distributing them across both conditions of the study it does not eliminate them. Finally, the fact that there was only one monolingual English-

speaking tester in this study could be considered a limitation. However, both the bilingual and monolingual testers in this study completed rigorous training requirements before being able to administer the neuropsychological tests to research participants, which helps to minimize this concern.

The generalizability of the study findings is restricted to healthy Spanish speakers with limited English proficiency in San Juan, Puerto Rico. Whether or not the observed findings extend to other populations, including patients with brain damage and neurological disease or individuals from other countries or language groups, remains unknown. It is possible, perhaps even likely, that interpreter effects are more pronounced in patient populations. It also may be the case that interpreter effects differentially impact populations from various immigrant or language backgrounds. For example, Puerto Ricans are considered United States citizens with legal rights and privileges. American influence and presence also has been fairly prominent on the islands since 1898 when Spain ceded Puerto Rico to the United States as an outcome of the Spanish-American War. It is possible that limited-English-proficiency speakers from other backgrounds, particularly those who are undocumented immigrants, may respond differently to an interpreter's presence.

Future Research

In many respects this study was intended to serve as a pilot experiment for future research. The primary emphasis of this work was to determine whether or not interpreter use affected neuropsychological test scores, but it was not designed to address how or why. Future studies are planned to examine the underlying reasons for the observed findings. For example, data from the video recordings will be transcribed for the Vocabulary and Similarities subtests so that participant responses can be compared to interpreter translations for accuracy and to determine the effects of interpretation errors on the test scores. The video data also will provide an opportunity to examine differences in the rate at which interpreters read numbers on the Digit Span subtest or the words they

select to translate the CVLT items to determine their relationships to test scores. Future studies also are planned to investigate the effects of interpreter use on neuropsychological test scores with patient populations and to investigate whether or not such effects influence diagnostic and treatment decisions as part of the more integrated process of neuropsychological assessment. Finally, it would be interesting and important to study interpreter experience and training as a more dynamic variable that potentially could affect neuropsychological test scores.

Conclusions and Main Findings

Taken together, the study findings have important implications for the practice of clinical neuropsychology. Results clearly indicate that use of an interpreter affects scores on verbal measures of neuropsychological functioning, although further research is needed to elucidate the mechanisms that influence the direction of the effects. The observation that interpreter use increases variability in test scores is an equally if not more important finding for clinical practice. It suggests that although interpreter use may not always influence mean score differences at the group level, they could have substantial impact on test scores for individual patients. The fact that significant results were obtained in this study despite using a sample of healthy Puerto Rican adults and relatively well trained interpreters also is important. It suggests that the effects may be even greater with clinical populations and/or untrained interpreters. Results indicate that neuropsychologists should avoid using interpreters to facilitate testing. Unless interpreter-mediated testing in the only practical option (i.e. documented attempts to find appropriate referrals have not been successful), in which case test selection should be limited largely to nonverbal measures that minimize reliance on an interpreter. Moreover, larger confidence intervals should be used when interpreting the clinical significance of observed scores, particularly for verbal tests.

REFERENCES

Abe-Kim, J., Takeuchi, D., Hong, S., Zane, N., Sue, S., Spencer, M.S., Appel, H., Nicdao, E., & Alegria, M. (2007). Use of mental health-related services among immigrant and US-born Asian Americans: Results from the National Latino and Asian American Study. *American Journal of Public Health*, 97, 91-98.

Acevedo, M., Reyes, C.J., Annett, R.D., & Lopez, E.M. (2003). Assessing language competence: Guidelines for assisting persons with limited English proficiency in research and clinical settings. Journal of Multicultural Counseling and Development, 31, 192-204.

Acosta, F.X., & Christo, M. (1982). Bilingual-bicultural interpreters as psychotherapeutic bridges: A program note. *Journal of Community Psychology*, 10. 54-56.

Alegria, M., Mulvaney-Day, N., Torres, M., Polo, A., Cao, Z., & Canino, G. (2007). Prevalence of psychiatric disorders across Latino subgroups in the United States. *American Journal of Public Health,* 97, 68-75.

American Psychological Association (2002). Ethical Principles of Psychologists and Code of Conduct. *American Psychologist*, 57, 1060-1073.

