



Published in final edited form as:

Clin Neuropsychol. 2012 January ; 26(1): 88–101. doi:10.1080/13854046.2011.640641.

Interpreter-Mediated Neuropsychological Testing of Monolingual Spanish Speakers

Rachel Casas,
University of Iowa

Edmarie Guzmán-Vélez,
University of Puerto Rico, Rio Piedras

Javier Cardona-Rodriguez,
University of Puerto Rico, Rio Piedras

Nayra Rodriguez,
University of Puerto Rico, Rio Piedras

Gabriela Quiñones,
University of Puerto Rico, Rio Piedras

San Juan,
Puerto Rico

Borja Izaguirre, and
University of California, Los Angeles

Daniel Tranel
Department of Neurology (Division of Behavioral Neurology and Cognitive Neuroscience),
University of Iowa College of Medicine

Abstract

The primary objective of this study was to investigate empirically whether using an interpreter to conduct neuropsychological testing of monolingual Spanish speakers affects test scores. Participants included 40 neurologically normal Spanish-speakers with limited English proficiency, ages 18–65 years ($M=39.7$, $SD=13.9$), who completed the Vocabulary, Similarities, Block Design, and Matrix Reasoning subtests of the Wechsler Adult Intelligence Scale-III in two counterbalanced conditions: with and without an interpreter. Results indicated that interpreter use significantly increased scores on Vocabulary and Similarities. However, scores on Block Design and Matrix Reasoning did not differ depending upon whether or not an interpreter was used. In addition, the findings suggested a trend toward higher variability in scores when an interpreter was used to administer Vocabulary and Similarities; this trend did not show up for Block Design or Matrix Reasoning. Together, the results indicate that interpreter use may significantly affect scores for some tests commonly used in neuropsychological practice, with this influence being greater for verbally mediated tests. Additional research is needed to identify the types of tests that may be most affected as well as the factors that contribute to the effects. In the meantime, neuropsychologists are encouraged to avoid interpreter use whenever practically possible, particularly for tests with high demands on interpreter abilities and skills, with tests that have not been appropriately adapted and translated into the patient's target language, and with interpreters who are not trained professionals.

Hispanics represent the largest and fastest growing ethnic minority group in the United States (United States Census Bureau, 2009). Among Hispanics ages five and older, an estimated 76% speak Spanish at home and approximately 47% are believed to require

assistance in English in at least some situations (United States Census Bureau, 2010). These demographic changes have placed new demands on the health care workforce (Health Resources and Services Administration, 2003) including psychology (Clay, 2009a, 2009b; Daw, 2002). Health services which require specialty training, such as clinical neuropsychology, may be especially vulnerable to new challenges associated with demographic shifts because fewer bilingual individuals seek the additional education and training requirements necessary for specialization (Rivera Mindt, Byrd, Saez, & Manly, 2010). Delivery of culturally and linguistically appropriate neuropsychological services, particularly for the growing population of Spanish-speaking patients with limited English proficiency (LEP), is among the biggest challenges facing clinical neuropsychology, as there is a critical shortage of bilingual neuropsychologists available to meet the demand for services from Spanish-speaking patients.

The ability to speak more than one language is becoming increasingly important in clinical practice. According to the American Psychological Association's Center for Workforce Studies (American Psychological Association, 1999, 2010), new psychologists consistently identify bilingualism and the ability to work in multiple languages as a "most useful" skill. However, the best available evidence suggests that the number of bilingual neuropsychologists in practice pales in comparison to the demand for services from Spanish-speaking patients. A national survey of clinical neuropsychologists' assessment and treatment practices with Hispanic patients in the United States found that more than 88% of respondents were unable to read, write, or speak Spanish either "adequately" or "fluently" (Echemendia, Harris, Congett, & Puente, 1997). Since the survey data were published more than a decade ago, the Hispanic population has continued to increase, contributing to more than half of the total United States population growth (Fry, 2008). A significant body of research has simultaneously shown that Hispanics are at increased risk for medical illnesses with known cognitive sequelae, such as diabetes, heart disease, and HIV/AIDS (e.g. Centers for Disease Control and Prevention, 2010; Cooper et al., 2000; Venkat Narayan, Boyle, Thompson, & Williamson, 2003). Taken together, these data suggest that Spanish speakers with LEP may be particularly in need of neuropsychological services, but there are not enough bilingual neuropsychologists available to meet the demand for referrals.

