

8-1-2012

Construct and Criterion Validity of the Rey Auditory Verbal Learning Test-Spanish Version in Adults with Traumatic Brain Injury

Cristobal Neblina
University of Nevada, Las Vegas

Follow this and additional works at: <https://digitalscholarship.unlv.edu/thesesdissertations>



Part of the [Clinical Psychology Commons](#), and the [Psychiatry and Psychology Commons](#)

Repository Citation

Neblina, Cristobal, "Construct and Criterion Validity of the Rey Auditory Verbal Learning Test-Spanish Version in Adults with Traumatic Brain Injury" (2012). *UNLV Theses, Dissertations, Professional Papers, and Capstones*. 1686.

<http://dx.doi.org/10.34917/4332667>

This Thesis is protected by copyright and/or related rights. It has been brought to you by Digital Scholarship@UNLV with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Thesis has been accepted for inclusion in UNLV Theses, Dissertations, Professional Papers, and Capstones by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.

CONSTRUCT AND CRITERION VALIDITY OF THE REY AUDITORY VERBAL
LEARNING TEST-SPANISH VERSION IN ADULTS
WITH TRAUMATIC BRAIN INJURY

By

Cristobal Neblina

Bachelor of Arts
University of Nevada, Las Vegas
2008

A thesis submitted in partial fulfillment
of the requirements for the

Master of Arts in Psychology

Department of Psychology
College of Liberal Arts
The Graduate College

University of Nevada, Las Vegas
August 2012



THE GRADUATE COLLEGE

We recommend the thesis prepared under our supervision by

Cristobal Neblina

entitled

Construct and Criterion Validity of the Rey Auditory Verbal Learning Test-Spanish
Version in Adults with Traumatic Brain Injury

be accepted in partial fulfillment of the requirements for the degree of

Master of Arts in Psychology
Department of Psychology

Daniel Allen, Ph.D. Committee Chair

Christopher L. Heavey Ph.D. Committee Member

Jefferson W. Kinney, Ph.D. Committee Member

Liza San Miguel-Montes, Ph.D. Committee Member

Merrill Landers, Ph.D. Graduate College Representative

Tom Piechota, Ph.D., Interim Vice President for Research &
Dean of the Graduate College

August 2012

Abstract

Construct and Criterion Validity of the Rey Auditory Verbal

Learning Test-Spanish Version in Adults

with Traumatic Brain Injury

By

Cristobal Neblina, B.A.

Dr. Daniel N. Allen, Examination Committee Chair

Professor of Psychology

University of Nevada, Las Vegas

The Rey Auditory Verbal Learning Test (RAVLT) is among the most commonly used English-language neuropsychological tests of verbal learning and memory.

Previous research supports the validity and clinical utility of adaptations of the RAVLT into many diverse languages. In the United States, Hispanics represent the largest and fastest-growing ethnic minority group. As the Hispanic populace continues to grow, so does the need for empirically validated Spanish-language neuropsychological measures. In 2002, a Spanish adaptation of the RAVLT was developed in Puerto Rico (Acevedo-Vargas, 2002). However, validation studies have not been undertaken with clinical samples, and little is known regarding its psychometric properties when used to evaluate Hispanic individuals with traumatic brain injury (TBI).

Using archival data, this study examined the construct and criterion validity of the Rey Auditory Verbal Learning Test-Spanish (RAVLT-S) when used to evaluate Spanish-speaking adults with TBI. Participants included 106 Spanish-language dominant adults (Mean age = 39.3 years, $SD = 17.9$; 50.0% male) selected from a consecutive series of cases referred to a neuropsychology consultation service at the Neurology Section of the

University of Puerto Rico Medical School. Measures included the RAVLT-S and Spanish adaptations of the Wechsler Adult Intelligence Scale-third edition (WMS-III), Wechsler Memory Scale (WMS), Controlled Oral Word Association Test (COWAT), and Trail Making Test parts A and B (TMT A & B). A split-half procedure was used to examine internal consistency. To examine criterion validity, TBI group performance was compared to the English-language standardization sample (NS; Schmidt, 1996) and to the DEP group. Sensitivity, specificity, positive predictive power and negative predictive power were calculated. Confirmatory factor analyses were conducted to evaluate the underlying structure of the RAVLT-S. Construct validity was further evaluated by examining correlations between RAVLT-S scores and age and education, while convergent and discriminant validity were examined by correlations with the other tests of cognitive abilities. It was hypothesized that (a) the RAVLT-S would demonstrate acceptable reliability; (b) mean RAVLT-S scores for the TBI group would be selectively reduced as compared to the standardization sample; (c) sensitivity, specificity, and positive and negative predictive power would exceed chance; (d) the RAVLT-S would be composed of two factors; (e) raw RAVLT-S scores would yield expected patterns of associations with demographic variables; and (6) standardized RAVLT-S scores would be strongly correlated with other measures of verbal learning and memory, less so with verbal measures that lack an explicit memory component, and insignificantly to measures of perceptual and motor abilities.

Split-half correlations yielded excellent reliability ($r = 0.95$). MANCOVA comparing age-corrected z scores for the DEP and TBI groups while covarying education indicated a significant overall effect, $F(8, 96) = 7.01, p < .001, \eta_p^2 = .37$, as well as a

significant effect for education, $F(8,96) = 3.08, p = .004, \eta_p^2 = .20$, and diagnosis, $F(8,96) = 2.22, p = .032, \eta_p^2 = .16$. Repeated measures MANCOVA indicated a significant effect for trial, $F(7, 721) = 3.03, p = .004, \eta_p^2 = .029$, a significant effect for group, $F(1, 103) = 4.10, p = .046, \eta_p^2 = .038$, and a significant trial by group interaction effect, $F(7, 721) = 3.61, p = .001, \eta_p^2 = .034$. Single sample t tests indicated that the TBI group performed significantly worse ($p < .001$) than the NS group on all RAVLT-S trials. Classification statistics were modest. Confirmatory factor analyses indicated that a two-factor model provided improved fit and parsimony over a one-factor model. Only Trial 1 was significantly correlated with age ($r = -.195, p < .05$), while all RAVLT-S trials were significantly positively correlated with education ($p < .01$). RAVLT-S factor scores correlated significantly with nearly all other measures.

In light of the growing need for linguistically diverse neuropsychological measures, these analyses examined the construct and criterion validity of the RAVLT-S when used to evaluate Hispanic adults with TBI. With some exceptions, the RAVLT-S yielded results consistent with our hypotheses, providing initial support for its validity and clinical utility. Limitations of our study include sample size, use of archival data, potential selection bias, and the use of English language norms. A major strength of this study is its empirical approach to evaluating the validity of the RAVLT-S, which was carried out in accordance with numerous recommendations outlined in the *Standards for Educational and Psychological Testing* (AERA et al., 1999). Further analyses with normal and clinical populations of children and adults are needed, as are factor analytic studies with larger sample sizes. The effects of bilingualism on RAVLT-S performance should also be explored.

Acknowledgements

The completion of this thesis is dedicated to my sons Greco and Henri, and to my wife Tracy, for their genuine encouragement, love and good humor, and for the many personal sacrifices they made as I pursued a higher education.

I would also like to thank my mentor and advisor, Daniel N. Allen, for his kindness and sound stewardship throughout my time at UNLV.

Table of Contents

Abstract.....	iii
Acknowledgements.....	vi
Chapter 1.....	1
Chapter 2.....	5
Literature Review.....	5
Traumatic Brain Injury.....	5
Definition.....	5
Incidence.....	6
Neuropathology.....	7
Severity classification.....	9
Neuropsychological Assessment.....	12
Level of performance.....	13
Pattern of performance.....	13
Right-left comparisons.....	13
Pathognomonic signs.....	14
Diagnostic versus clinical utility.....	15
Cultural Factors.....	16
Definition of culture, ethnicity and race.....	16
Cultural competence.....	18
Culture and theories of intelligence.....	19
Level of acculturation.....	25
Psychopathology.....	27
Culturally appropriate assessment.....	29
Rey Auditory Verbal Learning Test.....	33
Rey Auditory Verbal Learning Test-Spanish.....	34
Purpose of Study.....	37
Hypotheses.....	39
Chapter 3.....	40
Method.....	40
Participants.....	40

Measures	41
Rey Auditory Verbal Learning Test-Spanish (RAVLT-S).....	41
Wechsler Adult Intelligence Scale–third edition (WAIS-III).....	42
Wechsler Memory Scale (WMS).....	43
Trail Making Test (TMT).	43
Controlled Oral Word Association Test (COWAT)	44
Procedure	45
Data Analysis.....	45
Data screening.....	45
Internal consistency.	46
Criterion validity: sensitivity and specificity.....	46
Construct validity: factor structure.	47
Construct validity: correlations.....	50
Chapter 4.....	51
Results	51
Preliminary Analyses.....	51
Reliability	54
Hypothesis 1 internal consistency: split-half correlations.	54
Validity	55
Hypothesis 2 criterion validity: group comparisons.	55
Hypothesis 3 criterion validity: classification statistics.....	58
Hypothesis 4 construct validity: confirmatory factor analysis.	59
Hypothesis 5 construct validity: correlations with demographic variables.	60
Hypothesis 6 construct validity: correlations with other tests.	61
Chapter 5.....	62
Discussion	62
Appendix.....	75
References.....	86
Curriculum Vitae	107

Chapter 1

Historically, psychology has been seated in Western and biological perspectives and assumptions that do not fully consider the impact of race and cultural socialization on individual behavior (APA, 2003). In 2003, the APA published its “Guidelines on Multicultural Education, Training, Research, Practice, and Organizational Change for Psychologists” in recognition of changes in the cultural and sociopolitical landscape of the United States. These guidelines are based on emerging data that suggest that persons of diverse ethnic or racial heritage present unique needs and challenges within American mental healthcare settings. Still, there exists a dearth of psychological research conducted with ethnic minority youth (Artiles, Trent, & Kuan, 1997; Huey & Polo, 2008; Mash & Barkley, 2006) and adults (Areán & Gallagher-Thompson, 1996; Arnett, 2008; Sue, 1999; Sue, 2009), and concerns continue to be raised regarding the gate-keeping functions of psychological tests, the limited availability of cross-culturally validated diagnostic assessment measures, and problems in the application of best practices with members of minority groups (APA, 2012; Suzuki & Ponterotto, 2008).

In the field of clinical psychology, as in other areas of the health care industry, demographic shifts are underway that influence the English language proficiency of segments of the U.S. patient populace. This has been observed within the growing Hispanic population, which to varying degrees retains the use of its native Spanish language. Spanish is the de facto language in many nations that provide migration opportunities to the United States – from Europe (e.g., Spain) to the Caribbean (e.g., Dominican Republic) and from North America (e.g., Mexico) to South America (e.g., Argentina). The Hispanic populace is therefore ethnically, culturally, and linguistically

diverse, and the Spanish language is widely and increasingly spoken throughout the U.S. In recognition of the solidarity that shared ancestry, cultural practices and linguistic roots confer on Hispanic identity, the U.S. Department of Health and Human Services noted over a decade ago that “the Spanish language and culture forge common bonds for many Hispanic Americans, regardless of whether they trace their ancestry to Africa, Asia, Europe or the Americas” (DHHS, 2001, p. 129). For clinicians, this is an important consideration that affects social and clinical transactions with Hispanic clients, including aspects of neuropsychological assessment related to validity, such as test selection, administration and interpretation.

Currently, Hispanics represent the largest and fastest-growing minority group in the United States. According to the Census Bureau, there were over 35 million people of Hispanic origin living in the U.S. as of the year 2000, of which approximately 10.5 million were 14 years of age or younger. Eight years later, the Hispanic population was estimated at over 44 million, representing 14.8% of the total U.S. population (Pew Hispanic Center, 2008). Currently, Census Bureau data indicates that 50.5 million Hispanics reside in the U.S. (representing 16.3% of the population), and that an additional 3.7 million Hispanics reside in Puerto Rico (U.S. Census Bureau, 2011). As of 2010, the percentage of Hispanic-origin people in the U.S. of Mexican descent was 63%, followed by 9.2% Puerto Rican, 3.5% Cuban, 3.3% Salvadoran, 2.8% Dominican, and the remainder from various origins (U.S. Census Bureau, 2011). As mentioned above, many immigrant and native (i.e., U.S.-born) Hispanics retain the use of their native Spanish language. One statistic indicates that approximately 11% of Mexico’s native-born (presumably Spanish-speaking) population is currently living in the United States

(Pew Hispanic Center, 2008). Another confirms that Spanish is the second-most widely spoken language in the U.S., constituting 62.3% of people who speak a language other than English at home (Shin & Kominski, 2010). In the Commonwealth of Puerto Rico, where Spanish and English are both official languages, 95% of the population over 5 years of age speaks Spanish at home, as compared to 4.5% who speak only English (U.S. Census Bureau, 2010).

Given these figures, it seems evident that as the Hispanic population continues to grow, so does the likelihood that psychologists will encounter clients in their clinical practice whose dominant language is Spanish. Despite the significant growth of the Hispanic populace, however, this group has not received adequate attention in the rehabilitation literature (Sharma & Kerl, 2002). Indeed, many Hispanics underutilize mental health services in general, and specialty mental health services in particular (APA, 2012; DHHS, 2001). In the realm of psychological assessment, debates regarding the appropriateness of psychological instruments for use with ethnic and racial minorities bear a long-standing, often contentious history, and continue in earnest. Until fairly recently, however, the well-documented differences in level of performance observed among ethnic minorities on neuropsychological tests were largely ignored, and the field of neuropsychology was relatively immune from criticism on that front (Reynolds, 2000). Yet from a psychometric standpoint, construct irrelevant variance in any form (as may or may not arise from factors related to ethnicity or race), represents a potential threat to the validity of test scores and must be addressed empirically. Such endeavors present both challenges and opportunities for researchers and clinicians working with Hispanic clients, since relatively few Spanish-language neuropsychological measures are currently

available. Combined with the growing patient populace, the limited availability of empirically validated Spanish-language measures presents a considerable diagnostic challenge for clinical psychologists, who are often charged with assessing the cognitive sequelae of traumatic brain injury in Hispanic individuals.

Chapter 2

Literature Review

Traumatic Brain Injury

Definition. Traumatic brain injury (TBI) is a physical insult to the brain caused by external mechanical forces. Such forces may initiate changes in the level of consciousness, and cause temporary or lasting impairment in physical, cognitive, behavioral, or emotional domains. The U.S. Department of Defense defines the event of traumatic brain injury (TBI) as follows:

A traumatically induced structural injury and/or physiological disruption of brain function as a result of an external force that is indicated by new onset or worsening of at least one of the following clinical signs, immediately following the event:

- any period of loss of or a decreased level of consciousness;
- any loss of memory for events immediately before or after the injury;
- any alteration in mental state at the time of the injury (confusion, disorientation, slowed thinking, etc.);
- neurological deficits (weakness, loss of balance, change in vision, praxis, paresis/plegia, sensory loss, aphasia, etc.) that may or may not be transient; or
- intracranial lesion (Casscells, 2007, p. 1).

Traumatic brain injury can thus be conceptualized as any clinically significant neurological impairment that occurs as a consequence of external forces acting suddenly on the brain. These injuries can be the result of any number of extrinsic agents, including

blows to the head, the head striking an object, acceleration and deceleration forces (such as may occur in falls, sports injuries and vehicular collisions), and can occur in the presence or absence of structural damage to the skull itself. In addition to the cognitive and motor impairments listed above, moderate to severe TBI often results in significant changes in behavioral and emotional functioning, such as impulsivity, agitation, depression, anxiety and aggression (Casscells, 2007; Rosenthal & Ricker, 2000).

Incidence. Traumatic brain injury (TBI) is among the most common causes of brain dysfunction in children and adults. In the U.S. alone, an estimated 1.7 million people sustain TBI annually. Of these, approximately 52,000 die, 275,000 are hospitalized, and 1.3 million are treated and released from emergency room settings (Faul, Xu, Wald, & Coronado, 2010). For the period from 1997-2007, firearm (34.8%), motor-vehicle (31.4%), and fall-related (16.7%) TBIs were the leading causes of TBI-related death, and the rate of TBI deaths was three times higher among males than females (CDC, 2011). Although during the same period Hispanics had the lowest overall national rates of TBI deaths for both males and females (CDC, 2011), in some areas Hispanics represented the ethnic group most prone to TBI deaths (Washington State Department of Health, 2009), perhaps as a result of employment in high-risk agriculture and construction industries. Other reports indicate higher than average incidence rates of TBI among Hispanics (e.g. 262 per 100,000 versus a national average of 200 per 100,000) and note that risk rates among Hispanics may be influenced by ethnic minority status, culture-specific health behaviors, and dangerous occupational and living environments (Arango-Lasprilla et al., 2007).

Importantly, an estimated yearly 124, 600 people who sustain TBI experience long-term impairment or disability from their injury, exacting high emotional and financial costs for the individual, their families and society (Faul et al., 2010). Research indicates that Hispanics are disproportionately at risk for poorer long-term functional outcomes following TBI (Gary, Arango-Lasprilla, & Stevens, 2009). At 1-year post-injury, for example, Hispanics showed lower physical functioning, cognitive functioning, and community integration when compared to European American TBI survivors, despite similarities in functional status at admission and discharge and after controlling for age, length of posttraumatic amnesia, injury severity, disability score at admission, functional assessment score at admission, and pre-injury educational level (Arango-Lasprilla et al., 2007).

Neuropathology. The neuropathology of traumatic brain injury is well understood. Modern postulations regarding the physics of brain injury, including the vulnerability of the brain to shearing versus compression strains, were proposed by Holbourn in the mid-1940's (1943, 1945; as cited in Adams, Graham, Murray, & Scott, 1982). Since at least 1982, the mechanisms of moderate to severe injury have been known experimentally (Ricker, 2004). Early comparative research demonstrated that diffuse axonal shearing in primates produced by coronal head acceleration was identical to that observed postmortem in severely head-injured humans (Gennarelli, Thibault, & Adams, 1982), and vice versa (Adams, Graham, Murray, & Scott, 1982). Such early investigations helped to parse out the effects and extent of initial trauma from sequelae secondary to impact, and to clarify the susceptibility of specific brain regions to different types of traumatic forces.

Injury occurs when contact or inertial forces strain the brain tissue beyond its structural tolerance, causing compressive, tensile, or shear deformations (Segun, 2011). Injuries are classified as primary or secondary in nature. Primary injuries occur at the moment of impact, and manifest as focal and/or diffuse trauma. Focal brain injury occurs as a direct result of the brain colliding with the rough interior surface of the cranium, causing lacerations or contusions. The most common sites of focal brain injury are the orbitofrontal region of the frontal lobes and the anterior two-thirds of the temporal lobes, where the greatest bone-brain interfaces occur (Horton & Wedding, 2008). Centers within these areas of the brain (e.g. prefrontal cortex, hippocampal complex) support critical executive, learning, and memory functions which are commonly impaired following TBI. Focal injuries include but are not limited to skull fractures, lacerations, coup and contracoup contusions, intracranial hemorrhage, and neurosensory disturbances such as hearing loss or vestibular dysfunction.