American Educational Research Association; American Psychological Association; National Council on Measurement in Education. (2002). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review,* 15, 185-195.

Arroyo, J.A. (1996). Psychotherapist bias with Hispanics: An analog study. *Hispanic Journal of Behavioral Sciences*, 18, 21-28.

Artiola y Fortuny, L., Heaton, R.K., & Hermosillo, D. (1998). Neuropsychological comparisons of Spanish-speaking participants from the U.S.-Mexico border region versus Spain. *Journal of the International Neuropsychological Society*, 4, 363-379.

Axelrod, B.N., Vanderploeg, R.D., & Schinka, J.A. (1999). Comparing methods for estimating premorbid intellectual functioning. *Archives of Clinical Neuropsychology*, 14, 341-346.

Bailey, D.S. (2004). Number of Psychology Ph.Ds declining. *Monitor on Psychology,* 35, 18-19.

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

Baker, D.W., Hayes, R., & Fortier, J.P. (1998). Interpreter use and satisfaction with interpersonal aspects of care for Spanish-speaking patients. *Medical Care*, 36, 1461-1470.

Baker, F.M. (2004). Diagnosing depression in African Americans. *Community and Mental Health Journal*, 37, 31-38.

Bernal, G. & Enchautegui-de-Jesus, N. (1994). Latinos and Latinas in community psychology: a review of literature. *American Journal of Community Psychology*, 22, 531-557.

Bernstein, J., Bernstein, E., Dave, A., Hardt, E., James, T., Linden, J., Mitchell, P., Oishi, T., & Safi, C. (2002). Trained medical interpreters in the emergency department: Effects on services, subsequent charges, and follow-up. Journal *of Immigrant Health*, 4, 171-176.

Bischoff, A., Perneger, T.V., Bovier, P.A., Loutan, L., & Stalder, H. (2003). Improving communication between physicians and patients who speak a foreign language. *The British Journal of General Practice,* 53, 541-546.

Bond, Charles F., Titus, L. J. (1983). Social facilitation: a meta-analysis of 241 studies. *Psychological Bulletin.* 94(2), Sep 1983, 265-292.

Brickman, M., Cabo, R., & Manly, J. (2006). Ethical issues in cross-cultural neuropsychology. *Applied Neuropsychology*, 13, 91-100.

Bush, S.S., & NAN Policy & Planning Committee (2005). Independent and court-ordered forensic neuropsychological examinations: Official statement of the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology,* 20, 997-1007.

Bush, S.S. & Lees-Haley, P. (2004). Threats to the validity of forensic neuropsychological data: Ethical considerations. *Journal of Forensic Neuropsychology*, 4, 45-66.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Constantinou, M. (2005), Effects of a third-party observer during neuropsychological assessment when the observer is a video camera. *Journal of Forensic Neuropsychology*, 4, 39-48.

Echemendia, R., & Harris, J. (2004). Neuropsychological test use with Hispanic/Latino populations in the United States: Part II of a national survey. *Applied Neuropsychology*, 11, 4-12.

Echemendia, R.J., Harris, J.G., Congett, S.M., Diaz, L.M., & Puente, A.E. (1997). Neuropsychological training and practices with Hispanics: A national survey. *The Clinical Neuropsychologist*, 11, 229-243.

Eytan, A., Bischoff, A., Rrustemi, I., Loutan, L., Gilbert, M., & Bovier, P.A. (2002). *Australian and New Zealand Journal of Psychiatry*, 36, 499-503.

Farooq, S., Fear C., & Oyebode, F. (1997). An investigation of the adequacy of psychiatric interviews conducted through an interpreter. *Psychiatric Bulletin*.

Flores, G. (2005). The impact of medical interpreter services on the quality of health care: A systematic review. *Medical Care Research and Review*, 62, 255-299.

Gavett, Brandon E; Lynch, Julie K; McCaffrey, Robert J.(2005). Third party observers: The effect size is greater than you might think. *Journal of Forensic Neuropsychology*, 4, 49-64.

Ginsberg C, Martin D, Andrulis D, Shaw-Taylor Y, McGregor C (1995). *Interpretation and Translation Services in Health Care: A Survey of US Public and Private Teaching Hospitals.* Washington, DC: National Public Health and Hospital Institute.