Under-representation of ethnic minorities in neuropsychology is a realistic concern that has garnered increased attention over the past few years (e.g. Byrd, Razani, Lafosse, Manly, & Attix, 2010; Manly, 2008; Rivera Mindt et al., 2010). Recent efforts have focused on the so-called "broken pipeline" within neuropsychology and the fact that, as a field, we have not prioritized strategic initiatives to identify, recruit, retain, and support potential neuropsychologists from diverse backgrounds early on in their careers. Several recommendations have been identified to increase representation of ethnic minority students in clinical neuropsychology training programs. The hope is that such endeavors will diversify the neuropsychology workforce and improve access to quality neuropsychological services for patients from various cultural backgrounds, including those with LEP. However, it will likely take years, perhaps even decades, to fix the broken pipeline in the field. Moreover, although cultural and linguistic diversity are related, they are not synonymous. For example, it is possible to increase the pipeline of culturally diverse neuropsychologists without increasing the number of Spanish-speaking clinicians and vice versa. Specific strategies aimed at improving the Spanish-speaking abilities of neuropsychologists in practice are also urgently needed. In the meantime, clinical neuropsychologists continue to confront practical dilemmas related to neuropsychological service delivery for Spanish-speaking patients who do not speak English fluently.

It is probably not controversial to suggest that under ideal circumstances most neuropsychologists would agree that Spanish-speaking patients with LEP should be

evaluated by Spanish-speaking neuropsychologists rather than through an interpreter. Some have even argued that it is almost obligatory (Artiola i Fortuny & Mullaney, 1997), citing various Standards of the *Ethical Principles of Psychologists and Code of Conduct* (American Psychological Association, 2002). Standard 9.03(c), for example, states that, “psychologists use assessment methods that are appropriate to an individual’s language preference and competence.” However, the ethics code does not specify whether or not the use of interpreters constitutes an appropriate, albeit alternative, method for providing psychological services that are consistent with a patient’s language preference. The *Guidelines for Providers of Psychological Services to Ethnic, Linguistic, and Culturally Diverse Populations* (American Psychological Association, 1990) offer more specificity, stating that, “psychologists interact in the language requested by the client, and, if this is not feasible, make an appropriate referral.” In addition, an educational paper from the National Academy of Neuropsychology asserts that, “ideally, the examinee is evaluated in his or her preferred and/or best language” (Judd, Capetillo, Carrion-Baralt, Marmol, San Miguel-Montes, Navarrete, Puente, Romero, Vales, 2009). However, there is still ambiguity. Standard 2.02 of the ethics code 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.”

It appears as though there is general consensus regarding what constitutes “best practice” for neuropsychological assessment of LEP patients. However, the practical challenge remains in that there are simply not enough bilingual neuropsychologists available to meet the demand for services. Making an appropriate referral may not always be a viable option and can present significant financial, logistical, and/or health-related hardships for some patients. As an example, the nearest bilingual neuropsychologist may be several hours away. For patients who are physically too ill or otherwise unable to travel long distances this simply may not be a viable option. Even when distance is not an obstacle other barriers may exist. Because bilingual neuropsychologists are in such demand, they may have waiting lists of up to several months. Postponing a neuropsychological evaluation for an extended period of time could have detrimental health consequences for some patients, e.g., those who require cognitive evaluations prior to neurosurgery or organ transplant. Additional factors, such as insurance coverage restrictions, may also limit patients’ abilities to seek services from more linguistically appropriate referrals. The referral source may also lack the experience, expertise, or competence to evaluate the patient’s presenting concerns. Neuropsychology is a specialty field with advanced education, training, and supervision requirements. However, even within this specialized area of practice, there are more finely graded areas of expertise. As an example, a neuropsychologist who has worked exclusively with older adults may not be competent to evaluate children. The benefits of conducting an evaluation in a patient’s native language need to be weighed carefully against the expertise and clinical competence of the clinician, particularly for high-stakes evaluations such as WADA testing, intra-operative brain mapping, or forensic evaluations. Regardless of the circumstances, neuropsychologists have an ethical obligation to provide LEP patients with options for linguistically appropriate referrals (American Psychological Association, 1990, 2002). However, clients can decline or refuse referrals, and for these and other legitimate reasons neuropsychologists are sometimes left with no realistic option but to conduct neuropsychological evaluations of LEP patients with the assistance of an interpreter.