Secondary injuries refer to cellular degradation that occurs after the initial impact. Secondary injuries are mediated by neurochemical mechanisms that regulate swelling, vacuolization, and neuronal death (Segun, 2011), primarily through a cascade of events initiated by the excitatory neurotransmitter glutamate. Cellular damage secondary to the initial impact may or may not be readily apparent, as it can manifest from within hours to after several days. Epidural, subdural, or intracerebral hematomas, for example, may raise intracerebral pressure (Horton & Wedding, 2008), which in turn may lead to cerebral hypoxia, cerebral ischemia, cerebral edema, hydrocephalus, and brain herniation (Segun, 2011). Hypoxia, hypotension, poor cerebral perfusion, and intracranial pressure can represent serious complications. Brown and colleagues (2008), for example, indicate

that even a single episode of hypotension (systolic blood pressure < 90mmHg) or hypoxia (arterial oxygen < 60mmHg) is associated with increased morbidity and mortality. Open-head injuries are also susceptible to infection, which represents a serious threat to cell survival as inflammatory processes are initiated by the immune system. As compared to primary injuries, secondary injuries are characterized by more generalized damage and diffuse cognitive deficits. Careful monitoring of the traumatically brain injured person can help minimize or even prevent secondary injuries.

Since head injury may impact any area of the brain, the domains of cognitive impairment resulting from TBI are many. As previously noted, executive functioning and memory abilities are especially susceptible to structural damage. Tissue damage due to compression, tension and shearing may also disrupt communication circuitry between distal areas of the brain, producing secondary deficits in cognitive domains driven by areas of the brain not directly compromised by trauma. These domains include visuospatial abilities, language, attention, speed of processing, and general intelligence.

Severity classification. By convention, TBI is classified as mild, moderate, or severe on the basis of physical findings, level of consciousness, and post-injury cognitive functioning. Mild TBI is often referred to as a concussion, and is characterized by relatively brief (e.g. 30 minutes) alterations of consciousness such as confusion and disorientation. Other symptoms may be present, such as headaches, memory disturbances and concentration/attentional difficulties. Nearly 90% of all TBIs incurred in the United States are mild in severity (A. W. Brown, Elovic, Kothari, Flanagan, & Kwasnica, 2008).

At the other extreme, severe TBI is characterized by death, severe physical trauma, or enduring functional impairments in physical (e.g. paresis, plegia, seizures, vestibular problems), cognitive (e.g. impairments in memory, executive functioning, language, processing speed), emotional (e.g. depression, anxiety, anger, lability), or behavioral (e.g. agitation, impulsivity, aggression, disinhibition, distractibility) domains (A. W. Brown et al., 2008; Casscells, 2007; Curtiss, Vanderploeg, Spencer, & Salazar, 2001; Donders & Nesbit-Greene, 2004; Faul et al., 2010; Greenwald, Burnett, & Miller, 2003; Horton & Wedding, 2008; Jacobs & Donders, 2008; Lezak, 2004; Reitan & Wolfson, 1986; Rosenthal & Ricker, 2000; Segun, 2011; Uomoto, 2004; Vakil, 2005). Moderate TBI is characterized by impairments between the mild and severe classification ranges.

Gross evaluation of the calvarium is the first step in the classification process, and determines whether an open- or closed-head injury was sustained. This determination is based on the structural integrity of the skull and on the potential for exposure to external elements. Generally speaking, open-head injuries are more severe than closed-head injuries, because in addition to tissue damage incurred as a result of acceleration, deceleration, and rotational forces acting within the skull, the brain is susceptible to meningeal and intracerebral lacerations, contusions, hemorrhages, or hematomas caused by bone fragmentation. In addition, open-head injuries are more susceptible to infection and thus cell death associated with an immunological response. On the other hand, as compared to closed-head injuries, open head injuries may incidentally reduce the potential for compressive tissue deformation secondary to inflammation by providing room for the swollen brain to expand.

In addition to open- versus closed-head designations, duration of loss of consciousness (LOC) is a simple and very commonly used method to classify TBI severity. Within this system, a change in mental status or LOC that lasts less than 30 minutes is classified as mild; mental status changes or LOC that last from 30 minutes to within 6 hours are considered moderate; and mental status changes or LOC lasting 6 hours or more are considered severe. Diffuse axonal injury is the predominant cause of LOC (Greenwald et al., 2003). In general, longer durations of unconsciousness are associated with more extensive brain injuries and more enduring impairments (Horton & Wedding, 2008). Post-traumatic amnesia (PTA) is another common measure of TBI severity. PTA refers to the period from initial time of injury to the point at which ongoing memory for events is stabilized (Horton & Wedding, 2008). Under current conceptualizations, a PTA of 0 to 1 day is classified as mild, a PTA of more than 1 day and less than 7 days is classified as moderate, and a PTA of 7 or more days is classified as severe (Casscells, 2007). In general, longer PTAs are associated with poorer long-term outcomes, but the relationship between PTA and functional outcome is neither absolute nor linear.

Finally, the Glasgow Coma Scale (GCS; Teasdale & Jennett, 1974) is widely used to classify TBI, especially in emergency medical settings. In this system, a numerical score between 3 and 15 is computed by adding the scores in three domains of neurological functioning: motor response (1 to 6 points), verbal response (1 to 5 points), and eye opening (1 to 4 points), with higher scores representing higher levels of functioning. Scores of 13 to 15 constitute mild TBI, 9 to 12 moderate TBI, and 3 to 8

severe TBI. Under this classification system, 90% of patients with a score of 8 or less are comatose (Horton & Wedding, 2008).

Neuropsychological Assessment

Because of the neuropathology resulting from TBI, deficits in neuropsychological functions are common, particularly when severity of TBI is moderate to severe (Lezak, 2004; Reitan & Wolfson, 1986; Uomoto, 2004). Memory impairment, in particular, is among the most commonly observed neuropsychological deficits following traumatic brain injury (Curtiss et al., 2001; Jacobs & Donders, 2008; Vakil, 2005). Proper diagnosis, prognostication, and treatment of TBI therefore rely on valid measures of the psychological processes that underlie memory and other neuropsychological abilities. Accordingly, comprehensive neuropsychological assessments are generally accomplished through the use of standardized neuropsychological tests and batteries, which assess the constructs that underlie cognition (such as memory, attention, processing speed, and executive functioning). This approach allows for multiple levels of inference and therefore provides some control over false-positive and false-negative errors in diagnosis (Horton, 2008). Classic and widely used examples of such batteries include the Halstead-Reitan Neuropsychological Test Battery (Reitan & Wolfson, 1985) and the Luria-Nebraska Neuropsychological Battery (Golden, Hammeke, & Purisch, 1980).

Interpretation of these standardized batteries proceeds using multiple levels of inference, and vary to some extent based on the tests administered and the approach of the examiner. However, in all cases, a key element of test interpretation is to initially establish the level of cognitive abilities prior to injury, which serves as a reference point in interpreting test performance. In this sense, average performance on a particular test

is, in and of itself, meaningless with regard to the possible effects of brain injury on cognitive ability. As an example, an average score for someone who prior to injury was functioning in the superior range suggests the presence of decline in cognitive function.

Level of performance. Within this general framework, other approaches to data interpretation are commonly used (Reitan & Wolfson, 1993). Level of performance, although not diagnostic in and of itself (for the aforementioned reason), is certainly considered in test interpretation. In this method, comparisons are made between the test performance of individuals with suspected brain injury or dysfunction to that of individuals without brain dysfunction. Such norm-referenced approaches (i.e. the comparison of a group or individual against a standardization sample) allow evaluative conclusions to be drawn on the basis of empirical data, since many psychological tests have no inherent or predetermined standards of passing or failing – that is, raw test scores are meaningless in the absence of some type of normative data (Anastasi & Urbina, 1997).

Pattern of performance. A second approach to test interpretation is examining test scores for unusual patterns of performance. Large discrepancies between performance on individual tests, such as large differences between verbal and performance IQ, may suggest the presence of brain injury, because it is assumed that most abilities develop at comparable levels within the individual. The meaningfulness of these discrepancies is often determined based on the magnitude of the discrepancies, such that larger discrepancies are more indicative of brain damage than small discrepancies.

Right-left comparisons. A third approach to test interpretation is examination of performances on each side of the body. This method is limited to tests that are

administered to both sides of the body, such as motor tests that are performed separately with each hand. However, it is a very useful method of test score interpretation because it allows for inferences to be made regarding lateralization, and to some extent, localization of cerebral injury. The rationale underlying this approach is based on known association between the right and left side of the body. For example, the dominant side (the right side in most individuals) is usually somewhat stronger and faster than the non-dominant side of the body, so test subjects should be somewhat faster on motor tests with their dominant hand compared to their non-dominant hand. For basic sensory abilities, such as vision, audition, and tactile sensation, equal sensitivity is expected for both sides of the body. Thus, individuals who are right hand dominant, but exhibit poorer tactile sensation in their right hand compared to their left, and also have slower motor performance in their right hand compared to their left, would have likely sustained damage to their left hemisphere.

Pathognomonic signs. A fourth approach to test interpretation is to examine test scores for pathognomonic signs. These signs represent errors on tasks that are rarely made by individuals who do not have brain damage, and so are, in and of themselves, strong indicators of brain damage. An example of a pathognomonic sign would be a response on a block design task that fails to maintain the gestalt of the design. Another would be an inability to accurately reproduce simple geometric designs, such as a square or a cross, with a pencil and paper. Physical anomalies can also be pathognomonic. In TBI, for example, a fracture of the skull is a pathognomonic sign of head injury, though its clinical importance can be overemphasized since damage sustained by the brain (rather than the skull) ultimately determines the clinical outcome (Adams et al., 1982).

Visual field defects are also strong indicators of brain damage. While these signs, independent of corroborating data, do not provide direct information about overall impairment in particular cognitive domains such as memory, they are strong indicators of brain damage and so are commonly used in interpretation of neuropsychological test performance.

Diagnostic versus clinical utility. It is important to note that neuropsychological norms do not represent “ideal” performance (any more than a norm on an emotional adjustment inventory corresponds to an absence of unfavorable or maladaptive behaviors), rather average performance and relative frequency of deviation above and below the average (Anastasi & Urbina, 1997). Norm-referenced tests are thus constructed such that the standardization sample resembles the population on as many meaningful variables as possible, in order to reduce the potential unwanted effects of confounding variables. Such measures are often improved by stratifying the normative comparison group on the basis of demographic variables such as gender and socioeconomic status. Age and education, in particular, have been clearly demonstrated to influence the results of neuropsychological tests (Del Ser Quijano et al., 2004; Fernández & Marcopulos, 2008; Gómez-Pérez & Ostrosky-Solís, 2006; Knight, McMahon, Green, & Skeaff, 2006; Leòn-Carriòn, 1989; Pena-Casanova et al., 2009; Perianez et al., 2007; Reitan & Wolfson, 1995; Schmidt, 1996; Schoenberg et al., 2006).

The diagnostic validity of a given instrument is determined by the degree to which it accurately discriminates between impaired and non-impaired individuals, in terms of the distribution of scores. The clinical utility of a test, which is a function of its false negative and false positive rate (Schoenberg, 2006), relates to the diagnostic

meaning that a particular score has for a particular person (Smith, Ivnik, & Lucas, 2008). Within this framework, the probability that a test will correctly identify a neurologically impaired individual is referred to as sensitivity (SENS), and is defined by the ratio of true positive to true positives plus false negatives. The probability that a test will correctly identify a non-neurologically impaired individual as non-impaired (i.e., indicate a negative result) is referred to as specificity (SPEC), and is defined by the ratio of true negatives to true negatives plus false positives. Given that the neurological status of patients is often not initially known, however, it can be argued (e.g., Schoenberg 2006) that a more useful index of a test's diagnostic utility is its positive predictive power (PPP), which refers to the likelihood that an individual with a positive test result actually has neurological impairment. PPP is defined by the ratio of true positives to true positives plus false positives. Inversely, negative predictive power (NPV) refers to the likelihood that a negative test result correctly identifies a non-impaired person. NPV is defined by the ratio of true negatives to false negatives plus true negatives. Thus, positive and negative predictive values answer the most critical question facing neuropsychologists once the diagnostic validity of a measure has been established, which is whether a patient does or does not have the condition of interest (Smith et al., 2008).

Cultural Factors

Definition of culture, ethnicity and race. There are no universally agreed-upon definitions of culture, ethnicity or race. Considerable debate exists regarding the extent of overlap between these constructs, and whether clearly delineated boundaries between them are necessary, valid, or ethical. Against this backdrop, and in recognition of the need for a common nomenclature to facilitate communication, the following abbreviated

definitions from the APA's "Guidelines on Multicultural Education, Training, Research, Practice, and Organizational Change for Psychologists" are used:

Culture is defined as the belief systems and value orientations that influence customs, norms, practices, and social institutions, including psychological processes (language, caretaking practices, media, educational systems) and organizations (media, educational systems) . . . all individuals are cultural beings and have a cultural, ethnic, and racial heritage . . . informed by the historical, economic, ecological, and political forces on a group . . . culture is fluid and dynamic and that there are both cultural universal phenomena and culturally specific or relative constructs.

Ethnicity is defined as the acceptance of the group mores and practices of one's culture of origin and the concomitant sense of belonging . . . individuals may have multiple ethnic identities that operate with different salience at different times.

Race is considered to be socially constructed rather than biologically determined. Race . . . is the category to which others assign individuals on the basis of physical characteristics, such as skin color or hair type, and the generalizations and stereotypes made as a result. (APA, 2003, p. 380)

Within this framework, the composition of traditionally defined groups, including Hispanics and European Americans, is by definition culturally, ethnically, and racially diverse. Variations in linguistic expression are a natural component of such diversity. In the U.S. (unlike in countries where Hispanics represent a majority and Spanish is either the de jure or de facto national language), Hispanics represent an ethnic minority group.

Unlike other ethnic minorities, however, such as Asian- and African Americans, Hispanics have not traditionally been defined as a race (Puente & Ardila, 2000).

Cultural competence. Cultural competence refers to the aptitude and duty of a service provider (e.g. psychologist, physician, social worker, educator, etc.) to deliver appropriate, individually-tailored service to a member of a cultural group that differs from their own, in a manner that acknowledges, respects, and incorporates said group's unique cultural characteristics and values. The APA Ethics Code Standard 2.01b, "Boundaries of Competence", states that psychologists must obtain the training, experience, consultation or supervision necessary to ensure the competence of their services when scientific knowledge establishes that an understanding of cultural factors (e.g. language, gender identity, race, ethnicity, etc.) is essential for effective implementation of services (APA, 2010). Within this paradigm, psychologists are charged with assuming a *culture-centered* focus by recognizing the influence of different historical, ecological, sociopolitical and disciplinary contexts (APA, 2003). Maintaining cultural competence is therefore an ethical obligation that requires ongoing education, training, and supervised experience, as well as empathy and self-awareness. Currently, the preferred method of ensuring a culturally and clinically competent evaluation is for the clinician to possess the competencies required to provide such services, including speaking the client's primary language, or to refer the client to a provider who does (Judd et al., 2009). Data derived from a national survey of neuropsychologists in the U.S., however, indicates that 82% of respondents rated their level of competence to work with Hispanic/Spanish-speaking populations as inadequate (Echemendia, Harris, Congett, Diaz, & Puente, 1997; Echemendia & Harris, 2004; Judd et al., 2009); that published

Spanish-language adaptations of English measures are rarely used; and that neuropsychological test performances of monolingual Spanish speakers are often evaluated on the basis of clinical judgment alone, rather than in relation to normative data (Echemendia & Harris, 2004).

Culture and theories of intelligence. Intelligence is a culturally-defined construct. In the U.S., “the conventional notion of intelligence is built around a loosely consensual definition of intelligence in terms of generalized adaptation to the environment” (Sternberg, 2003, p. 139). David Wechsler defined intelligence as the “aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment” (Wechsler, 1944, p. 3). Edwin Boring famously declared that intelligence is what intelligence tests test (Boring, 1923; as cited in Neisser, 1979). Others have argued that intelligence itself does not exist, except as it relates to or resembles a prototypically intelligent person, and suggest that an adequate verbal definition of intelligence is thus impossible (Neisser, 1979). By relying on observable behavior and acknowledging the gains or limits imposed by environmental factors, these psychologists and others seem to imply that intelligence varies not only as a function of the particular cognitive properties of the individual, but as a function of temporal, evaluative judgments made external to the individual (i.e., those assigned to a given behavior, at a particular point in time, by a given community or audience). Thus, while the natural distribution of biological mental processes underlying cognition (e.g., memory, processing speed, visuospatial acuity) and the degree of vertical integration among them (e.g., reasoning ability, creativity) are of primary importance, the evaluative component of intelligence as a function of culture cannot be ignored. It can thus be

argued that the degree to which intelligence can be defined universally is in part constrained to areas of overlapping behaviors among cultures that correspond to valued forms of self-representation.

Given its multidimensionality, Neisser (1979) concluded that two possibilities arise with respect to measuring intelligence: not to measure it at all, or to measure it inadequately. In educational and psychological settings, intelligence is routinely defined psychometrically and expressed in terms of distributions of scores derived from standardized, norm-referenced tests. These operationalizations derive from the work of Alfred Binet and Theodore Simon (Binet & Simon, 1905; Binet & Simon, 1916), and later, Lewis Terman (Terman, 1916), who originally developed measures to identify educable children in school settings. Scores on such measures (e.g., IQ) are heavily influenced by skills and abilities associated with academic achievement (Anastasi & Urbina, 1997), and have become synonymous with intelligence in popular culture.

Other conceptualizations of intelligence trace their origins to the theoretical and psychometric framework provided by Charles Spearman, who postulated that a general factor, *g*, underlies human intelligence (Spearman, 1904). In the 1940s, Cattell refined Spearman's theory by presenting a model of intelligence that distinguished "fluid" or novel problem-solving abilities from "crystallized" or knowledge-based abilities (Cattell, 1943). Carroll, in turn, elaborated on Cattell's theory in a major factor analytic study of hundreds of cognitive ability test scores, whereupon he concluded that three strata (consisting of "general", "broad", and "narrow" abilities) underlie human intelligence (Carroll, 1993; Carroll, 1997). Contemporary psychometric models of intelligence are exemplified by what is now known as Cattell-Horn-Carroll (CHC) theory (see Flanagan,

Ortiz, & Alfonso, 2007), which synthesizes the aforementioned work with that of John Horn (e.g., Horn & Cattell, 1966; Horn, 1989) and others, expanding the hierarchical framework through increasingly sophisticated psychometric analyses and procedures. In its current form, CHC theory postulates 10 broad and over 70 narrow abilities, or facets of intelligence.