Goff, D.C., Nichaman, M.Z., Chan, W., Ramsey, D.J., Labarthe, D.R., & Ortiz, C. (1997). Greater incidence of hospitalized myocardial infarction among Mexican-Americans than non-Hispanic Whites. *Circulation,* 95, 1433-1440.

Hagan, R., Durham, T., & Shannon, D. (2006). Administration of digit span on the Wechsler and Binet: Differences that matter. *Journal of Clinical Psychology*, 33, 480-482.

Hernandez, A., Plant, E.A., Sachs-Ericsson, N., & Joiner, T.E. (2004). Mental health among Hispanics and Caucasians: Risk and protective factors contributing to prevalence rates of psychiatric disorders. *Journal of Anxiety Disorders*, 19, 844-860.

Holtzer, R., Verghese, J., Wang, C., Hall, C.B., & Lipton, R. (2008). Within-person across-neuropsychological test variability and incident dementia. *Journal of the American Medical Association,* 300, 823-830.

Horvath, A.O., & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology*, 64, 561-573.

Hwa-Froelich, D.A. & Westby, C. (2003). Considerations when working with interpreters. *Communication Disorders Quarterly*, 24, 78-85.

Kapborg, I. & Bertero, C. (2002). Using an interpreter in qualitative interviews: does it threaten validity? *Nursing Inquiry*, 9, 52-56.

Kline, F., Acosta, F., & Austin, W. (1980). The misunderstood Hispanic patient. *American Journal of Psychiatry*, 137, 1530-1533.

LaCalle, J.J. (1987). Forensic psychological evaluations through an interpreter: Legal and ethical issues. *American Journal of Forensic Psychology*, 5, 29-43.

Lawson, W.B., Hepler, N., Holladay, J., & Cuffel, B. (1994). Race as a factor in inpatient and outpatient admissions and diagnosis. *Hospital and Community Psychiatry,* 45, 72-74.

Manly, J., Jacobs, D.M., Sano, M., Bell, K., Merchant, C.A., Small, S.A., & Stern, Y. (1999). Effect of literacy on neuropsychological test performance in nondemented, education-matched elders. *Journal of the International Neuropsychological Society,* 5, 191-202.

Marcos, L. (1979). Effects of interpreters on the evaluation of psychopathology in non-English-speaking patients. *American Journal of Psychiatry*, 136, 171-179.

McLaren, P. (1995). White terror and oppositions agency: Towards a critical multiculturalism. In Theo-Goldberg (Eds.), *Multiculturalism: A critical reader* (pp. 56-57). Cambridge, MA: Blackwell.

Merriam-Webster's collegiate dictionary (10th ed). (1993). Springfield, MA: Merriam-Webster.

Miller, K., Martell, Z.L., Pazdirek, L., Caruth, M., &Lopez. D. (2005). The role of interpreters in psychotherapy with refugees: An exploratory study. *American Journal of Orthopsychiatry*, 75, 27-39.

Morales, A. & Hanson, W.E. (2007). Language brokering: An integrative review of the literature. *Hispanic Journal of Behavioral Sciences*, 27, 471-503.

Muhkerjee, S., Shukla, S., Woodle, J., Rosen, A.M., & Olarte, S. (1983). Misdiagnosis of schizophrenia in bipolar patients: A multiethnic comparison. *American Journal of Psychiatry*, 140, 1471-1574.

Newton, R.L. (1950). A comparison of two methods of administering the digit span test. *Journal of Clinical Psychology*, 6, 409-412.

Okazaki, S. & Sue, S. (1995). Methodological issues in assessment research with ethnic minorities. Psychological Assessment, 7, 367-375.

Pavlik, V.N., Hyman, D.J., & Doody, R. (2005). Cariovascular risk factors and cognitive function in adults 30-59 years of age (NHANES-III). *Neuroepidemiology*, 24, 42-50.

Perez-Arce, P. & Puente, P. (1984). Neuropsychological assessment of ethnic minorities: The case of assessing Hispanics living in North America. In Sbordone, R.J. & Long, C.J. (Eds). *Ecological Validity of Neuropsychological Testing*. Boca Raton, FL: CRC Press.

Pfeffer, N. (1998). Theories in health care and research theories of race, ethnicity, and culture. *British Medical Journal,* 317, 1381-1384.