To the best of our knowledge, the survey by Echemendia and colleagues (1997) is the only one of its kind to have published data regarding neuropsychologists’ use of “translators.” However, use of the term “translator” is probably inappropriate in this context, and arguably, provides evidence in itself of the nascent state of scientific inquiry and investigation into this area of work within the neuropsychological literature. The roles of

interpreters and translators are not synonymous, and yet the terms are often used interchangeably in the limited literature that exists within the field. Interpreters communicate oral information from one language into a different language using spoken words. Translators, on the other hand, communicate written information from one language into another using text. The distinction is important because although both interpreters and translators require excellent linguistic abilities in the source and target languages, they are employed in different contexts and rely on unique skill sets (Gile, 1995). Interpreters are sometimes asked to provide “sight translations” of documents and instruments on the spot, which may best reflect the experiences that Echemendia and colleagues were trying to capture in their study. Results of the survey were striking and revealed that more than 70% of respondents reported using “translators” to provide services to Spanish-speaking patients at least “sometimes.” A majority of these neuropsychologists, 80%, reported that the “translators” they used had not received any formal psychological or neuropsychological training.

The Current Study

Given the paucity of published data regarding the use of interpreters in neuropsychological settings and the ethical debate surrounding their use, research in this area is urgently needed. Careful empirical work, in particular, is needed to objectively determine the effects of interpreter use on various neuropsychological outcomes. The primary objective of the current study was *to determine whether using an interpreter to conduct neuropsychological testing of Spanish-speaking patients with LEP would affect neuropsychological test scores. Our goal was to use a rigorous experimental design and to examine the effects 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 advanced training in neuropsychology (the so-called “Houston model” training). 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 armamentarium, namely, neuropsychological tests. Our goal was to focus on a circumscribed battery of neuropsychological tests with high utility, applicability, and flexibility for neuropsychologists in clinical practice. A review of the existing literature regarding patterns of neuropsychological test use revealed that the Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III) was ranked by practicing neuropsychologists as the most frequently used instrument in their assessment batteries (Rabin, Barr, & Burton, 2005). Subtests of the WAIS-III were consistently rated among the top tests used to assess specific cognitive domains such as memory, attention, and executive functioning as well as to evaluate readiness to return to work. Nearly 90% of respondents also identified intelligence assessment as a common referral question in their practice. Furthermore, neuropsychologists use estimates of premorbid intelligence to help formulate interpretations about the meaning of scores on other neuropsychological tests, making it paramount that scores on such measures are reliable and valid. For these reasons, we focused on a set of tests from the WAIS-III in this initial study of interpreter-mediated neuropsychological testing.

Method

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 participants was 14.00 years ($SD=1.99$). All but one of the participants completed their education in Puerto Rico.

Participants in this study were native Spanish-speakers with LEP who were between the ages of 18–64. Adults older than 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). Individuals who self-reported a history of head trauma, stroke, seizures, dementia, or other neurological conditions or psychiatric history also were excluded. Limited English Proficiency was defined as having Limited, Very Limited, or Negligible English language skills on the Oral Language cluster of the Woodcock-Munoz Language Survey – Revised (WMLS-R).

Materials

Four subtests from the Wechsler Adult Intelligence Scale -3rd Edition (WAIS-III) were included in this study - Vocabulary, Similarities, Block Design, and Matrix Reasoning. These four subtests from the WAIS-III have parallel forms on the Wechsler Abbreviated Scale of Intelligence (WASI), an instrument that ranked as number 22 on the “Top 40” list of most frequently used neuropsychological tests (Rabin et al, 2005). These four subtests are especially useful because an estimate of full scale IQ, Verbal IQ, and Performance IQ can be derived from their scores within approximately 30 minutes, an important consideration when time constraints and other factors prohibit a lengthy test battery. Additionally, because interpreter use could differentially affect scores on verbal versus nonverbal tests it was important to include both types of measures in this initial investigation. The verbal tests were administered without regard to reverse or discontinuation rules. Other studies have demonstrated that the hierarchy of item difficulty on some verbal neuropsychological tests is not equivalent across languages (Kohnert, Hernandez, & Bates, 1998). Thus, we reasoned that it was an appropriate first step to administer all of the verbal test items and to use raw scores when making comparisons between study conditions.