Aspects of these abilities are captured to varying degrees by numerous measures of intellectual functioning, including the Wechsler, Stanford-Binet, Woodcock-Johnson, and Kaufmann series of assessment tools (see Flanagan et al., 2007), and are typically expressed as standardized scores. David Wechsler is credited with emphasizing the clinical utility of intellectual assessment with patient populations as early as the 1930s. Today, the Wechsler series are the most widely used individually-administered measures of intelligence for children, adolescents, and adults (Lichtenberger & Kaufman, 2009). These are routinely incorporated into neuropsychological batteries used to evaluate both English and Spanish speakers. In fact, the Wechsler Adult Intelligence Scale is reported to be the single most frequently used measure by neuropsychologists when evaluating both bi- and monolingual Hispanics (Echemendia & Harris, 2004). Similarly, for both groups of Hispanics, the Wechsler Memory Scale is among the top five instruments used (Echemendia & Harris, 2004).

Other, more pluralistic theories of intelligence have also been proposed. Gardner, for example, has put forth a theory of Multiple Intelligences (MI), which argues against the all encompassing view of intellect assumed by *g*, and against the proposition that conventional psychometric instruments adequately assess human intelligence (Gardner, 1983; Gardner, 1993; Gardner, 1999). In his theory, Gardner proposes the existence of 8

to 9 relatively autonomous intelligences (e.g., kinesthetic, musical, linguistic, and spatial), each with distinct characteristic processes and comprised of constituent subunits of that particular intelligence. These intelligences are posited to develop as a byproduct of genetic potential, personal motivation, and quality of instruction. Gardner emphasizes the role of culture in the expression of particular intelligences, noting that different societies value different intelligences, and in this way selectively encourage or deter their development.

Sternberg also offers an alternative to psychometric models of intelligence. Sternberg's theory of Successful Intelligence (SI), or triarchic theory of intelligence (Sternberg, 1985; Sternberg, 2003), suggests that conventional, academically-laden definitions of intelligence are too narrow. In this theory, SI is comprised of three sub-theories of intelligence (componential, experiential, and contextual), each reflecting different aspects of intelligent behavior (information processing, creativity/novel reasoning, and practical problem-solving in everyday life). Within this framework, Sternberg argues that the definition of intelligence is idiographic and contingent on an individual's personal goals and standards of success. The behaviors that constitute successful intelligence operate within and are inseparable from an individual's specific sociocultural context. Thus, whereas the processes that underlie intelligence may not change across contexts, the determination as to whether a particular behavior is or is not intelligent may. Finally, according to Sternberg, SI is accomplished by capitalizing on strengths and compensating for weaknesses; by adapting the self to environmental changes; by shaping or modifying the environment itself; and/or by selecting different environments that increase the chances for personal success.

Still others stress the cross-cultural variation of intelligence, noting that definitions of intelligence are eco-culturally constrained (Greenfield, 1997; Serpell, 2000). Like Anastasi (1997), Serpell (2000) notes that in contemporary, industrialized societies, intelligence is strongly associated with individual excellence on tasks emphasized by academic curricula. In communities without schools, he argues, those indicators have no indigenous meaning. He also contends that because the prevailing psychometric practices in the U.S. reflect a Western view of intelligence (i.e., emphasizing decontextualization, quantification, and biologization), the definition of intelligence is restricted in ways that limit its applicability to other social groups and environments. Greenfield (1997) similarly views testing instruments not as universal metrics, but as specific cultural genres that presuppose frameworks of shared values, knowledge, and forms of communication. In the absence of convergent views of intelligence, she argues, one culture's criterion for intelligent behavior can be another's criterion for foolishness (Glick, 1968; as cited in Greenfield, 1997).

Clearly, the abbreviated list of theories and conceptualizations of intelligence presented here are neither exhaustive nor mutually exclusive. Gardner's Multiple Intelligences theory, for example, does not contest the existence of *g*, rather its "province and exploratory power" (Gardner, 1999, p. 87). Greenfield (1997) acknowledges that where cultural definitions of intelligence are equivalent and culture-specific content is removed, a translation of a test of cognitive ability may be "perfectly valid" (p. 1117). Wechsler acknowledged the role of non-intellective factors, reportedly expressed frustration that factor analyses rarely accounted for more than 60% of the total variance of his tests, and assumed that residual variance was accounted for by aspects of

personality such as persistence and anxiety that facilitated or inhibited intelligent behavior (Kaufman & Kaufman, 2001). Conceptualizations of intelligence that stem from cross-disciplinary perspectives (e.g., philosophical, anthropological, evolutionary) are also unaccounted for here, as are the merits or limitations of common heuristics and implicit judgments about intelligence that laypeople make every day. Finally, historical controversies regarding the definition, scope, measurement, and implications of intelligence (and intelligence testing) remain salient, and the intensity with which the subject of human intelligence is debated endures (e.g., R. T. Brown, Reynolds, & Whitaker, 1999; Helms, 1992; Herrnstein & Murray, 1994; Jensen, 1980; Neisser et al., 1996; Sanchez, 1932; Sternberg, Grigorenko, & Kidd, 2005; Williams, 1975).

Given the multiplicity of domains and definitions of intelligent behavior, it is not surprising that a universally accepted cross-cultural metric of intelligence is unavailable. This can present unique challenges where cultural conceptions of intelligence differ between client and practitioner. Differences in cognitive orientations, for example, can negatively influence IQ scores, as has been observed in cultures that use functional, rather than taxonomical classification systems to organize test responses (Suzuki & Ponterotto, 2008). Differences in relational styles can also affect assessment outcomes. Sharma and Kerl (2005), for example, note that Mexican American culture is more relational than the broader American culture in which it is embedded. Within this interdependent relational context, the family unit – which transcends the immediate and extended family unit to include friends – is often valued over its individual members. Whether such interdependence is considered adaptive, or even permissible (e.g., in the classroom, or during group administered tests) however, depends on contextual factors.

Greenfield (1997), for example, referred to differences in test-taking behavior between individuals from individualistic versus collectivistic cultures as evidence of cultural variation in epistemology. This point was illustrated in a study with Mayan children, who were perplexed when forbidden to collaborate with parents or others on test questions (Sternberg & Grigorenko, 2008). As the authors noted, whereas such behavior would be considered adaptive (i.e., intelligent) in Mayan culture, such collaboration would more likely be viewed as cheating in the United States. Varying degrees of acculturation, acceptance of and orientation to the host culture, exposure to educational and occupational opportunities, and English language proficiency also influence interactions between the client and practitioner during the assessment of intelligent behavior. Ultimately, measures of intelligence necessarily reflect responses, abilities, characteristics, cognitive orientations, and approaches to test-taking that are valued within the test developer's culture. These factors represent important psychometric considerations for clinicians.

Level of acculturation. Underlying many of the cultural idiosyncrasies that arise when working with Hispanic clients is the level of acculturation they have attained. Broadly speaking, acculturation refers to the push/pull phenomena of assimilation, separation, marginalization, and integration that occur while adapting to a host culture (Suzuki & Ponterotto, 2008). One complicating factor is that while a Hispanic person's external behavior may reflect a high degree of behavioral adaptation, it cannot be assumed that s/he has internalized the values of the mainstream culture itself (Suzuki & Ponterotto, 2008). As an example, Mexican American children whose migration history is generations old may appear to identify exclusively with European American culture.

However, even Mexican Americans who have lived in the U.S. for generations often have values, language usage and behaviors that differ from the dominant culture (Sharma & Kerl, 2002). Intergenerational differences in rates of acculturation may also alter the normal development of child/parent relationships, such that over time immigrant parents understand less of their children's experiences outside of the home environment, and children rely less on the assistance of their parents to navigate the demands of the host culture (APA, 2012).

A related dilemma may arise when children develop English-speaking skills more quickly than their parents, and are asked to serve as interlocutors between their parents and other adults. While bilingual service providers or the use of professional interpreters or translators is ideal (Judd et al., 2009), this is often impracticable, if not impossible. Similarly, language barriers can complicate access to and delivery of mental health services in many settings, but may be especially problematic in rural areas, and for clients in need of highly specialized practitioners (e.g. psychologists specializing in OCD who are willing to conduct in vivo exposure and response prevention outside of the office setting, in a language other than English). Against this backdrop, Hispanic children are often expected to interpret for their parents. While this is not necessarily problematic in every setting (e.g. the grocery store), the nature of the communications between psychologists and parents preclude the use of a child interpreter. It would be ethically untenable, for example, to use a child to facilitate communication with a purported or potential abuser. In general, therefore, psychologists and other health professionals are encouraged to refrain from using children as interpreters. Some even argue that such a practice constitutes a violation of the child's civil rights (Suleiman, 2003). Others simply

note that the potential for errors is compounded when family members are used as translators (Biever et al., 2002). For psychologists, such factors call for a heightened degree of flexibility, inquisitiveness, and alertness when assessing and developing interventions with Hispanic clients.

Psychopathology. Special attention must be paid to cultural factors that can influence the diagnosis and treatment of psychopathology. Research indicates, for example, that Hispanics tend to express depressive symptomatology differently than their European American counterparts, with depression being more typically associated with somatic rather than cognitive symptoms (Blaney & Millon, 2009). Somatic explanations for psychological disturbances may therefore be more prominent in Hispanic communities, where such explanations are sanctioned. Studies have also demonstrated that 1) Hispanic adolescents report significantly more depressive symptoms than European American adolescents; 2) Hispanic adolescents report greater depressed mood than African American, Asian American, and European American adolescents, independent of SES; and 3) Mexican American youths manifest higher rates of major depressive disorder than 8 other ethnic minority groups, even after adjusting for age, gender and SES (Mash & Barkley, 2007). These factors may present added difficulties when seeking to isolate TBI-specific impairment, which is characterized by heterogeneous symptomatology spanning cognitive, motor, behavioral and emotional domains. Similarly, Hispanics tend to underutilize mental health resources (APA, 2012; DHHS, 2001), potentially complicating TBI outcomes by protracting the amount of time between injury and assessment.

Psychologists must also be sensitive to ways in which low levels of acculturation may influence presenting symptomatology. Many immigrants leave their home country as a result of economic or political pressures, and must immediately and simultaneously navigate and adapt to the U.S. culture, often with limited financial resources and psychosocial support. Even highly educated and skilled adults may experience a loss of status when they immigrate, often finding dramatically fewer employment opportunities leading to unemployment, underemployment or downward mobility – problems compounded by ethnic or racial minority status (APA, 2012). For poor, newly arrived immigrants, the migratory experience itself can represent a major life crisis: The language barrier, unfamiliar culture, possible climatic changes, and losses of routine and social supports can create very significant stress and/or exacerbate existing psychopathology (Hancock, 2005). In the case of Mexican immigrants, for example, the patriarch often ventures to the U.S. alone (disrupting the traditional family environment, sometimes for many months), secures employment, and sends for his family later. If the patriarch and/or his family are illegal immigrants, the situation is considerably more serious, as adults and children live under constant threat of deportation and/or separation from family members. In addition, many U.S.-born and immigrant-origin Hispanic children alike are subject to cyclical upheavals as their parents migrate regionally to maintain agricultural and other seasonal employment. Thus, the assessment of psychopathology in Hispanic immigrants must consider whether behavioral manifestations (e.g., depression, anxiety, post traumatic stress, substance abuse and conduct problems) are related to migratory factors such as the loss or disruption of family relations, friendships, social support, and self-identity; or to associated sociocultural

factors such as discrimination and marginalization. Potentially traumatizing experiences should be carefully examined and assessed within a culture-specific context (Suzuki & Ponterotto, 2008), with careful consideration of legal, economic, and social issues related to ethnicity. Similarly, psychological interventions and recommendations, including those derived through neuropsychological assessment, should continuously take language and cultural factors into account (AERA, APA, & NCME, 1999; APA, 2003; Judd et al., 2009). Finally, many ethnic minorities demonstrate significant resilience and a level of optimism that belies the difficulties they encounter. Psychologists are advised to recognize and capitalize on the advantages that such strengths may confer on Hispanics in clinical, educational, and employment settings (APA, 2012).

Culturally appropriate assessment. Recognizing the cultural specificity of all behavior, no test can be universally applicable, as every test tends to favor persons from the culture in which it was developed (Anastasi & Urbina, 1997). Questions regarding the reliability and validity of psychological tests for use with Hispanics were raised and published in scientific journals over 75 years ago (Sanchez, 1932; as cited in Padilla & Borsato, 2008). Today, high-stakes decisions in educational, occupational, legal, and therapeutic settings are frequently made on the basis of psychological assessments. Results can determine access to services, employment, and competence to stand trial. Research in test bias and fairness in mental testing indicates that middle and upper class children and adults score higher on tests of cognitive ability than those from lower socioeconomic classes; that European Americans tend to score higher than minorities by 1 standard deviation on average; and that systematic differences exist between males and females depending on the measure used (Murphy & Davidshofer, 2005). Regarding

ethnic minorities, the different levels and patterns of performance on cognitive tests are most prominent on measures of IQ (Reynolds, 2000). While there are many potential sources of error unrelated to culture or ethnicity, where cultural bias in psychological testing exists, it contributes to test error variance (Horton, 2008). Such systematic, construct irrelevant variance obscures patients' true scores, complicating and potentially invalidating the diagnostic process.

Even when tests do not systematically discriminate against particular groups, the invalid use of a test can cause harm (Murphy & Davidshofer, 2005). Sample characteristics and instrumentation directly influence the process and outcome of psychological assessments. The use of standardized instruments not intended for individuals with limited English proficiency, for example, may adversely affect Hispanics by functioning as measures of language proficiency rather than valid measures of the construct in question (AERA et al., 1999; Puente & Ardila, 2000; Reynolds, 2000; Suzuki & Ponterotto, 2008). For instance, because language proficiency includes measures of reading ability, and 58% of Hispanic students score below the *National Assessment of Education Progress* basic level of proficiency (Mash & Barkley, 2006), Hispanic children are at greater risk for learning disability misdiagnoses and placement in remedial education courses. Bilingual Hispanic Americans are also susceptible to linguistic interference and may be penalized when English language testing materials and norms are used to the exclusion of Spanish language materials, and vice versa (AERA et al., 1999; Reynolds, 2000).

The *Standards for Educational and Psychological Testing*, (Standards; AERA et al., 1999), a joint collaboration among leading educational and psychological

organizations, also recognizes that test use with individuals with limited proficiency in the language of the test may introduce construct irrelevant components to the assessment process. The *Standards* (1999) thus prescribe that testing practices be designed to reduce threats to the reliability and validity of test score inferences that may arise due to differences in language. It cannot be assumed, for example, that the validity of even a well-validated English language measure is preserved when literally translated, because the translation may fail to account for important cultural and linguistic factors such as functional, metric or construct equivalence (AERA et al., 1999), the influence of acculturation or bilingualism (Ardila, Rosselli, & Puente, 1994; Puente & Ardila, 2000; Suzuki & Ponterotto, 2008), or the cultural salience of the test or test-taking practices. As an alternative, the use of non-verbal, performance, and culture reduced tests (i.e. tests that utilize objects, symbols, or information with which members of various cultures would be expected to be equally familiar), has been proposed, but such approaches have failed to reduce systematic differences in test scores between European American and minority subjects (Murphy & Davidshofer, 2005). Of particular importance to neuropsychological assessment, measures of neurocognitive functioning must be validated for use with specific clinical populations, as even tests with otherwise excellent psychometric properties may be insensitive to brain dysfunction (Bello, Allen, & Mayfield, 2008).

Whereas the *Standards* (1999) acknowledge that language differences are almost always concomitant with cultural differences that must *also* be taken into account (emphasis added), others warn against interpreting such differences as indicators of cultural bias. Brown and colleagues, for example, contend that the available empirical

evidence argues against claims that standardized cognitive tests are culturally biased, and that many criticisms (prevalent in sources ranging from academic journals and psychology textbooks to the popular media) are misinformed and inappropriate (R. T. Brown et al., 1999). Others suggest that Hispanics may indeed suffer bias from neuropsychological testing (Gary et al., 2009), as suggested, for example, by the effects of ethnicity on common neuropsychological measures of verbal and perceptual functioning (Donders & Nesbit-Greene, 2004).

In sum, the potential influence of cultural factors on clinical diagnosis warrants special attention. According to the APA, “for testing and assessment to be culturally appropriate, there needs to be a continuous, intentional, and active preoccupation with the culture of the group or individual being assessed” (APA, 2012, p. 7). Thus, when assessing Spanish-speaking clients, clinicians must consider the potential threat of linguistic and cultural confounds to both the construct and criterion validity of their chosen measures. Many legitimate objections to the use of psychological tests with Hispanics and other ethnic minorities (such as the use of inappropriate content, measurement of different constructs, and linguistic bias) resolve to issues of validity, which can be examined empirically (Reynolds, 2000). Ideally, psychometrically sound, Spanish-language neuropsychological measures should be used by clinicians and psychometrists that possess a command of the Spanish language commensurate with the level of training and education required to conduct psychological assessments. Unfortunately, very few commonly used neuropsychological tests are available in Spanish (Camara, Nathan, & Puente, 2000); limited data are available regarding the validity of those that are (De la Plata et al., 2009; Prifitera & Saklofske, 1998); non-

Spanish speakers routinely perform neuropsychological assessments with Spanish speaking clients (Echemendia & Harris, 2004); and even when Spanish language alternatives are available, unvalidated verbatim translations are used more frequently than culturally adapted translations (Echemendia & Harris, 2004). Such limitations illustrate the need for validating neuropsychological measures such as the RAVLT-S for use with Spanish speaking clinical populations, as examined here.

Rey Auditory Verbal Learning Test

The Rey Auditory Verbal Learning Test (RAVLT; A. Rey, 1958) is among the most commonly used neuropsychological tests of verbal learning and memory. It is a list learning task originally developed in French and then adapted to English (Lezak, 1983; Taylor, 1959). The RAVLT is used to evaluate rates of verbal learning and memory, proactive and retroactive inhibition, retention, recognition ability, encoding, retrieval, and subjective organization. Over the years, test performances of numerous patient samples have been evaluated and compiled (Bohlhalter, Abela, Weniger, & Weder, 2009; Estévez-González, Kulisevsky, Boltes, Otermín, & García-Sánchez, 2003; Jacova et al., 2008; Ryan, Paolo, & Skrade, 1992; Schmidt, 1996; Schoenberg et al., 2006; Steinberg, Bieliauskas, Smith, Ivnik, & Malec, 2005; Vakil, Greenstein, & Blachstein, 2010). In general, the RAVLT has been shown to be sensitive to neurological injury, including TBI, and insensitive to psychiatric illness such as depression and anxiety. However, there is some evidence that psychological distress (including depression, post-traumatic stress, and other anxiety disorders) has some effect on RAVLT performance (Spren & Strauss, 1998). The RAVLT is also influenced by demographic factors, including age, IQ/education, and possibly gender. Evidence indicates, for example, that RAVLT scores

improve as a function of age in children, tend to decrease in adults with advancing age, are positively influenced by higher IQ/education levels, and when gender differences are found, women outperform men on recall but not recognition trials (Spreeen & Strauss, 1998). Schmidt (1996) compiled normative data for the English language RAVLT using healthy individuals, which were then stratified by age.