Pollard, R. Q (2001, August). Mental health interpreting: A mentored curriculum. *Connections, 2*, 7-9.

Ponton, M.O. (2001). Research assessment issues with Hispanic populations. In Ponton, M.O.& Leon-Carrion, J. (Eds). *Neuropsychology and the Hispanic patient: A Clinical Handbook*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Rabin, L.A., Barr, W.B., & Burton, L.A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20*, 33-65.

Roselli, M. & Ardila, A. (2002). A Cross-linguistic comparison of verbal fluency tests. *International Journal of Neuroscience,* 112 (6), 759-776.

Rosenberg, E., Leanza, Y., & Seller, R. (2007). Doctor-patient communication in primary care with an interpreter: Physician perceptions of professional and family interpreters. *Patient Education and Counseling*, 67, 286-29.

Russell, G.L., Sue, S., Cheung, M., & Snowden, L.R. (1996). The effects of therapist-client ethnic match in the assessment of mental health functioning. *Journal of Cross-Cultural Psychology*, 27, 598-615.

Sabin, J.E.(1975). Translating despair. *American Journal of Psychiatry*, 132, 197-199.

Schofield, W. (1964). *Psychotherapy: The purchase of friendship*. Englewood Cliffs, N.J.: Prentice-Hall.

Simon, C., Zyzanski, S., Durans, E., Jimenes, X., & Kodish, E. (2006). Interpreter accuracy and informed consent among Spanish-speaking families with cancer. *Journal of Health Communication*, 11, 509-22.

Snowden, L., Masland, M.,& Guerrero, R. (2007). Federal civil rights policy and mental health treatment access for persons with limited English proficiency. *American Psychologist*, 62, 109-117.

Stern, MP, Mitchell, B.D. (1995). *Diabetes in Hispanic Americans*. In: Harris, MI, Cowie, C.C., Stern, M.P., Boyko, E.J., Reiber, G.E., Bennett, P.H. (Eds). Diabetes in America, 2nd ed. (pp.631-659). Bethesda, MD: National Institute of Diabetes and Kidney Diseases.

Sue, S. (1973). Training of 'third-world' students to function as counselors. *Journal of Counseling Psychology*, 1, 73-78.

Sue, D.W., Capodilupo, C.M., Torino, G.C., Bucceri, J.M., Holder, A.M., Nadal, K.L., & Esquilin, M. (2007). Racial microaggessions in everyday life: Implications for clinical practice. *American Psychologist*, 62, 271-286.

Sue, S. & Chu, J. (2003). The mental health of ethnic minority groups: Challenges posed by the supplement to the surgeon general's report on mental health. *Culture, Medicine, and Psychiatry*, 27, 447-465.

Sue, S., Fujino, D.C., Hu, L., & Takeuchi, D.T. (1991). Community mental health services for ethnic minority groups: A test of the cultural responsiveness hypothesis. *Journal of Consulting and Clinical Psychology*, 59, 533-540.

Sue, D.W., & Sue, S. (1987). Cultural factors in the clinical assessment of Asian Americans. *Journal of Consulting and Clinical Psychology*, 55, 479-487.

Sue, S., Allen, D.B., &Conoway, L. (1978). The responsiveness and equality of mental health care to Chicanos and Native Americans. *American Journal of Community Psychology*, 6, 137-146.

Sue, S., McKinney, H., Allen, D., & Hall, J. (1974). Delivery of community and mental health services to black and white clients. *Journal of Consulting and Clinical Psychology*, 42, 794-801.

Takeuchi, D., Sue, S., & Yeh, M. (1995). Return rates and outcomes from ethnicity-specific mental health programs in Los Angeles. *American Journal of Public Health*, 85, 638-643.

Taub, H. (1972). A comparison of young adult and old groups on various digit span tasks. *Developmental Psychology*, 6,0-65.

U.S. Department of Health and Human Services. (2001). *Mental health: Culture, race, and ethnicity.* In A Report of the Surgeon General. Washington, DC: Government Printing Office.

U.S. Bureau of the Census. 2000. We the people: *The Hispanic population in the United States.* In Census 2000 Special Reports. Washington, DC: Government Printing Office.

U.S. Bureau of the Census. 2007. *Selected Social Characteristics of the United States.* In 2007 American Community Survey. Washington, DC: Government Printing Office.