Procedure

A within-subjects repeated-measures design was used in this study; participants completed the four neuropsychological measures twice under different conditions. In one condition, a monolingual English-speaking tester administered the tests via a bilingual Spanish/English speaking interpreter (i.e. Interpreter-Mediated condition, abbreviated “I”). In the other condition, a psychometrist administered the measures directly in Spanish (No-Interpreter condition, abbreviated “No- I”). Both the monolingual ($n = 1$) and bilingual testers ($n = 4$) received rigorous training in the administration of the neuropsychological measures, which included providing explanations for the purposes of each test, specific and detailed instructions for their administration, observation of the principal investigator administering the tests, and supervised practice administering the tests on volunteers. After first demonstrating their competence to administer the tests in English by administering two flawless exams on two pilot participants, bilingual testers received additional training in their Spanish administration. They were also required to administer two flawless exams on two pilot participants in Spanish. An objective rubric was used to measure test administration competence (e.g. started at the appropriate test item, applied reverse and discontinuation rules appropriately, queried appropriately, etc.). Test instructions were

included in a manual which testers were required to use during each session. 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 conditions. The length of time between the first and second testing sessions ranged from 3 weeks to 5 months.

Participant Recruitment

A broad-based recruitment 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.

Several strategies were used to facilitate recruitment and retention of participants. To assist with transportation costs, participants were paid \$5.00 in cash on each day of testing. Participants also were compensated \$20.00 by check after completing each condition of the study. Snacks (cookies and water) were also provided during each testing session. Finally, reminder phone calls and thank you cards were mailed to participants.

A total of 193 individuals inquired about participation and completed telephone screening. Of these, 102 individuals met all of the eligibility criteria and were scheduled for a testing appointment. The majority of individuals were excluded because their English language proficiency was rated as too advanced, they reported that they were taking medication for a psychiatric condition, or they reported a history of head injury. To operationalize level of English language proficiency during the telephone screening, we asked participants several questions in English and instructed them to provide English responses. The questions, which were open-ended, were intentionally designed to elicit samples of their English language fluency and included questions such as, “Can you tell me what you did today before you called us?” and “Can you tell me how you would get from your house to the University of Puerto Rico at Rio Piedras?” Participants who were judged to be able to answer the questions (in English) with appropriate content, in complete sentences, without hesitation, and with a high level of subject-verb agreement were excluded. Of the 102 individuals who met the screening criteria and were enrolled in the study, 52 completed at least one testing condition. A total of 40 participants completed both testing conditions of the study.

Test Translations

Instructions for all tests were translated from English into Spanish using a consensus translation strategy. Stimuli for the Vocabulary and Similarities subtests were also translated from English into Spanish using consensus translation. Four bilingual psychometrists independently translated the test instructions and items from English into Spanish. Their translations were then entered into a database and compared to each other. Discrepancies in translation were discussed and collectively negotiated until consensus was achieved. We selected a consensus method of translation because studies have demonstrated that the quality of translations is better when undertaken by teams rather than individuals (Sumathipala & Murray, 2000). In addition, the quality and content validity of translated items is not necessarily improved using other methods, such as back-translation approaches. The purpose was to develop standardized language in the No-I condition of the study in order to determine the effect of interpreter use on test scores. We reasoned that it was not appropriate to use existing Spanish language tests in this initial study (e.g., the Escala de Inteligencia de Wechsler para Adultos – Tercera Edición, the Spanish version of the WAIS-III) because it would be impossible to determine if observed differences in tests scores were

attributed to the affects of interpreter use or to non-equivalence of the measures. The within-subjects, repeated measures design that was used in this study also enabled participants to serve as their own “control group,” thereby minimizing some of the challenges and concerns related to the equivalence of translated tests. Importantly, the test translations were not designed or intended for clinical use. Rather, the goal in this initial study was to isolate the experimental effect of interpreter use on neuropsychological test scores using an empirical design. Artioli i Fortuny and colleagues (2005) have described the inherent problems with using translated tests in clinical and research settings.

Testing Procedures for the Interpreter Condition

Interpreters met with the monolingual tester before every testing session for approximately 15–30 minutes. During that time, the tester introduced the interpreter to the neuropsychological tests and briefly reviewed the typical administration instructions. The intent was to model real-life clinical settings in which there is often limited preparation time before meeting with (and testing) patients. Interpreters were allowed to review the testing protocols during this “pre-session,” and they were encouraged to write translations for the items in the Vocabulary and Similarities subtests in advance. However, this was not always possible, and some interpreters provided translations for some words during the testing session itself. For all four subtests, the monolingual tester read the instructions in English and the interpreter communicated the directions back to the participant in Spanish using a consecutive style of interpretation. Similarly, the interpreter communicated participant responses from Spanish into English for the monolingual tester to record and score.