Rey Auditory Verbal Learning Test-Spanish. In 2002, a Spanish translation of the RAVLT (RAVLT-S) was developed by Acevedo-Vargas. This version was developed and evaluated in Puerto Rico with native Spanish-speaking participants, and is currently in use at the neuropsychology consultation service at the Neurology Section of the University of Puerto Rico Medical School. Previous research supports the validity and clinical utility of translations of the English RAVLT into languages as diverse as Greek, Portuguese, and Persian (Jafari, Moritz, Zandi, Kamrani, & Malyeri, 2010; Malloy-Diniz, Lasmar, Gazinelli, Fuentes, & Salgado, 2007; Messinis, Tsakona, Malefaki, & Papathanasopoulos, 2007). However, it has been reported that variations in translation and administration practices, as well as educational and cultural differences may invalidate comparisons of English-speaking North American samples to norms developed in Europe (Spreeen & Strauss, 1998). It has also been reported that the cross-cultural utility of the RAVLT is weakened by the inclusion of culturally salient items (e.g., turkey, ranger, curtain) with which other cultures may be unfamiliar (Maj, D'Elia, Satz, & Janssen, 1993).

Regarding its use in medical settings specifically (as is currently done in Puerto Rico with the RAVLT-S), research indicates that the English RAVLT is sensitive to diffuse neuropsychological changes in TBI patients, and taps not only specific verbal

learning and memory, but also global cognitive functions (Callahan & Johnstone, 1994). The RAVLT is also routinely applied in practices that evaluate patients with non-trauma related medical conditions. In the early 1990's, for example, the World Health Organization (WHO) launched an international initiative to study the neurological and psychiatric disorders associated with HIV infection (Maj et al., 1993). A central feature of that study was the assessment of cognitive functions. This resulted in the development of cross-cultural adaptations of major neuropsychological instruments, including the RAVLT (renamed WHO/UCLA AVLT), which were identified by the WHO as unsuitable for cross-cultural use in their original form.

As a result, a Spanish language version of the WHO/UCLA AVLT was later developed and included as part of the Neuropsychological Screening Battery for Hispanics, or NeSBHIS (Ponton et al., 1996). Data for this battery were derived from normal, predominantly monolingual Hispanics (75%) residing in the U.S., and approximated the distribution of Hispanics in the U.S. by country of origin. Subsequent research regarding the psychometric properties of the NeSBHIS has provided support for its construct validity with both normal participants (Ponton, Gonzalez, Hernandez, Herrera, & Higareda, 2000) and patients with epilepsy (Bender et al., 2009), although its diagnostic utility in distinguishing lateralized neuropsychological impairment in patients with temporal lobe epilepsy is limited (Barr et al., 2009).

The diverse nature of the Hispanic population, however, raises important questions regarding regional and cultural variations of the Spanish language. These concerns underscore the need for psychologists to exercise best efforts to utilize measures appropriate to specific patient populations given their particular language variant (e.g.,

Judd et al., 2009). For example, while U.S./Mexico borderland Spanish speakers performed similarly when compared to Spaniards on 16 Spanish-language neuropsychological measures, both percent of life span spent in the U.S. and the bilingual status of the borderland group were correlated with performance on some tests (Artiola, Heaton, & Hermosillo, 1998). Specifically, increased percent of life span spent in the U.S. was negatively correlated with performance on a Spanish word-generation task, and positively correlated with performance on the Wisconsin Card Sorting Task. These differences diminished, however, as education levels increased. Also, bilingual borderland subjects performed significantly better than monolingual speakers in a list-learning task, suggesting that multi-language development may confer a cognitive advantage related to verbal learning. In another study comparing performance by patients in the U.S. to patients in Columbia and Spain on Spanish variants of the Boston Naming Test, De la Plata et al. (2009) observed similar education-related phenomena, and suggest that clinicians and rehabilitation professionals consider using normative data from Spain when evaluating highly educated Spanish-speakers on neuropsychological measures. Finally, in the factor analytic study of the NeSBHIS by Ponton et al. (2000), Digit Span loaded on a Language rather than Attention factor, suggesting that the cognitive abilities tapped by neuropsychological instruments may vary as a function of language. As reported by the authors, the common observation that Hispanics recall fewer numbers on average than non-Hispanics may be explained by the hypothesis that Digit Span taxes attentional skill in English speakers and linguistic skills in Spanish speakers.

In contrast, using a Mexican Spanish translation of the RAVLT (Miranda, 1996), Miranda & Valencia (1997) found no significant difference in the performance between English and Spanish speakers, despite significant differences in Spanish word syllabic length. This somewhat counterintuitive finding was explained by the relative spoken word duration speed of many Spanish words, which appeared to counteract the effects of longer word length. This finding – that no significant differences in performance were found between English and Spanish versions of the RAVLT – underlies the rationale for the comparisons made in this study.

Purpose of Study

The purpose of this study was to examine the construct and criterion validity of the RAVLT-S (Acevedo-Vargas, 2002) when used to evaluate Spanish-speaking individuals with TBI. Three analyses were proposed. In the first analysis, to examine criterion validity, TBI group performance was compared to that of the English-language standardization sample and to a sample of individuals with Major Depressive Disorder. As discussed above, memory impairment is one of the most common consequences of TBI. Naturally, this phenomenon also disrupts learning. Thus, mean scores on Learning Trials 1 through 5 and on Immediate Recall, Delayed Recall, and Delayed Recognition trials were compared between groups. Classification statistics including sensitivity, specificity, positive predictive power and negative predictive power were also examined in order to determine the ability of RAVLT-S scores to differentiate TBI from the normative sample.

In the second analysis, the underlying factor structure of the RAVLT-S was examined using confirmatory factor analyses. RAVLT-S Learning Trials 1 through 5

provide information about individuals' ability to learn context-free auditory verbal stimuli over repeated trials, while Immediate Recall, Delayed Recall and Delayed Recognition (Trials 6 through 8) assess the integrity of long-term verbal memory. Prior factor analytic work with the RAVLT yielded mixed results (Baños, Elliott, & Schmitt, 2005; Ryan, Rosenberg, & Mittenberg, 1984; Talley, 1986; Vakil & Blachstein, 1993). However, it was hypothesized that a two factor model would fit the data well (Erickson & Scott, 1977; Vakil & Blachstein, 1993) reflecting *acquisition* and *retention* (Vakil & Blachstein, 1993, p. 886-887).

The influence of age and education on neuropsychological measures including the RAVLT is well-documented (Gómez-Pérez & Ostrosky-Solís, 2006; Pontón, Satz, Herrera, & Ortiz, 1996; Reitan & Wolfson, 1995; Schoenberg et al., 2006; Steinberg et al., 2005; Vakil et al., 2010). Thus, in the third analysis, construct validity was evaluated by examining correlations between RAVLT-S raw scores and age and education. In addition, convergent and discriminant validity was evaluated by examining the correlations between RAVLT-S age-corrected z scores with other tests of cognitive abilities. Convergent validity was evaluated by examining the strength and direction of the relationship between RAVLT-S scores on Learning Trials 1 through 5 and other measures of verbal learning, and between RAVLT-S scores on long-term memory trials (6 through 8) and measures of verbal memory. Conversely, the discriminant validity of the RAVLT-S was evaluated by examining the relationship between scores on Trials 1 through 5 and Trials 6 through 8 to neuropsychological measures that do not explicitly measure verbal learning and memory, respectively. These analyses were facilitated by converting raw RAVLT-S scores to z scores, creating Acquisition (Trials 1 through 5)

and Retention (Trials 6 through 8) factor-based scores, then conducting correlational analyses.

Hypotheses

Given the psychometric properties of the neuropsychological instruments used, the results of the single sample *t* test, factor analyses, and correlational analyses were expected to provide support for the construct and criterion validity of the RAVLT-S. Specifically, the following six hypotheses were made: (1) the RAVLT-S would demonstrate acceptable reliability consistent with similar English version tests; (2) mean RAVLT-S scores for the TBI group would be selectively reduced as compared to those of the healthy, age-controlled standardization sample (Schmidt, 1996); (3) sensitivity, specificity, and positive and negative predictive power would exceed chance; (4) the RAVLT-S would be composed of two factors, one assessing acquisition and another assessing retention; (5) raw RAVLT-S scores would yield expected patterns of associations with demographic variables, such that older and lower educated subjects would perform more poorly; and (6) standardized RAVLT-S scores would be strongly correlated with other measures of verbal learning and memory (WMS Immediate Recall and WMS Delayed Recall, respectively, and WAIS Working Memory Index), less so with verbal measures that lack an explicit memory component (WAIS Verbal Comprehension Index, COWAT Phonemic Fluency), and insignificantly to measures of perceptual and motor abilities (Trail Making Test Parts A and B, WAIS Perceptual Organization and Processing Speed indexes).

Chapter 3

Method

Participants

Participants included 106 adults (Mean age = 39.3 years, $SD = 17.9$; 50.0% male). All participants were Hispanic, were born and lived in Puerto Rico at the time of the assessment, and reported Spanish as their dominant language in expressive, comprehension and writing skills. Of these, 68 sustained traumatic brain injury (TBI group) and 38 were diagnosed with Major Depressive Disorder (DEP group). Participants in the TBI group were 30.3 years old on average ($SD = 12.0$), had 13.0 years ($SD = 3.1$) of education, and were 60.3% male ($n = 41$). They were included in the TBI group if they had sustained a traumatic brain injury with evidence of structural brain damage based on comprehensive neurological evaluation utilizing appropriate neuroimaging, laboratory, and examinational findings, and had been administered the RAVLT-S as part of their neuropsychological evaluation. Based on review of the medical records, all participants sustained complicated TBI that was moderate to severe in nature. Participants in the DEP group were 55.3 years old on average ($SD = 15.2$), had 15.0 years of education ($SD = 2.5$) and were 31.6% male ($n = 12$). They were included in the DEP group if they were diagnosed with DSM-IV-TR (APA, 2000) Major Depressive Disorder (MDD), had no other co-existing neurological or neurodevelopmental disorder, and were administered the RAVLT-S as part of their neuropsychological evaluation. Diagnosis of MDD was made by a licensed psychologist or psychiatric, based on routine psychological and psychiatric examination.

Measures

Rey Auditory Verbal Learning Test-Spanish (RAVLT-S). The RAVLT-S (Acevedo-Vargas, 2002) is presented in Table 1. The RAVLT-S is an auditory test of verbal learning and memory that provides a measure of the ability to encode, consolidate, store, and retrieve verbally acquired information. The standard format starts with a list of 15 words (List A), which are read aloud by the examiner at the rate of one word per second. The individual's task is to repeat all the words s/he can remember, in any order. This procedure is carried out a total of five times (Learning Trials 1 through 5) without delay between trials. The examiner then presents a distractor set (List B) which consists of a different list of 15 words, and the individual is allowed a single attempt at recall. Immediately following the distractor task, the individual is asked to remember as many words as possible from List A (Trial 6). In contrast, the Delayed Recall (Trial 7) requires the individual to remember as many words from the original list after a period of 20 minutes. Delayed Recognition (Trial 8) requires the individual to confirm whether words read aloud by the examiner were or were not on the original list, following a 20 minute delay.

The primary metric produced by the RAVLT-S is the amount of words recalled for each of the five learning trials and each of the delayed tasks. Data yielded by list learning tasks such as the RAVLT can also be used to evaluate learning curve patterns, the effects of intrusions or perseverations, cognitive strategies, and patterns of recall such as primacy or recency effects (Ardila et al., 1994; Spreen & Strauss, 1998). Age-controlled metanormative data on healthy, English-speaking individuals were compiled for the RAVLT by Schmidt (1996), and used for comparative purposes in this study.

Wechsler Adult Intelligence Scale–third edition (WAIS-III). The WAIS-III is an individually-administered battery designed to assess general intellectual functioning and cognitive strengths and weaknesses in adults. In 2008, a Spanish adaptation, the *Escala de Inteligencia Wechsler para Adultos–Tercera Edición* (EIWA-III; Wechsler, 2008), was developed in collaboration with the Ponce School of Medicine, Puerto Rico. This revision made explicit use of culturally-relevant stimuli and included demographically adjusted norms compiled with Puerto Rican census data from 2000. The EIWA-III can be administered to Spanish-speaking individuals from 16 to 64 years of age, and retains the overall structure of the WAIS-III. Individual abilities are grouped into four global areas: Verbal Comprehension Index (VCI), which provides a measure of verbal ability; Perceptual Organization Index (POI), which involves the manipulation of concrete materials or processing of visual information to solve problems nonverbally; Working Memory Index (WMI), which provides a measure of auditory short-term memory; and Processing Speed Index (PSI), which provides a measure of cognitive processing speed and efficiency. These four Composite Indexes comprise the Full Scale IQ (FSIQ) which serves as an estimate of general intellectual ability. Like the WAIS-III, the EIWA-III yields two additional sets of summary scores: a verbal scale or Verbal IQ (VIQ) comprised of the VMI and WMI indices, and a performance scale or Performance IQ (PIQ) comprised of the POI and PSI indices.

Each Composite Index and the FSIQ yields a standard score with an average of 100 and standard deviation of 15. The subtests that constitute each of the indices have an average score of 10 and a standard deviation of 3. Percentile ranks may also be reported for each score.

Wechsler Memory Scale (WMS). The WMS (Wechsler, 1945) is a standardized memory scale originally developed for use with clinical populations. It is comprised of seven subtests that measure visual and auditory learning and memory. Normative data are available for adults ranging in age from 16 to 89. The WMS provides measures of general memory capacity, recognition ability, the ability to remember visual and auditory information, and the capacity to remember and manipulate information in short-term memory.

In 1994, a Spanish adaptation was developed and normative data published as part of a research program undertaken in Columbia (Ardila et al., 1994). Among the available measures included in the Spanish adaptation (and used in this study) are subtests that comprise the Logical Memory (LM) indexes. LM indexes provide measures of verbal auditory memory by requiring examinees to recall stories presented orally by the examiner. Logical Memory Immediate (LMI) provides a measure of short-term memory by requiring the examinee to recall as much information as possible immediately following the recitation of the story. Logical Memory Delayed (LMD) assesses long-term memory by requiring the examinee to recall information after a 30-minute interval. Standard scores and percentile ranks are used for making comparisons between subtests, across age ranges, and with other tests of cognitive functioning.

Trail Making Test (TMT). The TMT (Army Individual Test Battery, 1944; Reitan & Wolfson, 1985; Reitan & Wolfson, 1993) is a commonly used neuropsychological measure of visuomotor ability, including visual attention, speed of processing, motor speed and executive function. The TMT is composed of two parts, A and B. Each part requires the individual to connect dots on an 8 ½ x 11 inch sheet of

paper, in sequential order, by drawing lines from one dot to the next. Part A requires individuals to draw lines connecting 25 consecutive numbers that are arranged on the sheet of paper with no apparent pattern. Part B is essentially the same, but requires the individual to alternate between consecutive numbers and letters in alphabetical order (i.e. lines must be drawn in the sequence 1→A→2→B→3→C) and is therefore more difficult. Notably, TMT Part B has been shown to differentiate between individuals with and without cerebral brain damage (Reitan, 1958).

The primary metric produced by the TMT is the amount of time required to complete each part. Raw scores for each part are represented by total time in seconds, and can be converted to standardized scores for normative comparisons. Age- and education-controlled Spanish language normative data have been compiled (Pena-Casanova et al., 2009; Perianez et al., 2007).

Controlled Oral Word Association Test (COWAT) Phonemic Fluency

subtest. The COWAT is one of seven subtests comprising the Multilingual Aphasia Examination (MAE; Benton & Hamsher, 1989). The MAE is a relatively brief battery designed to evaluate the presence and severity of aphasic disorders. Tests of oral expression, oral verbal comprehension, reading comprehension, and spelling/writing are used to evaluate receptive and expressive speech. The COWAT provides a measure of phonemic verbal fluency and executive functioning (organized retrieval) that requires examinees to produce as many words as possible, within a 60-second timeframe, that begin with each of three different letters (originally C, F, and L; or P, R, and W; now most commonly F, A, and S). Proper nouns, numbers, and the same word with alternate suffixes are disallowed. In 1991, a Spanish adaptation of the MAE was developed

(MAE-S; G. J. Rey, 1990). The MAE-S was standardized and normed using data from 234 Spanish speaking adults, is designed for use with adults through age 69, and has been reported to be a sensitive and accurate measure of language disturbances in Hispanics with TBI (G. J. Rey et al., 2001). In its adaptation, the traditionally-used English-language COWAT letters were replaced with the letters P, T, and M, in order to maintain similar frequencies in the respective languages (G. J. Rey et al., 2001). This version of the Phonemic Fluency subtest was used in this study.

Procedure

Participants were selected from a consecutive series of cases that were referred for neuropsychological assessment to a neuropsychology consultation service at the Neurology Section of the University of Puerto Rico Medical School. A licensed neuropsychologist or doctoral level graduate student with extensive training and appropriate supervision related to the tests administered conducted all evaluations. All tests were administered according to standardized procedures.

Data Analysis

Data screening. Prior to conducting the analyses, neuropsychological data were examined for outliers using descriptive statistics and box plots. Outliers were defined as scores 2.5 *SDs* above or below the mean. When outliers were identified, scores were adjusted using standard procedures (Reitan, 2008; Tabachnick & Fidell, 2007) to decrease their influence on measures of central tendency. Skewness and kurtosis analyses were conducted to determine whether variables were normally distributed. The extent and pattern of missing values within the dataset were also examined.

Internal consistency. A split-half procedure was used to examine internal consistency. This approach to estimating internal consistency has been applied to other verbal list learning tasks, such as the California Verbal Learning Test–II (Delis, Kramer, Kaplan, & Ober, 2000). It was expected that the reliability estimate for the RAVLT-S would be comparable to that reported for the CVLT-II. As with the CVLT-II, a split-half correlation was calculated by computing odd-even correlations between immediate free recall Trials 1 + 3 versus Trials 2 + 4, and Trials 2 + 4 versus Trials 3 + 5. The correlation between Trials 1 + 3 versus Trials 3 + 5 was also calculated. However, because the reliability of a test depends in part on its length, and improves as length increases (Franzen, 2000), splitting the RAVLT-S effectively underestimated its reliability relative to its actual length. To address this problem, the Spearman-Brown formula was applied to the average of these correlations with a lengthening factor of 2.5, in order to extrapolate the reliability coefficient appropriate to the full-length test. These internal consistency estimates were calculated for each group separately (TBI and DEP), for the entire sample, and for the entire sample by gender.