Vega, W.A., Kolody, B., Aguilar-Gaxiola, S., Alderete, E., Catalano, R., & Caraveo-Anduaga. (1998). Lifetime prevalence of DSM-III-R psychiatric disorders among urban and rural Mexican-Americans in California. *Archives of General Psychiatry*, 55, 771-778.

Wallin, A. & Ahlstrom, G. (2006). Cross-cultural interview studies using interpreters: systematic literature review. *Journal of Advanced Nursing*, 55, 723-235.

Weible, J.A., Nuest, B.D., Welty, J., Pate, W.E., & Turner, M.L. (2002). Demonstrating the effects of presentation rate on aging memory using the California Verbal Learning Test (CVLT). *Aging, Neuropsychology, and Cognition,* 9, 38-47.

Woodcock, R. W. & Munoz-Sandoval, A.F. (1993). *Woodcock-Munoz language survey comprehensive manual.* Chicago, IL: Riverside Publishing Company.

Zhang, A., Snowden, L., & Sue, S. (1998). Differences between Asian and White Americans' help seeking and utilization patterns in the Los Angeles area. *Journal of Community Psychology*, 26, 317-326.

APPENDIX A.

SELF-IDENTIFIED ENGLISH PROFICIENY QUESTION

How well do you speak English?

Very Well

Well

Not Well

Not at All

APPENDIX B

TELEPHONE SCREENING QUESTIONS

What was your first language?

Do you speak English and Spanish?

Have you ever been unconscious? If so, for how long?

Have you ever had a seizure?

Have you ever had any type of neurological condition?

Have you ever taken psychiatric medication?

Have you ever been in psychological therapy?

APPENDIX C

METAL HEALTH INTERPRETING: A MENTORED CURRICULUM

Chapter 1: About This Curriculum

Chapter 2: The Role of the Interpreter

Types of Interpreters

The Evolving Role of Sign Language Interpreters

Chapter 3: Ethics in Mental Health Interpreting

Confidentiality

Counseling, Advising, and Adding Personal Opinion

Chapter 4: Mental Health Professionals

Psychiatrists

Psychologists

Social Workers

Psychiatric Nurses

Students, Interns, and Residents

Other Mental Health Workers

Chapter 5:Mental Illnesses and the DSM System

Diagnoses and the DSM System

Common Axis I Diagnoses and Interview Questions

Chapter 6: Mental Health Settings and Clinicians'

Acute Care and Emergency Rooms

Crisis Services

Inpatient Services

Outpatient Services

Other Clinician Objectives

Chapter 7: Interpreting and Dysfluent Patients

Summaries of the Eleven Vignettes

Introductory Video

An example of a recommended pre-session, in which a Spanish language interpreter and a clinician discuss what each other needs to know to prepare for an upcoming intake.

Cultural Bonds

A Russian-speaking patient makes several private side-comments to the interpreter, who then takes a moment to explain his role to her.

Cultural Attitudes Toward Mental Illness

A two-part scene. In part 1, Chinese cultural values factor heavily in a consumer's reluctance to be cooperative with a clinician. In part 2, the interpreter conducts a post-session, explaining the issues and providing useful advice and information to the clinician.

I Can't Do Your Job For You

A clinician inappropriately pressures a Russian interpreter for clinical guidance. He responds only on language and culture issues.

It's a Small Community

An interpreter seeks confidential supervision to deal with questions and stress pertaining to outside knowledge of a psychiatric patient.

Gender, Age, and Culture

In a pre-session, a Vietnamese language interpreter provides a clinician with guidance on how to handle gender and age issues likely to come up in the forthcoming session.

Linguistic and Cultural Barriers to Translation

A clinician conducting a mental status exam runs into several language and cultural problems that hinder the assessment. The Spanish language interpreter provides explanation and some guidance on the spot.

Language and Psychosis

A two-part scene on how to (and how not to) handle interpreting for patients' whose language is disrupted by mental illness. In part 1, the interpreter cannot provide coherent translations and just gives up. In part 2, the interpreter shifts to third person and provides additional language information.

Embarrassing Moments

A two-part scene showing alternate ways to translate a patient's comments that are offensive and embarrassing (a common interpreter concern). In part 1, the Spanish language interpreter translates verbatim; in part 2, by using some third person distance.