Interpreter Language Proficiency

Interpreters ($n = 6$) were native Puerto Ricans who had completed their elementary and high school education in Spanish, but were not professional interpreters. Interpreters were required to pass the English version of the Oral Proficiency Interview (OPI) of the American Council of Teaching and Foreign Languages (ACTFL) at a minimum level of Advanced-Low. According to the ACTFL Proficiency Guidelines, English speakers at the Advanced-Low 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.”

Interpreter Training

In an effort to enhance ecological validity and model the reality of interpreter use in many neuropsychological settings, interpreters were not given explicit a priori training on how to administer the neuropsychological measures. However, there has been a recent trend to provide interpreters who work in health care contexts with more general training for how to interpret in mental health settings. Interpreters in this study received general training on effective strategies for mental health interpreting through a curriculum developed by the Deaf Wellness Center at the University of Rochester titled *Mental Health Interpreting: A Mentored Curriculum* (e.g. Pollard, 2001). 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. 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.

Results

Raw scores were used for all statistical analyses. Because our experimental manipulation required a deviation from standardized administration of the neuropsychological measures, the validity of applying any normative data set to our sample was jeopardized. In addition, there are inherent challenges regarding the selection of appropriate norms for this population of monolingual Spanish-speakers. Thus, raw scores were deemed most appropriate.

Scores for each of the dependent variables were calculated by summing the total number of correct items on each measure. Total scores were then compared between the “I” and “No-I” conditions of the study. Data were analyzed using a repeated-measures multiple analysis of variance (MANOVA) with one independent variable (test condition) and four dependent variables (total scores on Vocabulary, Similarities, Block Design, and Matrix Reasoning). Order of test condition administration was included as a between-subjects factor to determine the effects of counterbalancing the study conditions.

Because participants completed the same battery of neuropsychological tests twice, it also was important to examine the potential influence of practice effects. Thus, in a separate analysis 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 versus Time 2). A repeated measures MANOVA with testing session as the independent variable and four dependent variables was conducted. Univariate analyses were computed as follow up statistics for significant omnibus MANOVA results.

Descriptive statistics for each of the four dependent variables are included in Table 1. Results revealed a significant main effect of test condition, Pillai’s Trace = .625, $F(4, 35) = 14.60$, $p < .001$, indicating that total scores were different between the “I” and “No-I” conditions of the study on at least one neuropsychological test. Follow-up univariate analyses revealed that mean scores were higher when an interpreter was used to administer both Vocabulary, $F(1,38) = 48.51$, $p < .001$, $d = 1.16$, and Similarities, $F(1,38) = 9.92$, $p = .003$, $d = .42$, but not Block Design, $F(1,38) = .053$, $p = .82$, or Matrix Reasoning, $F(1,38) = .880$, $p = .35$.

The main effect of order of test administration was not significant, Pillai’s Trace = .165, $F(4, 35) = 1.74$, $p = .17$. The interaction between test condition and order of test administration also was not significant, Pillai’s Trace = .034, $F(4,35) = .034$, $p = .87$.

There was no significant effect of practice on neuropsychological test scores, Pillai’s Trace = .070, $F(4, 36) = .681$, $p = .61$.

Examination of Table 1 revealed that standard deviations were larger when an interpreter was used to administer three of the four neuropsychological tests (Matrix Reasoning, Similarities, and Vocabulary). F-tests were therefore used to objectively determine the equality of variances for each dependent measure across the study conditions. There was a trend toward higher variability in the I condition of the study for Vocabulary, $F(39, 39) = 1.62$, $p = .07$, and a more marginal effect for Similarities, $F(39,39) = 1.47$, $p = .12$. There were no significant differences in variability when an interpreter was used to facilitate Block Design, $F(39,39) = 1.03$, $p = .46$ or Matrix Reasoning, $F(39,39) = 1.19$, $p = .30$.