Criterion validity: sensitivity and specificity. In order to examine the criterion validity of the RAVLT-S, the TBI group was compared to the DEP group and the Normative Sample (NS). Comparisons between the TBI and the DEP group were accomplished using multivariate analysis of covariance (MANCOVA), where age-corrected RAVLT-S z scores served as the dependent variables and diagnosis served as the between-subjects variable. Given the potential differences between the groups on education, this variable was included in the analysis as a covariate. Follow-up univariate analyses of covariance (ANCOVA) were used to examine differences between the groups

on the individual RAVLT-S trial scores, given that the overall MANCOVA was significant. To examine differences between the TBI and DEP groups across trials, age-corrected z scores were then subjected to a repeated measures MANCOVA, where RAVLT-S trials served as the repeated measure, diagnosis as the between-subjects factor, and education as a covariate. Comparisons were also made between the TBI group, DEP group, and age-corrected metanormative data for healthy, English-speaking individuals compiled by Schmidt (1996). These comparisons were conducted using single sample t tests. Finally, z -scores were used to calculate sensitivity, specificity, positive predictive power, and negative predictive power for each of the RAVLT-S trials, with the DEP group serving as a comparison sample. In these analyses, performance of 1.5 SD s or more below the meta-normative sample mean were used to indicate the presence of neurological dysfunction, as is commonly reported in neuropsychological literature (e.g., Schmidt, 1996; Schoenberg et al., 2006).

Construct validity: factor structure. Confirmatory factor analysis (CFA) was used to examine construct validity, evaluate the structure of the domain, and summarize variables. CFA's were conducted to determine whether a one- or two-factor model best fit the RAVLT-S data in our sample. The first model (VM₁₋₈) was a one-factor model in which all eight RAVLT-S trials were specified to load on a single factor. This model was used to evaluate whether learning (List A) and memory (Immediate Recall, Delayed Recall, and Recognition) trials assess a single latent trait we termed *verbal memory*. The second model (A₁₋₅R₆₋₈) represents our hypothesized two-factor model (Figure 3). In this model, Trials 1 through 5 were specified to load on the first factor, which we termed acquisition, because these trials are specifically designed to reflect learning by presenting

the 15-item list to the examinee sequentially and without delay. Trials 6 through 8, in contrast, were specified to load on a factor termed retention because they are administered after either the administration of a distractor set (Immediate Recall) or after a 20 minute delay (Delayed Recall, Delayed Recognition) and are hypothesized to reflect the retention of information in long-term memory. As there are many fit indices for evaluating how well a particular model reproduces the original variance-covariance or correlation matrix, the use of multiple indices is recommended to reduce the possibility of Type I and Type II error under various conditions (Hu & Bentler, 1999; Kenny & McCoach, 2003). Four goodness-of-fit statistics that evaluate different aspects of model fit (Bentler, 1990; Kline, 2005) were thus examined.

First, the maximum-likelihood (ML) chi-square test estimates the probability that a dataset will be observed given a particular hypothesized model. It tests the null hypothesis that a hypothesized model (with fewer path coefficients) is as likely to reproduce the data as the saturated model (in which there is a direct path from each variable to the other). A non-significant chi-square indicates good fit. While a significant chi-square can be used as evidence of poor fit between the sample data and hypothesized model, this statistic is sensitive to large sample sizes (Bentler & Bonett, 1980). With very large samples, even well-fitting models differ significantly from the saturated model, and may be rejected. Chi-square is also susceptible to Type II error when sample sizes are small (Kenny & McCoach, 2003). Nevertheless, because chi-square remains a popular test statistic, and provides the mathematical basis for most other fit statistics, it is reported here.

The second index examined was the comparative fit index (CFI; Bentler, 1990). This is an incremental fit index that evaluates the fit of the hypothesized model relative to a baseline (independence) model in which all paths (i.e., path coefficients) between parameters are removed. In this way, CFI compares the observed covariance matrix to a null model that assumes that all latent variables are uncorrelated. CFI values range from 0 to 1, with values near zero indicating very poor fit and values equal to or greater than 0.95 indicating good fit (Hu & Bentler, 1999).

The third index evaluated was the root mean square error of approximation (RMSEA; Steiger & Lind, 1980; as cited in Steiger, 1990), which estimates how well the hypothesized model fits the population covariation matrix. Because the RMSEA is based on the chi-square to *df* ratio (with smaller ratios indicating better fit), it serves as a parsimony index by favoring the model with fewer parameters. RMSEA values also range from 0 to 1. Unlike the CFI, however, values closer to zero indicate better fit. RMSEA values of .05 or less indicate good fit, while values from .06 to .08 suggest adequate fit (Jöreskog & Sörbom, 1993). The RMSEA has the added benefit of making confidence interval calculations possible.

Lastly, the Akaike information criterion (AIC; Akaike, 1974) was used. Unlike the ML, CFI, and RMSEA goodness-of-fit statistics, AIC is not a test in the conventional sense, but a relative measure of fit. AIC derives from Information Theory (Shannon, 1948) and evaluates the relationship between Kullback-Leibler information (i.e., the distance between two probability distributions; Kullback & Leibler, 1951) and maximized log-likelihood (i.e., parameter estimation; Akaike, 1974). Within this framework, AIC quantifies the amount of information lost by each of two hypothesized

models, relative to each other, given the same set of data. AIC penalizes models by a factor of two for every parameter used, and thus also functions as a measure of parsimony. Lower values indicate better fit.

Construct validity: correlations. Correlations were calculated between RAVLT-S raw scores and age and education. Factor-based scores were created for RAVLT-S data by deriving the means for age-corrected z scores for Trials 1 through 5 (Acquisition) and Trials 6 through 8 (Retention) for the TBI group. Correlational analyses between Acquisition and Retention factor scores and explicit measures of learning and memory (WMS Immediate & Delayed recall, WAIS Working Memory Index) were then performed. Correlational analyses were also performed between RAVLT-S factor scores and verbal measures that lack an explicit memory component (WAIS Verbal Comprehension; COWAT Phonemic Fluency total score). Lastly, correlational analyses were performed between the factor scores and measures of perceptual and motor abilities (Trails A & B, WAIS Perceptual Organization and Processing Speed Indexes). Bonferroni corrections were used to control for Type I error.

Chapter 4

Results

Preliminary Analyses

Descriptive statistics for each neuropsychological variable are presented in Table 2. Missing RAVLT-S data, which represented less than 5% of the entire dataset, were replaced with the mean (Tabachnick and Fidell, 2007) for the following number of data points: two Trial 6 (Immediate Recall) scores; four Trial 7 (Delayed Recall) scores; three Trial 8 (Recognition) scores, and two education scores. In one case, a cut-off score of 300 for TMT B (Reitan, 2008) was assigned to one participant on the basis of that participant's extreme score on TMT A. Missing data for the remaining neuropsychological variables were found to be significant and non-randomly distributed. The pattern of missing data indicated that the DEP group was administered significantly fewer neuropsychological tests than the TBI group. This pattern is attributable to the clinical nature of the data gathered. Unlike experimental studies, where random assignment is possible and examiner controls are considerable, clinical patients are generally only administered tests relevant to their presenting condition. It follows that TBI patients were administered a greater number of neuropsychological tests than their DEP counterparts. Given the purpose and design of this study, therefore, data for neuropsychological variables were combined when analyses were not specifically evaluating differences between DEP, TBI and NS groups on the RAVLT-S, for which complete data were available for each group.

Skewness and kurtosis estimates within ± 1.0 were considered appropriate for statistical analysis. As Table 2 indicates, 5 of the 17 variables had skewness and/or

kurtosis estimates exceeding the ± 1.00 criteria (most markedly Trails A and Trails B scores, which displayed both skewness and kurtosis). Box plots indicated the presence of univariate outliers for the following variables: RAVLT-S Trial 1 and Recognition, Phonemic Fluency, and WMS Immediate and Delayed; extreme outliers were identified for both Trails A and Trails B scores. One WAIS index, Perceptual Organization, was kurtotic. Data were reviewed for imputation errors and determined to reflect valid data. Skewness and kurtosis were thus addressed as follows.

Option 1: the clinical literature was reviewed to determine the existence of applicable cut-off scores. When available, these were applied to the raw data. Option 2: because outliers were determined to reflect valid scores that provide clinically relevant information, raw scores were replaced with a score one unit larger or smaller than the next most extreme score in the distribution (Tabachnick & Fidell, 2007). This approach has the benefit of minimizing the influence of outliers by reducing the magnitude of differences between them and the rest of the distribution while preserving the order of legitimate scores. Option 3: if neither of the first two approaches were possible, extreme scores were replaced with the mean or deleted.

In each case, application of one of these approaches brought skewness and kurtosis estimates to within acceptable limits. The single exception to this was the WAIS POI, which was platykurtic. However, because scores were determined to represent legitimate data, no outliers were present, and the kurtosis estimate (-1.241) was within ± 2 standard error of kurtosis (SEK = .788; see Table 2), no transformations were deemed necessary. Where adjustments to scores were undertaken (i.e., RAVLT-S Trial 1 and Recognition, Phonemic Fluency, WMS Immediate and Delayed, TMT A & B),

correlational analyses were conducted at each step to evaluate the impact on the relationship among neuropsychological variables. In general, whereas removing scores decreased the strength of correlations, adjusting scores increased homogeneity and thereby strengthened correlations. These effects were uniformly negligible, however, having no appreciable or significant effect on the correlations of interest. As a result, original scores were retained for all analyses, because they provide important clinical information regarding the variability of scores at the low end of the distribution for TBI patients.

Demographic and clinical data are presented in Table 3. Examination of the demographic and clinical data indicates a 50/50 gender distribution within the entire sample of participants. Mean differences in gender, age, and education between the TBI and DEP groups were assessed with a one-way ANOVA. Levene tests of homogeneity did not indicate significant violations of the assumption of homogeneity of variance. The ANOVA was significant for gender, $F(1, 104) = 8.54, p = .004, \eta^2 = .07$, age, $F(1, 104) = 87.25, p < .001, \eta^2 = .46$, and education, $F(1, 104) = 11.87, p = .001, \eta^2 = .10$. The gender distribution within each clinical group varied such that the TBI sample was composed of more men (60.3%) than women, while the DEP sample was composed of more women (68.4%) than men. This is consistent with the increased prevalence of TBI in males and DEP in females observed in the general population. Age differences were also evident. The mean age of DEP participants exceeds that of TBI participants by a margin of 20 years, with large *SDs* observed within each group. The TBI group was composed of predominantly younger participants, with 82.4% being 40 years of age or younger, while 84.2% of participants in the DEP group were over 40. The groups

differed in terms of average years of education, such that participants in the DEP group were more educated than participants in the TBI group. However, given that education tends to increase as a function of age, and TBI is prevalent in younger age groups (as reflected in mean age differences between our TBI and DEP samples), this was not surprising. Finally, 13.2% of the TBI group was comprised of participants with ≤ 10 years of education, while none of the DEP participants had less than 11 years of schooling.

Reliability

Hypothesis 1 internal consistency: split-half correlations. Verbal learning and memory tasks pose unique challenges to estimating reliability due to item interdependence within and between trials. Because of this, it was expected that internal consistency might be lower than that of other types of neuropsychological tests, as has been reported in other studies examining the psychometric properties of list learning tasks (Delis et al., 2000). To address these unique issues, internal consistency was determined across the five immediate free recall trials using a split-half correlation. This estimate served as the primary measure of reliability on the RAVLT-S, given that scores on these trials are direct indicators of verbal learning and memory. Correlations between the RAVLT-S immediate free recall trial pairs for each clinical group are presented in Table 4. Reliability for the entire sample was excellent ($r = .95$), as was reliability for the DEP ($r = .94$) and TBI ($r = .95$) groups separately. Additionally, when the entire sample was divided based on gender, there were no differences in internal consistency between male ($r = .96, n = 53$) and female ($r = .94, n = 53$) participants. Results of these analyses are presented in Table 5.

Validity

Hypothesis 2 criterion validity: group comparisons. Examination of raw data demonstrated incremental improvement in learning from Trial 1 to Trial 5 for both groups, from a mean of 4.96 (TBI) and 4.34 (DEP) for Trial 1 to a mean of 8.51 (TBI) and 9.89 (DEP) for Trial 5. Raw data also yielded the expected pattern of performance on the remaining trials (Spreen & Strauss, 1998), such that fewer words were recalled during Trial 6 (Immediate Recall) than Trial 5, fewer still during Trial 7 (Delayed Recall), and the most during Trial 8 (Recognition).

Descriptive statistics for age-corrected RAVLT-S z scores are presented in Table 6. MANCOVA comparing the DEP and TBI groups while covarying education indicated a significant overall effect, $F(8, 96) = 7.01, p < .001, \eta_p^2 = .37$, as well as a significant effect for education, $F(8, 96) = 3.08, p = .004, \eta_p^2 = .20$, and diagnosis, $F(8, 96) = 2.22, p = .032, \eta_p^2 = .16$. Results of follow-up ANCOVAs controlling for education differences between the DEP and TBI groups are also presented in Table 6. As can be seen from the Table in the DEP versus TBI columns, results of the ANCOVAs indicated significant differences between the groups on Learning Trials 3 and 5, and on Immediate and Delayed recall trials. Covarying out the effects of education had the effect of decreasing the TBI group's scores relative to the DEP group, and in this way increased the magnitude of differences between the groups. Comparisons between the TBI group and the normative sample (NS) using single sample t tests are also presented in Table 6, in the TBI versus NS columns. Results indicated that the TBI group performed significantly worse than the NS on all RAVLT-S trials.

Age-corrected z scores were then subjected to a repeated measures MANCOVA with RAVLT-S trials serving as the repeated measure, clinical group as the between-subjects factor, and education as a covariate. Results indicated a significant effect for trial, $F(7, 721) = 3.03, p = .004, \eta_p^2 = .029$, a significant effect for group, $F(1, 103) = 4.10, p = .046, \eta_p^2 = .038$, as well as a significant trial by group interaction effect, $F(7, 721) = 3.61, p = .001, \eta_p^2 = .034$. The trial by education effect was not significant, $F(7, 721) = 1.37, p = .22, \eta_p^2 = .013$. The trial by group interaction effect was primarily accounted for by a decreased rate of learning in the TBI group compared to the DEP group. To visualize this interaction, raw scores corrected for age and education for Learning Trials 1 through 5 are presented in Figure 1.

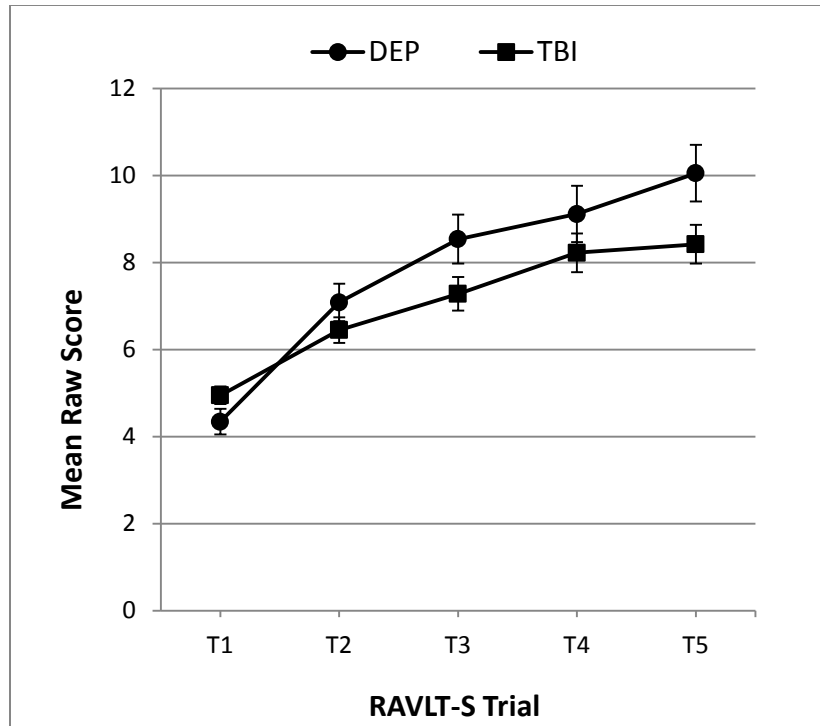


Figure 1. Rey Auditory Verbal Learning Test-Spanish age- and education-corrected raw scores for the Depression and Traumatic Brain Injury groups. DEP = Depression group; TBI = Traumatic Brain Injury group; T1 = Trial 1; T2 = Trial 2; T3 = Trial 3; T4 = Trial 4; T5 = Trial 5. Error bars represent standard error.

Finally, comparisons were made between the TBI and DEP groups relative to metanormative data on healthy, English-speaking adults (Schmidt 1996). Figure 2 illustrates the differences in age-corrected z scores for the DEP, TBI and NS groups. As the Figure indicates, the DEP group performed approximately 1 SD below the NS mean across most of the RAVLT-S trials. The main exception was the Recognition trial, for which DEP scores fell approximately 1.5 SD s below the NS mean. The TBI group initially performed like the DEP group. However, the TBI group's performance dropped off more steeply across the learning trials, suggesting a decrement in learning. This resulted in a Total score for Trials 1 through 5 that was approximately 2.5 SD s below the

NS mean. Scores on the Immediate and Delayed Recall trials were also depressed relative to the DEP group. Like the DEP group, TBI group performance declined sharply on the Recognition trial, averaging over 3 SDs below the NS mean.

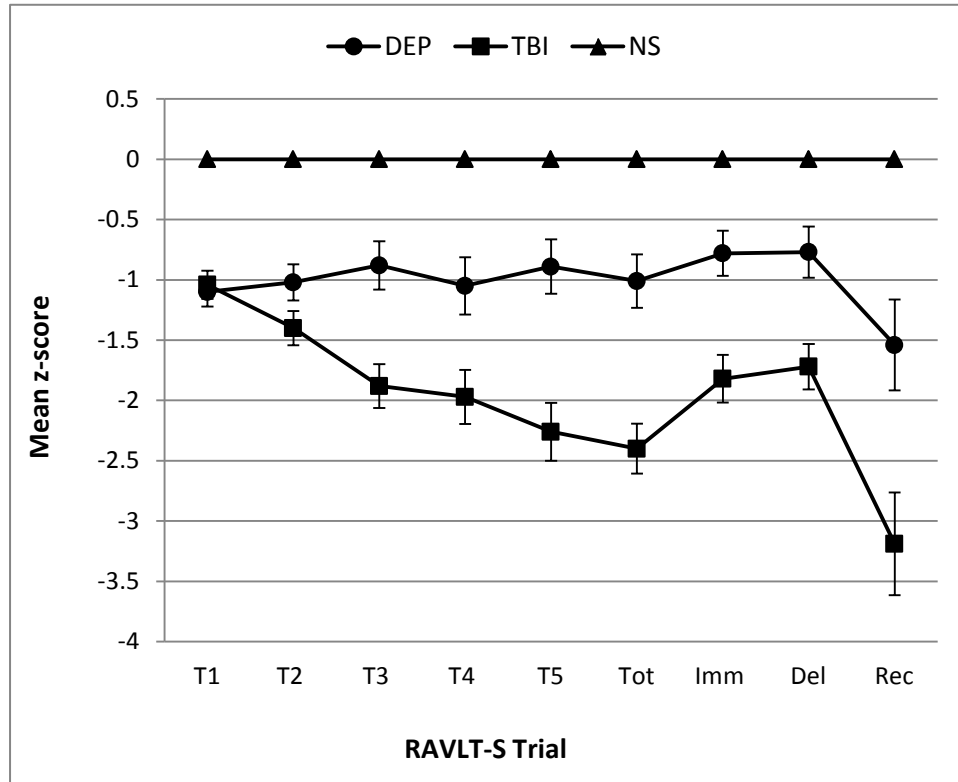


Figure 2. Rey Auditory Verbal Learning Test-Spanish age-corrected z scores for the Depression, Traumatic Brain Injury and Normative Sample groups. DEP = Depression group; TBI = Traumatic Brain Injury group; NS = Meta-normative Sample; T1 = Trial 1; T2 = Trial 2; T3 = Trial 3; T4 = Trial 4; T5 = Trial 5; Tot = Total (Trials 1-5); Imm = Immediate Recall; Del = Delayed Recall; Rec = Recognition. Error bars represent standard error.

Hypothesis 3 criterion validity: classification statistics. Classification statistics based on RAVLT-S age-corrected z scores are presented in Table 7. These include sensitivity, specificity, positive predictive power, negative predictive power, and other

recommended diagnostic efficiency statistics for standardized reporting (Kessel & Zimmerman, 1993). Sensitivity was highest for the Total score (.65) and lowest for Trial 1 (.42). Specificity ranged from .58 to .76, and was highest for Immediate and Delayed recall. Positive predictive power was highest for Immediate Recall (80.4) and negative predictive power was highest for the Total score (52.9). Overall correct classification rate (HR) ranged from .53 to .67, with HR highest for the Total score.

Hypothesis 4 construct validity: confirmatory factor analysis. Confirmatory factor analyses (CFA) were conducted using SPSS Amos (Arbuckle, 2006). Results of the four goodness-of-fit statistics described above for each of our two models are presented in Table 9. To increase sample size, these analyses were conducted using the entire sample as well as the TBI sample. As the table indicates, the one-factor model (VM₁₋₈) yielded significant Chi-squares when either sample was evaluated. Although in both cases the VM₁₋₈ model had CFIs above .95, suggesting good fit, it also had the highest AICs, and RMSEA values above .08, which are generally considered inadequate. Overall, the two-factor model (A₁₋₅R₆₋₈) provided improved fit and parsimony, yielding smaller Chi-squares (non-significantly so in the TBI sample), CFIs \geq .98, smaller AICs, and RMSEA values \leq .08. When considering the A₁₋₅R₆₋₈ model among the two samples, the TBI sample yielded the most favorable goodness-of-fit statistics. While the small size ($n = 68$) of the TBI sample calls the stability of the solution into question, it was included here because it did not differ meaningfully from the results of the entire sample. All RAVLT-S trials (with the exception of Trial 1) had excellent loadings on their respective factors. The A₁₋₅R₆₋₈ model thus represented the optimal model. Results of path analyses for the TBI sample versus TBI + DEP sample are illustrated in Figure 3.

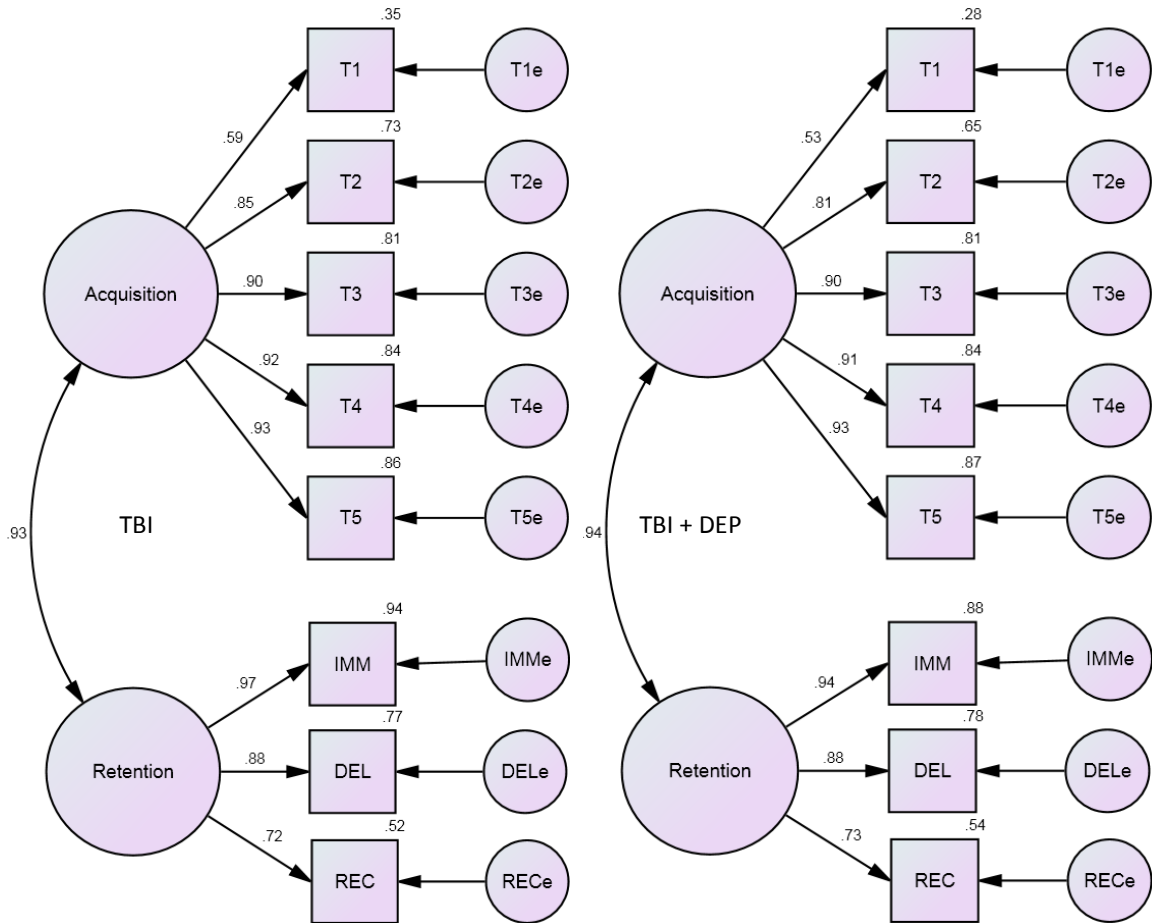


Figure 3. Path diagram of two-factor model of verbal learning and memory in the Rey Auditory Verbal Learning Test-Spanish. T1 = Trial 1; T2 = Trial 2; T3 = Trial 3; T4 = Trial 4; T5 = Trial 5; IMM = Immediate Recall; DEL = Delayed Recall; REC = Recognition; e = error variable. TBI = traumatic brain injury group; DEP = depression group; TBI + DEP = combined TBI and DEP groups.

Hypothesis 5 construct validity: correlations with demographic variables.

Correlations between RAVLT-S raw scores and age and education for the entire sample are presented in Table 8. As the Table shows, only Trial 1 was significantly correlated with age. This correlation was negative, indicating that as age increased, Trial 1 scores

decreased. In contrast, all RAVLT-S trials were significantly positively correlated with education, indicating that as education increased, performance improved.

Hypothesis 6 construct validity: correlations with other tests. Factor scores were created for RAVLT-S data by deriving the means for age-corrected z scores for Trials 1 through 5 (Acquisition) and Trials 6 through 8 (Retention) for the TBI group. The correlations among RAVLT-S factor scores and Phonemic Fluency, WAIS-III indices, WMS Long-term Memory Immediate and Delayed subtests, and TMT A & B scores for the entire sample ($N = 106$) are presented in Table 10. As can be seen in the table, the Acquisition factor was significantly correlated with all neuropsychological measures at the $p < .05$ level of significance after controlling for Type I error. Only Trail Making Test A & B scores were negatively correlated. The Retention factor was also significantly correlated with 7 of the 9 neuropsychological measures. As with the Acquisition factor, TMT A & B scores were significantly negatively correlated with Retention. Unlike the Acquisition factor, however, the Retention factor was not significantly correlated with either the Verbal Comprehension or Working Memory indexes of the WAIS-III, which underlie Verbal IQ.

Chapter 5

Discussion

There is a growing need for validated, linguistically diverse neuropsychological measures in the United States. These analyses were undertaken to examine the construct and criterion validity of a Spanish translation of the RAVLT when used to evaluate Hispanic TBI patients. It was hypothesized that RAVLT-S scores for TBI patients would be selectively reduced as compared to the standardization sample, reflecting the presence of brain dysfunction. It was also hypothesized that expected patterns of associations with demographic variables and neuropsychological measures would be observed, such that RAVLT-S scores would be influenced by age and education level, and be more strongly associated with other direct measures of learning and memory. Lastly, it was hypothesized that the underlying factor structure of the RAVLT-S would parallel that of its English-language counterpart, such that two factors reflecting acquisition and retention would represent the data well. The results of our analyses provide initial support for the construct and criterion validity of the RAVLT-S.

Internal consistency estimates of the RAVLT-S were excellent and comparable to those derived by the same method for commonly used English language measures of verbal learning and memory (e.g. CVLT-II, $r = .94$). Regarding associations with demographic variables, results were mixed. No significant correlations between age and raw RAVLT-S scores were observed for either the TBI or DEP group. When these groups were combined, only Trial 1 was significantly negatively correlated with age (see Table 8). Although this relationship occurred in the expected direction, overall findings are inconsistent with extant research on the RAVLT, which tends to support a negative

correlation between age and performance in adults across most trials. However, Reitan (1985, 1986, 1993) has argued that the presence of brain injury disrupts the relationships typically observed between neuropsychological test performance and age, such that severity of brain injury, rather than age, becomes the main predictor of test performance. Education, on the other hand, was significantly positively correlated with all RAVLT-S trials whether evaluating raw or age-corrected z scores. When considering TBI age-corrected z scores separately, all but the Delayed and Recognition trials ($r = .194, p = .11$ and $r = .214, p = .08$, respectively) were significantly positively correlated with education. These results are consistent with research evaluating the effects of education on RAVLT scores, which indicates a positive correlation between the number of years of education and higher scores.

Regarding criterion validity, single sample t tests using age-corrected z scores revealed that the TBI group had significantly lower scores ($p < .001$) than the standardization sample on all RAVLT-S trials, indicating impairment in both learning rate and retention. Similarly, on all but Trials 1 and 2, the TBI group had significantly lower scores than the DEP group, which itself had scores significantly lower than the standardization sample. While the overall pattern of performance observed (i.e. NS scores > DEP scores > TBI scores) was expected, our results differ somewhat from what is typically reported for the English version RAVLT, inasmuch as it is generally insensitive to psychiatric illnesses such as depression (Schmidt, 1996; Schoenberg et al., 2006). When the effects of education were controlled for, the differences between the TBI and DEP groups appeared to diminish further, such that differences on three of the learning trials and one of the long-term memory trials were no longer statistically

significant. While this suggests a greater similarity in scores than expected between the two groups, it also likely reflects the use and evaluation of non-normalized data in our analyses, since z scores were calculated using age group means derived from the standardization sample, and were not normalized. Nevertheless, given the memory and concentration deficits associated with depression, such findings are plausible. On the other hand, these results may be influenced by the atypical lack of significant correlations observed between age and RAVLT-S scores, since it would be expected that controlling for age would confer a relative benefit to older subjects such as those comprising the DEP group. Alternatively, our results may indicate sampling bias, since DEP participants were selected from a neuropsychological consultation service via the neurology section of a medical school. DEP participants might therefore be expected to have presented with more severe symptoms, including more substantial memory deficits, which may have prompted referral for neuropsychological evaluation. In light of the preliminary nature of our investigation regarding the psychometric properties of the RAVLT-S, further analyses with larger DEP samples (for which severity measures of depressive symptomatology at time of testing are available, for example), are recommended. TBI group performance, on the other hand, was consistent with extant research on the English language version of the RAVLT, providing strong support for its efficacy in classifying neurological dysfunction.

The clinical utility of the RAVLT-S was further evaluated by examining the sensitivity, specificity, positive predictive power, and negative predictive power of the trial scores, with scores $\leq 1.5 SD$ below Schmidt (1996) age-matched metanorms classifying neurological dysfunction. There was substantial variability in the

classification statistics yielded in our study. Sensitivity and specificity estimates were modest (65% SENS for Total and 76% SPEC for both Immediate and Delayed Recall). Positive predictive power was good, with Total and Immediate Recall trials yielding PPP rates of 80% (indicating neurological dysfunction in 8 out of 10 cases). Negative predictive power was poor across all trials, with the Total yielding the highest NPP value (53%), which approached chance. However, NPP was higher than the base rate for non-TBI participants in this study (52.9 vs. 35.8). Nevertheless, normal scores cannot readily be used to rule out neurological dysfunction. These results are similar, though not in complete agreement with Schoenberg's (2006) classification rates, which also reflected substantial variation. For example, Schoenberg's study yielded 81% and 83% SPEC statistics for Immediate and Delayed trials, respectively, and a best HR statistic of 69% to our 67% (albeit for different trials, Delayed versus Total). As noted by the author, the reduced ability of the RAVLT to distinguish the presence of neurological dysfunction from psychiatric participants likely reflects the heterogeneity of neurological impairment that may occur with TBI. As in that study, no effort was made in our study to select patients whose TBI was suspected of adversely affecting learning and memory specifically. Finally, the PPP values reported here are most applicable to neuropsychological clinics with relatively high base rates of neurological impairment. An important consideration for clinicians is that changes in the base rate will appreciably affect PPP and NPP statistics. For example, a base rate of 80% results in an improved PPP value of .90, with a commensurate decrease in NPP to .34. Overall, the classification statistics of the RAVLT-S support its use as a part of the armamentarium used to confirm brain injury.

Factor analysis is a useful empirical approach to evaluating construct validity, because factors are presumed to represent the underlying processes that explain the relationships among variables. Thus, if we assert that performance on a particular test validly reflects a set of innate, underlying mental processes (as may be argued of neuropsychological tests), then we would expect the factor structure to remain constant from one population to another, regardless of the specific configuration of scores for a particular group (Hilliard, 1979; as cited in Reynolds, 2000). This would simultaneously yield evidence for the psychometric precision of the test and for the universality of the underlying mental processes measured – both important considerations when evaluating the role of culture and language on test performance. Yet as Reynolds (2000) argues, even if we reject assumptions of innateness, consistent factor analytic results across populations provide strong empirical support that individuals of different groups perceive and interpret the test materials in the same manner, and that the same construct is being measured from one population to another. Assuming other relevant factors such as language proficiency have been addressed during test selection and administration, clinicians can more confidently approach the diagnostic process when factor analytic data are available for diverse ethnic populations, including Hispanics.

Regarding our factor analyses, the underlying factor structure of the RAVLT-S was consistent with the hypothesis that learning trials and recall trials tap distinct but related cognitive abilities (Erickson & Scott, 1977; Vakil & Blachstein, 1993). A major limitation of the factor analytic portion of our study, however, is sample size. While there is some evidence that samples ranging from 50 to 150 may be adequate under certain conditions, and that given strong correlations and few distinct factors a small

sample size may be adequate, factor analysis remains a large sample technique for which a minimum of 300 cases is recommended (Tabachnick & Fidell, 2007). Factor analytic studies including the English language version of the RAVLT, however, indicate that significantly fewer than 300 cases are often analyzed (Baños et al., 2005; Talley, 1986; Vakil & Blachstein, 1993), and that at least one published study was conducted with data from as few as 108 subjects (Ryan et al., 1984). Thus, given the preliminary nature of our analyses, and the limited availability of empirical data for the specific clinical population in question, factor analysis was considered worthwhile.

In an effort to increase sample size and strengthen the stability of the solutions, data for the TBI and DEP groups were combined and analyzed. Overall, a two factor model ($A_{1-5}R_{6-8}$) reflecting Acquisition (defined by loadings from Learning Trials 1 through 5) and Retention (defined by loadings from Trials 6 through 8) provided optimal fit and parsimony (see Table 9). Although inferences are constrained by the limits of a small sample size, the TBI sample yielded results synonymous with those of the TBI + DEP group, with more favorable goodness-of-fit statistics. In both cases, all but Trial 1 of the RAVLT-S had excellent loadings on their respective factors. This finding has been replicated elsewhere and reported to be an indicator of attention or immediate short-term memory (Baños et al., 2005; Lezak, 1983; Spreen & Strauss, 1998; Talley, 1986). The Acquisition and Retention factors were strongly correlated (TBI = .93; TBI + DEP = .94).

Another limitation of the factor analysis portion of our study is the unavailability of List B data, which provides information regarding the effects of a distractor set (i.e., proactive interference) on the plasticity of verbal memory. Proactive interference is defined as the decrease in number of words recalled on List B as compared to Trial 1 of

List A. Conversely, retroactive interference is defined by losses on Trial 6 (Immediate Recall), as compared to Trial 5 as a result of the administration of List B. Because List B had, in fact, been administered (though data were unavailable), and retroactive interference has been reported to characterize TBI (Spreeen & Strauss, 1998; Vakil et al., 2010), retroactive interference was evaluated. One-way ANOVA using raw scores while covarying the effects of age and education, however, indicated no significant differences in retroactive interference between the DEP and TBI groups in our sample. It is possible, however, that comparisons between normal controls and TBI participants would yield more pronounced differences.

On the other hand, although the use of List B is common in English-speaking countries such as the U.S., its administration was not a component of the original RAVLT and is routinely not included by practitioners in other countries (Spreeen & Strauss, 1998; van den Burg & Kingma, 1999). It has also been reported that List B scores load with learning Trials 1 and 2 (Baños et al., 2005; Talley, 1986), and thus may not have appreciably altered the structure of our solutions. Regarding the number of factors identified, our results differed from those reported by Baños, Elliott, and Schmitt (2005) and Tally (1986), who reported three factor solutions. This may be explained by the data or population included in the respective analyses. Unlike the Baños, Elliott, and Schmitt (2005) study, for example, our study did not include data regarding Delayed Recall false positives. Said data singularly defined a third factor for Baños and colleagues labeled “inaccurate recall” in addition to a “general verbal memory” factor and “auditory attention” factor (p. 377). Baños and colleagues’ study was conducted with patients with spinal cord injury. The Tally (1986) study, conducted with learning

disabled children, included in its analyses data derived from the digit span subtests of the WISC-R, which largely accounted for a third, “short-term memory factor with low coding demands” (p. 315). Direct comparisons could not be made to Ponton et al.’s (2000) factor analytic study of the Spanish-language NeSBHIS (which identified a verbal learning factor comprised of the WHO/UCLA AVLT) because in that study only Trial 5, Immediate Recall, and Delayed Recall were entered into the analyses. Similarly, differences in test structure preclude comparisons of the RAVLT-S to other Spanish verbal serial learning tasks, such as the 10-item bi-syllabic format described by Ardila et al. (1994).