Discussion

Of the 45 million Spanish speakers currently living in the United States, slightly fewer than half of them are proficient in English (United States Census Bureau, 2007). The Hispanic population will continue to increase over the next several decades and the number of

Spanish speakers with LEP also is likely to grow. As a profession, neuropsychology is under-prepared to meet the marketplace demands from this culturally and linguistically diverse population. Although initiatives are currently underway to recruit more bicultural and bilingual neuropsychologists into the field, such efforts will likely take several years to reach fruition. In the meantime, many neuropsychologists are left with few other options but to use language interpreters to facilitate services and treatment of patients with LEP. However, little is known regarding the effects of this kind of non-standardized assessment on neuropsychological test scores. Clinicians and researchers have strongly cautioned against the use of interpreters to facilitate neuropsychological testing based on clinical experiences, observations, and anecdotal evidence. Some have even suggested that their use is unethical. However, there is a critical shortage of bilingual neuropsychologists available to meet the demand for services from Spanish-speaking patients, and there is an absence of empirical literature to guide their work. To the best of our knowledge, this study is the first to investigate the effects of interpreter use on neuropsychological test scores in a sample of 40 Spanish speakers with LEP using a rigorous experimental design.

As a first attempt to examine the question, this study focused on four commonly used verbal and nonverbal tests from the WAIS-III: Vocabulary, Similarities, Block Design, and Matrix Reasoning. Results revealed that interpreter use can affect scores on neuropsychological tests. Specifically, mean total scores were higher when an interpreter was used to administer the Vocabulary and Similarities tests, but not Block Design or Matrix Reasoning. This finding suggests that verbal tests may be more sensitive to measurement error associated with interpreter use than nonverbal tests. Neuropsychological measures that rely heavily on verbal expression increase communication demands on interpreter abilities, skills, and attention. Although many nonverbal tasks require minimal reliance on the interpreter beyond translation of the test instructions, most verbal measures cannot be administered without significant assistance in translating test items, participant responses, and examiner queries as well.

One potential explanation is that tests that require more engagement and interaction between the interpreter, patient, and examiner provide more opportunities for information to become “lost in translation” or otherwise miscommunicated. Effect sizes were higher for the Vocabulary test compared to the Similarities test, which lends some support to this hypothesis. Several items on the Similarities test can be answered correctly with a simple one-word response. For example, the first administered item on the test asks, “In what way are a piano and a drum alike?” A full-credit score on this item can be achieved with the answer, “instruments.” In comparison, the first administered item on the Vocabulary test states, “Tell me what winter means.” To obtain a full-credit score on this item, individuals must provide a more elaborate verbal response, such as, “the cold season of the year” or “the season of the year between fall and spring.” Although there are items on the Vocabulary test that can receive a full-credit score with a single-word response, there are many fewer and they tend to be more difficult, nuanced items.

Interpreters may also make more errors when demands on their attention or skills increase. Errors such as omitting, distorting, or elaborating upon participant responses or examiner instructions and feedback could have meaningful influences on test scores. As an example, interpreters may have “edited” participant responses on the Vocabulary and Similarities tests in a way that unintentionally biased their answers, providing examiners with inaccurate information upon which to base their scores. Differences between 0, 1, or 2 point responses on these tests can sometimes be subtle and difficult to distinguish. Minor substitutions in word choice can affect item level scores and slight changes in response content can make a difference between a higher or a lower score. In addition, participant responses can be vague. Without specialized training or familiarity with the tests, it can be very difficult to

understand and differentiate between types of responses. Whereas neuropsychologists and psychometricians receive intensive and explicit training to be able to detect nuanced differences in participant responses, interpreters typically are not even acquainted with the purposes of the tests let alone the specific details of their administration and scoring criteria.

Of equal if not more relevance to neuropsychologists in practice, results from this study suggest that scores on some neuropsychological tests may not be as reliable when administered via an interpreter. Scores on the Vocabulary and Similarities tests were inflated when an interpreter was used, but there was a trend for scores to be more variable as well, particularly for the Vocabulary test. This finding is particularly important because the Vocabulary test is used not only as a measure of receptive and expressive language, but as a neuropsychological “hold” test as well. The so-called “hold” tests consist of a small number of measures that are believed to be fairly robust to age-related cognitive decline and various types of brain injury. Such measures are commonly used to derive estimates of premorbid intelligence (Axelrod et al., 1999) and as a baseline for comparing scores from other tests in a neuropsychological battery (Lezak et al., 2011). Inflated and unreliable scores on the Vocabulary test, therefore, can have broad and clinically important effects on the interpretation of scores in an entire neuropsychological test battery. This study compared differences in mean scores on a circumscribed battery of tests that were administered both with and without an interpreter. However, the results suggest that future research should explicitly examine the affects of interpreter use on the psychometric properties of tests as well.