The results of our factor analyses were most consistent with those of Vakil and Blachstein (1993), conducted with neurologically normal, non-psychiatric Hebrew-speaking participants, inasmuch as a two factor solution fit the data well. Here it is important to mention, however, that while the two-factor, acquisition and retention model that provided optimal fit in our study is theoretically congruent with Erickson & Scott (1977) and reflected in the results of Vakil and Blachstein (1993), Vakil and Blachstein derived their factors largely through the use of score sets. Nevertheless, the score sets that define the retention factor in their study are significantly influenced by the delayed measures, and the acquisition factor is almost exclusively defined by the learning rate (defined as Trial 5 minus Trial 1, and corresponding well to our A₁₋₅ factor). It is also important to note that Vakil and Blachstein (1993) further parsed the retention factor to reflect storage and retrieval, yielding a three factor solution that also represented the data well.

Factor-based scores labeled Acquisition and Retention were created by deriving the means of age-corrected z scores for Trials 1 through 5 and Trials 6 through 8, respectively. A high degree of similarity was observed among the correlations between Acquisition and Retention and the other neuropsychological measures, underscoring the relatedness of the learning and memory functions measured by the RAVLT-S. The primary distinction between Acquisition and Retention was reflected in the relationship between these factor-based scores and the WAIS-III Verbal Comprehension and Working Memory indexes. Whereas VCI and WMI were significantly correlated with Acquisition, they were not significantly correlated with Retention. This is an interesting distinction given that VCI and WMI comprise the Verbal IQ (VIQ) index of the WAIS-III, which reflects cognitive abilities that we hypothesized would underlie significant relationships. While this was partially borne out by the Acquisition factor, it is unclear what cognitive processes, if any, selectively reduced the strength of the relationship between the VCI and WMI and the Retention factor. Our results also yielded larger correlations with Phonemic Fluency than reported in other analyses of Spanish language adaptations of the RAVLT (Ponton et al., 2000), although the statistical significance of those correlations were not reported. Finally, our hypotheses stated that the RAVLT-S would be insignificantly related to neuropsychological measures of perceptual and motor abilities (TMT A & B, WAIS Perceptual Organization and Processing Speed indexes). This was not the case. Not only were TMT A & B significantly negatively correlated to the RAVLT-S, but the largest correlations occurred between the Acquisition and Retention factors and the indexes underlying the WAIS Performance IQ (PIQ), which we predicted would be unrelated. This is a counterintuitive finding given the nature of the cognitive

demands of the POI and PSI, which are predominantly visuospatial, graphomotor, and abstract in nature.

Further consideration, however, suggests that the observed relationships between RAVLT-S scores and measures of perceptual motor abilities may provide both convergent validity and support for the clinical utility of the RAVLT-S if it functions as a *general* measure of brain dysfunction. Neither the TMT nor the tasks underlying the POI and PSI are pure or sole measures of perceptual or motor abilities. Poor performance on the TMT, for example, is a well-established neuropsychological indicator of brain dysfunction (Reitan, 1958, 1985, 1993, 2008) in part because it taps cognitive processes such as processing speed and executive functioning, which are often impaired in TBI and diminished in MDD. Because poor performances on the TMT are expressed as high scores (elapsed time in seconds), while poor performances on the RAVLT-S are expressed as low scores (fewer words recalled), it follows that a negative correlation would arise if the RAVLT-S operated as a general measure of brain dysfunction as opposed to an explicit measure of verbal learning and memory. This contingency would also explain the observed relationships between the RAVLT-S and WAIS measures, given that verbal abilities tend to be better preserved than processing speed and executive functioning in both TBI and MDD. While this possibility departs from our original hypotheses, and is tempered by the inclusion of the DEP group in our analyses (which performed poorer than expected, as described above), it provides tentative support for the construct validity of the RAVLT-S as a general measure of brain dysfunction and warrants further investigation.

Another consideration is that although factor analytic studies indicate that the RAVLT loads primarily with other verbal memory tests such as the WMS, there is also evidence that the RAVLT may measure a construct that is not singularly verbal, and that memory variables that include the RAVLT load together regardless of whether they are verbal or non-verbal in nature (see Spreen & Strauss, 1998). Thus, if we accept RAVLT-S factor-based scores as reflective of broader measures of memory functioning, the correlations observed here make more sense. This assumption, however, erodes support for its construct validity as a measure of verbal learning and memory. In this regard, the results of our correlational analyses provide limited and conditional support for the construct validity of the RAVLT-S. As summarized here, the expected patterns regarding the strength and direction of correlations were not neatly borne out. An important consideration, however, is that our analyses did not include data derived from the normal population, and may thus reflect associations unique to the clinical samples evaluated. While both the TBI and DEP groups were well-represented in the RAVLT-S data, for example, data for the remaining neuropsychological variables were overwhelmingly derived from TBI patients. Given the heterogeneous nature of neurological impairment common to TBI, the results of this portion of our study should be evaluated conservatively.

A limitation of this study is the archival nature of the data analyzed and the potential for selection bias. The fact that participants were all native Puerto Ricans selected from a series of consecutive cases referred to a neuropsychology consult service in a hospital-based setting may limit generalizability of findings to other Spanish speaking individuals (e.g., Mexican Americans living in the U.S.), including those with

different neurological conditions. Another significant limitation of this study is the comparison of data derived from a Spanish-language instrument to English-language norms. Although this approach is generally discouraged (AERA et al., 1999), it is a function of the preliminary nature of our analyses (results of which must be interpreted with caution). A potential solution to this problem is to re-evaluate these data relative to Hispanic norms such as those derived by Ponton et al. (1996). However, fidelity to the English language version of the RAVLT was deliberately reduced as part of the development of the WHO/UCLA AVLT, through the use of a standardized lexicon of 250 universally familiar concepts (Maj et al., 1993). This represents a significant departure from traditional adaptations of the RAVLT, and may alter the meaning and relationships of scores as they relate to Spanish language measures derived from the traditional English version. Similar subtle differences have been observed in the interpretation of scores, for example, between other verbal list-learning tasks such as the CVLT (which incorporates a substantial semantic component) and the RAVLT (Spren & Strauss, 1998). By relying on the norms derived by Schmidt (1996), this study may also be susceptible to construct irrelevant variance, as may arise as a result of differences in cognitive equivalence between the two languages. Given the results of prior research, however, which failed to identify significant differences between English and Spanish RAVLT scores (Miranda & Valencia, 1997), the information provided by this study may be quite useful. This study also does not address complex issues related to bilingualism or linguistic cross-pollination that may affect individual performance on the RAVLT-S. Future analyses evaluating the relationship between RAVLT-S performance and elapsed time between injury and assessment may also be illuminating.

In conclusion, validity is the most fundamental consideration in the development and evaluation of tests, and the legitimate interpretation of test scores requires a sound scientific basis (AERA et al., 1999). However, the scientific literature on neuropsychological research addressing cultural issues is weak to virtually non-existent (Horton, 2008). While Hispanic and other ethnic minority psychologists have raised many legitimate objections regarding the use of psychological tests with minorities, these are often based on rational rather than empirical grounds (Reynolds, 2000). A major strength of this study is that it empirically evaluates the construct and criterion validity of the RAVLT-S in a clinical sample of traumatically brain-injured, Spanish speaking adults. This has been carried out in accordance with numerous recommendations outlined in the *Standards for Educational and Psychological Testing*, including the provision of internal consistency estimates and the use of various sources of data to help illuminate different aspects of validity evidence (AERA et al., 1999). Given clinicians' increased need for linguistically-diverse and empirically supported neuropsychological measures, this type of research helps to inform the psychometric and diagnostic literature related to the assessment of Hispanics with TBI. Overall, the results of this study provide initial, qualified support for the construct validity of the RAVLT-S when used to evaluate Spanish-speaking adults in clinical settings.

Appendix

Table 1

Original and Spanish Translation of the Rey Auditory Verbal Learning Test (RAVLT)

Word List

Original		Spanish translation ^a	
List A	List B	List A	List B
1. Drum	1. Desk	1. Tambor	1. Escritorio
2. Curtain	2. Ranger	2. Cortina	2. Maestro
3. Bell	3. Bird	3. Campana	3. Pájaro
4. Coffee	4. Shoe	4. Café	4. Zapato
5. School	5. Stove	5. Escuela	5. Estufa
6. Parent	6. Mountain	6. Padres	6. Montaña
7. Moon	7. Glasses	7. Luna	7. Piso
8. Garden	8. Towel	8. Jardín	8. Puente
9. Hat	9. Cloud	9. Sombrero	9. Toalla
10. Farmer	10. Boat	10. Granjero	10. Nube
11. Nose	11. Lamb	11. Nariz	11. Cordero
12. Turkey	12. Gun	12. Pavo	12. Pistola
13. Color	13. Pencil	13. Color	13. Pluma
14. House	14. Church	14. Casa	14. Iglesia
15. River	15. Fish	15. Río	15. Pez

^aAcevedo-Vargas, J.J. (2002), adapted from Miranda, J.P. & Valencia, R.R. (1997).

Table 2

Descriptive Statistics for all Neuropsychological Variables for Entire Sample

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
RAVLT-S Trial 1	106	4.74	1.56	.374	.336
RAVLT-S Trial 2	106	6.68	2.25	.113	-.893
RAVLT-S Trial 3	106	7.74	2.99	-.448	-.546
RAVLT-S Trial 4	106	8.55	3.43	-.296	-.994
RAVLT-S Trial 5	106	9.01	3.44	-.380	-.764
RAVLT-S Immediate	106	6.99	3.65	-.046	-.871
RAVLT-S Delayed	106	6.78	3.81	-.210	-.944
RAVLT-S Recognition	106	10.86	3.80	-1.159 ^a	.900
Phonemic Fluency Total	67	24.15	11.16	.790	1.241 ^b
WAIS III - POI	34	85.85	13.05	.291	-1.154 ^c
WAIS III - WMI	36	82.89	15.61	.633	.291
WAIS III - PSI	34	78.91	12.06	.581	-.230
WAIS III - VCI	36	85.25	12.63	.685	.011
WMS - LMI	70	13.44	6.99	.624	.657
WMS - LMD	70	9.94	7.09	.704	.689
Trails A	73	62.62	35.34	2.005 ^d	4.605 ^e
Trails B	74	155.41	105.05	1.630 ^f	2.150 ^g

Note. Skewness and kurtosis estimates after adjustment for outliers (Tabachnick & Fidell, 2007) are provided in specific notes a through e. Skewness and kurtosis estimates after application of 300 s cut-off (Reitan, 2008) are provided in specific notes f and g.

RAVLT-S = Rey Auditory Verbal Learning Test-Spanish; WAIS III = Wechsler Adult Intelligence Scale-Spanish, 3rd edition (*Escala de Inteligencia Wechsler para Adulto–Tercera Edición*; EIWA); POI = Perceptual Organization Index; WMI = Working Memory Index; PSI = Processing Speed Index; VCI = Verbal Comprehension Index; WMS = Wechsler Memory Scale, LMI = Long-term Memory Immediate, LMD = Long-term Memory Delayed.

^a -1.030; ^b -.496; ^c no adjustment necessary (kurtosis estimate within ± 2 standard error of kurtosis, SEK = .788, with no outliers); ^d .728; ^e -.347; ^f .921; ^g -.296.

Table 3

Demographic Data for Entire Sample by Diagnostic Group

Variable	TBI (<i>n</i> = 68)	DEP (<i>n</i> = 38)	% of Total
Age			
16-25	31 (45.6)	3 (7.9)	32.1
26-40	25 (36.8)	3 (7.9)	26.4
41-55	9 (13.2)	12 (31.6)	19.8
56-70	2 (2.9)	15 (39.5)	16.0
≥71	1 (1.5)	5 (13.2)	5.7
Total Mean (<i>SD</i>)	30.3 (12.0)	55.3 (15.2)	
Gender			
Female	27 (39.7)	26 (68.4)	50.0
Male	41 (60.3)	12 (31.6)	50.0
Education			
≤10	9 (13.2)	0 (0.0)	8.5
11-12	25 (36.8)	9 (23.7)	32.1
13-14	13 (19.1)	10 (26.3)	21.7
15-16	18 (26.5)	13 (34.2)	29.2
≥17	3 (4.4)	6 (15.8)	8.5
Total Mean (<i>SD</i>)	13.0 (3.1)	15 (2.5)	

Note. Values represent frequencies with percentages in parentheses, unless otherwise noted.

Table 4

Correlation Matrix for Rey Auditory Verbal Learning Test-Spanish List A Trial Pairs by Group Membership

Trials	Total ($N = 106$)			DEP ($n = 38$)			TBI ($n = 68$)		
	T1 + T3	T2 + T4	T3 + T5	T1 + T3	T2 + T4	T3 + T5	T1 + T3	T2 + T4	T3 + T5
T1 + T3	1.00			1.00			1.00		
T2 + T4	.86**	1.00		.83**	1.00		.87**	1.00	
T3 + T5	.90**	.89**	1.00	.92**	.86**	1.00	.91**	.90**	1.00

Note. DEP = Depression group; TBI = Traumatic Brain Injury group; T1 = Trial 1; T2 = Trial 2; T3 = Trial 3; T4 = Trial 4; T5 = Trial 5. T1 + T3 = combined Trial 1 and 3 data; T2 + T4 = combined Trial 2 and 4 data; T3 + T5 = combined Trial 3 and 5 data.

** $p < .01$ (2-tailed).

Table 5

Correlation Matrix for Rey Auditory Verbal Learning Test-Spanish List A Trial Pairs by

Gender for Entire Sample

Trials	Female (<i>n</i> = 53)			Male (<i>n</i> = 53)		
	T1 + T3	T2 + T4	T3 + T5	T1 + T3	T2 + T4	T3 + T5
T1 + T3	1.00			1.00		
T2 + T4	.84**	1.00		.88**	1.00	
T3 + T5	.87**	.87**	1.00	.94**	.90**	1.00

Note. T1 = Trial 1; T2 = Trial 2; T3 = Trial 3; T4 = Trial 4; T5 = Trial 5. T1 + T3 = combined Trial 1 and 3 data; T2 + T4 = combined Trial 2 and 4 data; T3 + T5 = combined Trial 3 and 5 data.

***p* < .01 (2-tailed).

Table 6

Descriptive Statistics and Comparisons for the Depression (DEP), Traumatic Brain Injury (TBI), and Normative Sample (NS) Groups

RAVLT Score	DEP		TBI		TBI vs. DEP		TBI vs. NS	
	Mean	SD	Mean	SD	F	p	t	p
Trial 1	-1.10	.75	-1.04	.964	2.60	.110	8.90	<.001
Trial 2	-1.02	.92	-1.40	1.17	.55	.460	9.89	<.001
Trial 3	-.88	1.24	-1.88	1.50	4.89	.029	10.30	<.001
Trial 4	-1.05	1.47	-1.97	1.85	1.96	.164	8.79	<.001
Trial 5	-.89	1.38	-2.26	1.98	6.78	.011	9.39	<.001
Immediate	-.78	1.15	-1.82	1.62	5.22	.024	9.22	<.001
Delayed	-.77	1.31	-1.72	1.55	4.89	.029	9.12	<.001
Recognition	-1.54	2.33	-3.19	3.52	3.38	.069	7.49	<.001

Notes. TBI vs. DEP are comparisons between TBI and DEP group age-corrected z scores using ANCOVA with education as a covariate. TBI vs. NS are comparisons between TBI and NS age-corrected z scores using single sample *t* tests.

Table 7

Classification Statistics of the Rey Auditory Verbal Learning Test-Spanish Learning Trials 1-5, Trials 1-5 Total, Immediate, Delayed and Recognition Trials

Score	SENS	SPEC	PPP	NPP	HR	Kappa	χ^2	<i>p</i>
T1	.418	.737	73.9	41.2	.53	.13	1.74	
T2	.500	.684	73.9	43.3	.57	.16	2.66	
T3	.588	.711	78.4	49.1	.63	.27	7.56	.006
T4	.544	.658	74.0	44.6	.58	.18	3.22	
T5	.559	.684	76.0	46.4	.60	.22	4.84	.028
Tot	.647	.711	80.0	52.9	.67	.33	11.09	< .001
Imm	.544	.763	80.4	48.3	.62	.27	8.16	.004
Del	.515	.763	79.5	46.8	.60	.24	6.65	.009
Rec	.603	.579	71.9	44.9	.59	.17	2.55	

Note. SENS = sensitivity; SPEC = specificity; PPP = positive predictive power; NPP = negative predictive power; HR = Hit Rate (Overall Correct Classification); χ^2 = Yates chi-square for 2 x 2 contingency table; *p* = probability for the chi-square test if *p* < .05. T1 = Trial 1; T2 = Trial 2; T3 = Trial 3; T4 = Trial 4; T5 = Trial 5; Tot = Trials 1-5 Total; Imm = Immediate Recall Trial; Del = Delayed Recall Trial; Rec = Recognition Trial. Cut-off scores based on Schmidt (1996) age-matched metanorms. RAVLT-S performances ≥ 1.5 SD below the mean classified as neurological dysfunction.

Table 8

Correlations Between Rey Auditory Verbal Learning Test-Spanish Raw Scores and Age and Education for the Combined Depression and Traumatic Brain Injury Samples

Variable	<i>n</i>	Age	Education
Trial 1	106	-.195*	.263**
Trial 2	106	-.123	.268**
Trial 3	106	.005	.378**
Trial 4	106	-.015	.374**
Trial 5	106	.041	.360**
Immediate	106	.028	.361**
Delayed	106	.006	.268**
Recognition	106	.053	.251**

* $p < .05$; ** $p < .01$.

Table 9

Rey Auditory Verbal Learning Test-Spanish Goodness-of-Fit Indices for One- and Two-Factor Models in the Entire Sample and in the Traumatic Brain Injury Group

Group/Model	Fit Indices				
	χ^2	<i>df</i>	CFI	RMSEA [90% CI]	AIC
TBI + DEP ^a					
VM ₁₋₈	46.85***	20	.97	.113 [.071, .156]	78.85
A ₁₋₅ R ₆₋₈	31.79*	19	.98	.080 [.023, .127]	65.79
TBI ^b					
VM ₁₋₈	40.68**	20	.96	.124 [.068, .179]	72.68
A ₁₋₅ R ₆₋₈	25.27	19	.99	.070 [.000, .136]	59.27

Note. VM₁₋₈ = one-factor model; A₁₋₅R₆₋₈ = hypothesized two-factor model; CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; AIC = Akaike information criterion; TBI = traumatic brain injury group; DEP = depression group. TBI + DEP = combined TBI and DEP groups.

^a Chi-square for independence model = 789.10; *df* = 28; *n* = 106.

^b Chi-square for independence model = 546.50; *df* = 28; *n* = 68.

p* = .033; *p* = .004; ****p* = .001.