Strengths and Weaknesses

To the best of our knowledge, this study represents the first attempt to empirically investigate whether and how interpreter use affects scores on neuropsychological tests. A within-subjects, repeated-measures design was used to increase efficiency, reduce variability, and maximize power. Nevertheless, the sample size was not large, and the power to detect meaningful differences between treatment conditions may have been compromised. Additionally, although counterbalancing the order of test administration appeared to successfully minimize the confounding effects of practice, it may have attenuated interpreter effects in a way that is difficult to capture quantitatively. The generalizability of these findings is restricted to healthy Spanish speakers with limited English proficiency living in San Juan, Puerto Rico. Whether or not the observed findings extend to other populations is an empirical question, of course, one that should be investigated in future studies. Moreover, interpreters in this study were not professional interpreters, and future work might benefit from an examination of interpreter competence, which is distinct from bilingual competence, on neuropsychological test scores. It is possible, perhaps even probable, that professional interpreters are more reliable in their translations, which could make a meaningful difference, and this possibility should be rigorously investigated. The style of interpretation also may impact neuropsychological test outcomes. Interpreters in this study used a consecutive style of interpretation that included sight translations of some items that were developed and standardized for monolingual English-speakers. Whether or how interpreters can be used to administer properly translated and adapted tests, or tests that were constructed in Spanish, also needs to be further investigated. Finally, this study investigated reliability of interpreter use for administration of neuropsychological tests. Additional research is needed to address normative and validity issues related to test translation and adaptation that result from interpreter use.

Conclusion and Main Findings

Neuropsychologists who work with interpreters may be at risk for erroneously interpreting an individual’s performance as “normal” (false negative) or impaired (false positive) for

some neuropsychological tests. Precision and reliability of measurement may be compromised when an interpreter is used to administer neuropsychological tests, at least for some tests. Additional research is urgently needed to determine which tests are most susceptible to interpreter effects, under what conditions, and with what types of populations. This study used a sample of healthy adults, but interpreter effects may be even greater with clinical populations, where factors such as inattention, impaired comprehension, impaired expression, and impaired memory are frequently part of the equation. Effects may also differ as a function of interpreter training, interpretation style, language proficiency, interpersonal style, and individual variability. These factors combined with the preliminary results of this study suggest that—when practically feasible—neuropsychologists should avoid using interpreters to facilitate testing of LEP patients, particularly if the interpreters are not professionals and are required to perform consecutive interpretation with sight translations. If interpreter-mediated testing is the only practical option (i.e., attempts to find appropriate referrals have not been successful), test selection should be limited to measures that minimize reliance on the interpreter. As for any occasion when there is a need to deviate from standardized testing protocol, a detailed explanation of this departure and its limitations should be included in the neuropsychological report.

Acknowledgments

Supported by APA DPN Fellowship 5T32 MH18882 (R.C.) and NINDS P50 NS19632 (D.T.)

References

- American Psychological Association. Now that I have a degree in psychology, what skills are most useful? What new doctorates have to say. 1999. Retrieved January 12, 2011, from <http://www.apa.org/workforce/snapshots/1999/useful-skills.aspx>
- American Psychological Association. Ethical Principles of Psychologists and Code of Conduct. *American Psychologist*. 2002; 57:1060–1073. [PubMed: 12613157]
- American Psychological Association (Producer). What can I do with a degree in psychology? [Power Point Presentation]. 2010. Retrieved January 12, 2011 from <http://www.apa.org/workforce/presentations/2010-psychology-degree.pdf>
- Artiola i Fortuny L, Garolera M, Romo DH, Feidman E, Barillas HF, Keefe R, Lemaître MJ, Martín AO, Mirsky A, Monguió I, Morote G, Parchment S, Parchment LJ, Da Pena E, Politis DG, Sedó MA, Taussik I, Valdivia F, De Valdivia LE, Maestre KV. Research with Spanish-speaking populations in the United States: Lost in the translation: A commentary and a plea. *Journal of Clinical and Experimental Neuropsychology*. 2005; 27(5):555–564. [PubMed: 16019632]
- Artiola i Fortuny L, Mullaney HA. Assessing patients whose language you do not know: Can the absurd be ethical? *The Clinical Neuropsychologist*. 1998; 12:113–126.
- Axelrod BN, Vanderploeg RD, Schinka JA. Comparing methods for estimating premorbid intellectual functioning. *Archives of Clinical Neuropsychology*. 1999; 14:341–346. [PubMed: 14590588]
- Byrd D, Razani J, Lafosse JM, Manly J, Attix D. Diversity Summit 2008: Challenges in the recruitment and retention of ethnic minorities in neuropsychology. *The Clinical Neuropsychologist*. 2010; 24:1279–1291. [PubMed: 21108149]
- Centers for Disease Control and Prevention. Estimated lifetime risk for diagnosis of HIV infection among Hispanics/Latinos--37 states and Puerto Rico, 2007. *Morbidity and Mortality Weekly Report*. 2010; 59(40):1297–1301. [PubMed: 20948507]
- Clay RA. A new vision for American healthcare. *Monitor on Psychology*. 2009a; 40(4):16.
- Clay RA. Repairing psychology's leaky pipeline. *Monitor on Psychology*. 2009b; 40(10):56.
- Cooper R, Cutler J, Desvigne-Nickens P, Fortmann SP, Friedman L, Havlik R, Thom T. Trends and disparities in coronary heart disease, stroke, and other cardiovascular disease in the United States: Findings of the National Conference on Cardiovascular Disease Prevention. *Circulation*. 2000; 102:3137–3147. [PubMed: 11120707]
- Daw J. Responding to America's changing demographics. *Monitor on Psychology*. 2002; 33(1):72.

- Echemendia RJ, Harris JG, Congett SM, Puente AE. Neuropsychological testing and practices with hispanics: A national survey. *The Clinical Neuropsychologist*. 1997; 11(2):229–243.
- Fry, R. Latino settlement in the new century: Pew Hispanic Center. Oct 23. 2008 Retrived January 12, 2010 from <http://pewhispanic.org/reports/report.php?ReportID=96>
- Daniel, Gile. *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam/ Philadelphia: John Benjamins Publishing Company; 1995b.
- Health Resources and Services Administration. *Changing demographics: Implications for physicians, nurses, and other health workers*. Spring. 2003 Retrieved January 12, 2011, from <ftp://ftp.hrsa.gov/bhpr/nationalcenter/changedemo.pdf>
- Lezak, MD.; Howieson, DB.; Bigler, E.; Tranel, D. *Neuropsychological assessment*. 5. New York: Oxford University Press; in press
- Manly J. Critical issues in cultural neuropsychology: Profit from diverisy. *Neuropsychology Review*. 2008; 18:179–183. [PubMed: 18814033]
- Pollard RQ. Mental health interpreting: A mentored curriculum. *Connections*. 2001 Aug.2:7–9.
- Rivera Mindt M, Byrd D, Saez P, Manly J. Increasing culturally competent neuropsychological services for ethnic minority populations: A call to action. *The Clinical Neuropsychologist*. 2010; 24:429–453. [PubMed: 20373222]
- Rabin LA, Barr WB, Burton LA. 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*. 2005; 20:33–65. [PubMed: 15620813]
- Sumanthipali A, Murray J. New approach to translating instruments for cross-cultural research: a combination of qualitative and quantitative approach for translation and consensus generation. *International Journal of Methods in Psychiatric Research*. 2000; 9(2):87–95.
- United States Census Bureau. Census Bureau estimates nearly half of children under age 5 are minorities. May 14. 2009 Retrieved from U.S. Census Bureau Newsroom Releases website: <http://www.census.gov/newsroom/releases/archives/population/cb09-75.html>
- Venkat Narayan KM, Boyle JP, Thompson TJ, Williamson DF. Lifetime risk for diabetes mellitus in the United States. *The Journal of the American Medical Association*. 2003; 290(4):1884.

Table 1

Descriptive Statistics of the Dependent Variables Across the Study Conditions

Variables	Interpreter Condition			No-Interpreter Condition		
	<i>M</i>	<i>SD</i>	95% CI	<i>M</i>	<i>SD</i>	95% CI
Block Design	31.65	9.90	[28.58, 34.72]	31.70	10.15	[28.55, 34.85]
Matrix Reasoning	13.65	5.28	[12.28, 15.02]	14.18	4.83	[12.68, 15.68]
Similarities	20.65	5.32	[19.0, 22.3]	18.53	4.39	[17.17, 19.89]
Vocabulary	40.65	8.29	[38.49, 42.81]	31.98	6.50	[29.97, 33.99]

Note. *N*=40