Table 10

*Correlations Between Rey Auditory Verbal Learning Test-Spanish (RAVLT-S) Factor**Scores and Available Phonemic Fluency, WAIS-III measures, WMS measures, and TMT A & B Scores for Entire Sample*

Neuropsychological Tests	<i>n</i>	Mean (<i>SD</i>)	RAVLT-S Score Correlations	
			Acquisition	Retention
Phonemic Fluency	67	24.15 (11.16)	.512**†	.405**†
WAIS III - VCI	36	85.25 (12.63)	.469**†	.367*
WAIS III - WMI	36	82.89 (15.61)	.542**†	.354*
WAIS III - POI	34	85.85 (13.05)	.600**†	.543**†
WAIS III - PSI	34	78.91 (12.06)	.706**†	.633**†
WMS – LMI	70	13.44 (7.00)	.644**†	.579**†
WMS – LMD	70	9.94 (7.09)	.554**†	.555**†
Trails A	73	62.62 (35.34)	-.504**†	-.490**†
Trails B	74	155.41(105.05)	-.607**†	-.545**†

Note. Values required for statistical significance after Bonferroni correction: $p = .025$ for Phonemic Fluency, $p = .006$ for WAIS indices, $p = .012$ for WMS subtests, and $p = .012$ for Trails A & B. WAIS III = Wechsler Adult Intelligence Scale-Spanish, 3rd edition (*Escala de Inteligencia Wechsler para Adulto–Tercera Edición*; EIWA), VCI = Verbal Comprehension Index, WMI = Working Memory Index, POI = Perceptual Organization Index, PSI = Processing Speed Index; WMS = Wechsler Memory Scale, LMI = Long-term Memory Immediate, LMD = Long-term Memory Delayed.

* $p < 0.05$ (2-tailed); ** $p < 0.01$ (2-tailed); † $p < 0.05$ after Bonferroni correction.

References

- Acevedo-Vargas, J. (2002). *Estudio de validez concurrente y de contenido de la prueba de Aprendizaje Auditiva-Verbal Rey (RAVLT)*. (PhD, Universidad Carlos Albizu).
- Adams, J. H., Graham, D. I., Murray, L. S., & Scott, G. (1982). Diffuse axonal injury due to nonmissile head injury in humans: An analysis of 45 cases. *Annals of Neurology*, *12*(6), 557-563. doi:10.1002/ana.410120610
- AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716. doi:10.1109/TAC.1974.1100705
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, N.J: Prentice Hall.
- APA. (2000). *Diagnostic and statistical manual of mental disorders : DSM-IV-TR* (4th , text revision 2000 ed.). Washington, DC: American Psychiatric Association.
- APA. (2003). Guidelines on multicultural education, training, research, practice, and organizational change for psychologists. *American Psychologist*, *58*(5), 377-402. doi:10.1037/0003-066X.58.5.377
- APA. (2010). *Ethical principles of psychologists and code of conduct with the 2010 amendments*. <http://www.apa.org/ethics/code/index.aspx>
- APA. (2012). *Crossroads: The psychology of immigration in the new century. report of the APA presidential task force on immigration executive summary*. <http://www.apa.org/topics/immigration/executive-summary.pdf>

- Arango-Lasprilla, J., Rosenthal, M., DeLuca, J., Cifu, D. X., Hanks, R., & Komaroff, E. (2007). Functional outcomes from inpatient rehabilitation after traumatic brain injury: How do hispanics fare? [Abstract]. *Archives of Physical Medicine and Rehabilitation*, 88(1) 11-18.
- Arbuckle, J. L. (2006). *Amos (version 7.0)*. Chicago: SPSS.
- Ardila, A., Rosselli, M., & Puente, A. E. (1994). *Neuropsychological evaluation of the spanish speaker*. New York, NY US: Plenum Press.
- Areán, P. A., & Gallagher-Thompson, D. (1996). Issues and recommendations for the recruitment and retention of older ethnic minority adults into clinical research. *Journal of Consulting and Clinical Psychology*, 64(5), 875-880. doi:10.1037/0022-006X.64.5.875
- Army Individual Test Battery. (1944). *Manual of directions and scoring*. Washington, DC: War Department, Adjutant General's Office.
- Arnett, J. J. (2008). The neglected 95%: Why american psychology needs to become less american. *American Psychologist*, 63(7), 602-614. doi:10.1037/0003-066X.63.7.602
- Artiles, A. J., Trent, S. C., & Kuan, L. (1997). Learning disabilities empirical research on ethnic minority students: An analysis of 22 years of studies published in selected refereed journals. *Learning Disabilities Research & Practice*, 12(2), 82-91.
- Artiola, i. F., Heaton, R. K., & Hermsillo, D. (1998). Neuropsychological comparisons of spanish-speaking participants from the U.S.–Mexico border region versus spain. *Journal of the International Neuropsychological Society*, 4(4), 363-379.

- Baños, J. H., Elliott, T. R., & Schmitt, M. (2005). Factor structure of the rey auditory verbal learning test in adults with spinal cord injury. *Rehabilitation Psychology, 50*(4), 375-380. doi:10.1037/0090-5550.50.4.375
- Barr, W. B., Bender, H. A., Morrison, C., Cruz-Laureano, D., Vazquez, B., & Kuzniecky, R. (2009). Diagnostic validity of a neuropsychological test battery for hispanic patients with epilepsy. *Epilepsy & Behavior, 16*(3), 479-483. doi:10.1016/j.yebeh.2009.08.030
- Bello, D. T., Allen, D. N., & Mayfield, J. (2008). Sensitivity of the children's category test level 2 to brain dysfunction. *Archives of Clinical Neuropsychology, 23*(3), 329-339. doi:10.1016/j.acn.2007.12.002
- Bender, H. A., Cole, J. R., Aponte-Samalot, M., Cruz-Laureano, D., Myers, L., Vazquez, B. R., & Barr, W. B. (2009). Construct validity of the neuropsychological screening battery for hispanics (NeSBHIS) in a neurological sample. *Journal of the International Neuropsychological Society, 15*(2), 217-224. doi:10.1017/S1355617709090250
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.
- Benton, A. L., & Hamsher, K. (1989). *Multilingual aphasia examination manual of instructions* (2nd ed.). Iowa City: AJA Associates.
- Biever, J. L., Castaño, M. T., de, I. F., González, C., Servín-López, S., Sprowls, C., & Tripp, C. G. (2002). The role of language in training psychologists to work with

- hispanic clients. *Professional Psychology: Research and Practice*, 33(3), 330-336.
doi:10.1037/0735-7028.33.3.330
- Binet, A., & Simon, T. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, 11, 191-192-336.
- Binet, A., & Simon, T. (1916). Upon the necessity of establishing a scientific diagnosis of inferior states of intelligence. (L'année psych., 1905, pp. 163-191). In E. S. Kite (Ed.), *The development of intelligence in children (the binet-simon scale)*. (pp. 9-36). Baltimore, MD US: Williams & Wilkins Co. doi:10.1037/11069-001
- Blaney, P. H., & Millon, T. (2009). *Oxford textbook of psychopathology* (2nd ed.). New York: Oxford University Press.
- Bohlhalter, S., Abela, E., Weniger, D., & Weder, B. (2009). Impaired verbal memory in parkinson disease: Relationship to prefrontal dysfunction and somatosensory discrimination. *Behavioral and Brain Functions*, 5 doi:10.1186/1744-9081-5-49
- Boring, E. G. (1923). Intelligence as the tests test it. *The New Republic*, 35-36-37.
- Brown, A. W., Elovic, E. P., Kothari, S., Flanagan, S. R., & Kwasnica, C. (2008). Congenital and acquired brain injury. 1. epidemiology, pathophysiology, prognostication, innovative treatments, and prevention. *Archives of Physical Medicine and Rehabilitation*, 89(3, Supplement 1), S3-S8.
doi:10.1016/j.apmr.2007.12.001
- Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since bias in mental testing. *School Psychology Quarterly*, 14(3), 208-238.
doi:10.1037/h0089007

- Callahan, C. D., & Johnstone, B. (1994). The clinical utility of the rey auditory-verbal learning test in medical rehabilitation. *Journal of Clinical Psychology in Medical Settings, 1*(3), 261-268. doi:10.1007/BF01989627
- Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31*(2), 141-154. doi:10.1037/0735-7028.31.2.141
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY US: Cambridge University Press.
- Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, P. L. Harrison, D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues*. (pp. 122-130). New York, NY US: Guilford Press.
- Casscells, S. W. (2007). *Consolidation of traumatic brain injury initiatives in the department of defense*. (). Washington, DC: United States Department of Defense Health Affairs Memorandum Appendix 1, dated March 23, 2007.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*(3), 153-193. doi:10.1037/h0059973
- CDC. (2011). *Surveillance for traumatic brain injury--related deaths --- united states, 1997--2007*.
http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6005a1.htm?s_cid=ss6005a1_w
- Curtiss, G., Vanderploeg, R. D., Spencer, J., & Salazar, A. M. (2001). Patterns of verbal learning and memory in traumatic brain injury. *Journal of the International Neuropsychological Society, 7*(5), 574-585. doi:10.1017/S1355617701755051

- De la Plata, C. M., Arango-Lasprilla, J. C., Alegret, M., Moreno, A., Tárraga, L., Lara, M., . . . Cullum, C. M. (2009). Item analysis of three spanish naming tests: A cross-cultural investigation. *NeuroRehabilitation, 24*(1), 75-85.
- Del Ser Quijano, T., Garcia de Yebenes, M. J., Sanchez, F., Frades Payo, B., Rodriguez Laso, A., Bartolome Martinez, M. P., & Otero Puime, A. (2004). [Cognitive assessment in the elderly. normative data of a spanish population sample older than 70 years]. [Evaluacion cognitiva del anciano. Datos normativos de una muestra poblacional espanola de mas de 70 anos.] *Med Clin (Barc), 122*(19), 727-740.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California verbal learning test* (2nd ed.). San Antonio, TX: Psychological Corporation.
- DHHS. (2001). *Mental health: Culture, race, and ethnicity, A supplement to mental health: A report of the surgeon general*.
<http://www.surgeongeneral.gov/library/mentalhealth/cre/>
- Donders, J., & Nesbit-Greene, K. (2004). Predictors of neuropsychological test performance after pediatric traumatic brain injury. *Assessment, 11*(4), 275-284.
doi:10.1177/1073191104268914
- Echemendia, R. J., & Harris, J. G. (2004). Neuropsychological test use with Hispanic/Latino populations in the united states: Part II of a national survey. *Applied Neuropsychology, 11*(1), 4-12. doi:10.1207/s15324826an1101_2
- Echemendia, R. J., Harris, J. G., Congett, S. M., Diaz, M. L., & Puente, A. E. (1997). Neuropsychological training and practices with hispanics: A national survey. *The Clinical Neuropsychologist, 11*(3), 229-243. doi:10.1080/13854049708400451

- Erickson, R. C., & Scott, M. L. (1977). Clinical memory testing: A review. *Psychological Bulletin*, 84(6), 1130-1149. doi:10.1037/0033-2909.84.6.1130
- Estévez-González, A., Kulisevsky, J., Boltes, A., Otermín, P., & García-Sánchez, C. (2003). Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of alzheimer's disease: Comparison with mild cognitive impairment and normal aging. *International Journal of Geriatric Psychiatry*, 18(11), 1021-1028. doi:10.1002/gps.1010
- Faul, M., Xu, L., Wald, M. M. & Coronado, V. (2010). *Traumatic brain injury in the united states: Emergency department visits, hospitalizations and deaths, 2002-2006*. <http://www.cdc.gov/TraumaticBrainInjury/statistics.html>
- Fernández, A. L., & Marcopulos, B. A. (2008). A comparison of normative data for the trail making test from several countries: Equivalence of norms and considerations for interpretation. *Scandinavian Journal of Psychology*, 49(3), 239-246. doi:10.1111/j.1467-9450.2008.00637.x
- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). Hoboken, N.J: John Wiley & Sons.
- Franzen, M. D. (2000). *Reliability and validity in neuropsychological assessment* (2nd ed.). New York: Kluwer Academic/Plenum Publishers.
- Gardner, H. (1983). *Frames of mind : The theory of multiple intelligences*. New York: Basic Books.
- Gardner, H. (1993). *Multiple intelligences : The theory in practice*. New York, NY: Basic Books.

- Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York, NY US: Basic Books.
- Gary, K. W., Arango-Lasprilla, J., & Stevens, L. F. (2009). Do racial/ethnic differences exist in post-injury outcomes after TBI? A comprehensive review of the literature. *Brain Injury, 23*(10), 775-789. doi:10.1080/02699050903200563
- Gennarelli, T. A., Thibault, L. E., & Adams, J. H. (1982). Diffuse axonal injury and traumatic coma in the primate. *Annals of Neurology, 12*(6), 564-574.
- Golden, C. J., Hammeke, T. A., & Purisch, A. D. (1980). *The luria-nebraska neuropsychological battery : Manual*. Los Angeles, Calif. (12031 Wilshire Blvd., Los Angeles 90025): Western Psychological Services.
- Gómez-Pérez, E., & Ostrosky-Solís, F. (2006). Attention and memory evaluation across the life span: Heterogeneous effects of age and education. *Journal of Clinical and Experimental Neuropsychology, 28*(4), 477-494. doi:10.1080/13803390590949296
- Greenfield, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist, 52*(10), 1115-1124. doi:10.1037/0003-066X.52.10.1115
- Greenwald, B. D., Burnett, D. M., & Miller, M. A. (2003). Congenital and acquired brain injury. 1. brain injury: Epidemiology and pathophysiology. *Archives of Physical Medicine and Rehabilitation, 84*(3), S3-S7. doi:10.1053/apmr.2003.50052
- Hancock, T. U. (2005). Cultural competence in the assessment of poor mexican families in the rural southeastern united states. *Child Welfare: Journal of Policy, Practice, and Program, 84*(5), 689-711.

- Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, *47*(9), 1083-1101.
doi:10.1037/0003-066X.47.9.1083
- Herrnstein, R. J., & Murray, C. A. (1994). *The bell curve : Intelligence and class structure in american life*. New York: Free Press.
- Horn, J. L. (1989). Cognitive diversity: A framework of learning. In P. L. Ackerman, R. J. Sternberg, R. Glaser, P. L. Ackerman, R. J. Sternberg & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research*. (pp. 61-116). New York, NY US: W H Freeman/Times Books/ Henry Holt & Co.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253-270. doi:10.1037/h0023816
- Horton, A. M., Jr.(Ed.), & Wedding, D. (.). (2008). In Horton A. M., Jr., Wedding D. (Eds.), *The neuropsychology handbook (3rd ed.)*. New York, NY US: Springer Publishing Co.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling-a Multidisciplinary Journal*, *6*(1), 1-55. doi:10.1080/10705519909540118
- Huey, S. J., Jr., & Polo, A. J. (2008). Evidence-based psychosocial treatments for ethnic minority youth. *Journal of Clinical Child and Adolescent Psychology*, *37*(1), 262-301. doi:10.1080/15374410701820174

- Jacobs, M. L., & Donders, J. (2008). Performance discrepancies on the california verbal learning test--second edition (CVLT-II) after traumatic brain injury. *Archives of Clinical Neuropsychology*, 23(1), 113-118. doi:10.1016/j.acn.2007.09.003
- Jacova, C., Peters, K. R., Beattie, B. L., Wong, E., Riddehough, A., Foti, D., . . . Feldman, H. H. (2008). Cognitive impairment no dementia--neuropsychological and neuroimaging characterization of an amnesic subgroup. *Dementia and Geriatric Cognitive Disorders*, 25(3), 238-247. doi:10.1159/000115848
- Jafari, Z., Moritz, P. S., Zandi, T., Kamrani, A. A., & Malyeri, S. (2010). Psychometric properties of persian version of the rey auditory-verbal learning test (RAVLT) among the elderly. *Iranian Journal of Psychiatry and Clinical Psychology*, 16(1), 56-64.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International Inc.
- Judd, T., Capetillo, D., Carrión-Baralt, J., Mármol, L. M., San Miguel-Montes, L., Navarrete, M. G., . . . Valdés, J. (2009). Professional considerations for improving the neuropsychological evaluation of hispanics: A national academy of neuropsychology education paper. *Archives of Clinical Neuropsychology*, 24(2), 127-135.
- Kaufman, A. S., & Kaufman, J. C. (2001). Emotional intelligence as an aspect of general intelligence: What would david wechsler say? *Emotion*, 1(3), 258-264. doi:10.1037/1528-3542.1.3.258

- Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10*(3), 333-351. doi:10.1207/S15328007SEM1003_1
- Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. *Psychological Assessment, 5*(4), 395-399. doi:10.1037/1040-3590.5.4.395
- Kline, R. B. (2005). *Principles and practice of structural equation modeling (2nd ed.)*. New York, NY US: Guilford Press.
- Knight, R. G., McMahon, J., Green, T. J., & Skeaff, C. M. (2006). Regression equations for predicting scores of persons over 65 on the rey auditory verbal learning test, the mini-mental state examination, the trail making test and semantic fluency measures. *British Journal of Clinical Psychology, 45*(3), 393-402. doi:10.1348/014466505X68032
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*(1), 79-86. doi:10.1214/aoms/1177729694
- Leòn-Carriòn, J. (1989). Trail making test scores for normal children: Normative data from spain. *Perceptual and Motor Skills, 68*(2), 627-630.
- Lezak, M. D. (1983). *Neuropsychological assessment (2nd ed.)*. New York: Oxford University Press.
- Lezak, M. D. (2004). *Neuropsychological assessment (4th ed.)*. New York: Oxford University Press.

- Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV assessment*. Hoboken, N.J: Wiley.
- Maj, M., D'Elia, L., Satz, P., & Janssen, R. S. (1993). Evaluation of two new neuropsychological tests designed to minimize cultural bias in the assessment of HIV-1 seropositive persons: A WHO study. *Archives of Clinical Neuropsychology*, 8(2), 123-135. doi:10.1016/0887-6177(93)90030-5
- Malloy-Diniz, L., Lasmar, V. A. P., Gazinelli, L. d. S. R., Fuentes, D., & Salgado, J. V. (2007). The rey auditory-verbal learning test: Applicability for the brazilian elderly population. *Revista Brasileira De Psiquiatria*, 29(4), 324-329. doi:10.1590/S1516-44462006005000053
- Mash, E. J., & Barkley, R. A. (2006). *Treatment of childhood disorders* (3rd ed.). New York: Guilford Press.
- Mash, E. J., & Barkley, R. A. (2007). *Assessment of childhood disorders* (4th ed.). New York: Guilford Press.
- Messinis, L., Tsakona, I., Malefaki, S., & Papathanasopoulos, P. (2007). Normative data and discriminant validity of rey's verbal learning test for the greek adult population. *Archives of Clinical Neuropsychology*, 22(6), 739-752. doi:10.1016/j.acn.2007.06.002
- Miranda, J. P., Jr. (1996). *A spanish translation of the english version of the rey auditory-verbal learning test: Implications for the assessment of mexican-american males*. ProQuest Information & Learning. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 56(10-)

- Miranda, J. P., Jr., & Valencia, R. R. (1997). English and spanish versions of a memory test: Word-length effects versus spoken-duration effects. *Hispanic Journal of Behavioral Sciences, 19*(2), 171-181. doi:10.1177/07399863970192005
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing : Principles and applications* (6th ed.). Upper Saddle River, N.J: Pearson/Prentice Hall.
- Neisser, U. (1979). The concept of intelligence. *Intelligence, 3*(3), 217-227.
doi:10.1016/0160-2896(79)90018-7
- Neisser, U., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77-101. doi:10.1037/0003-066X.51.2.77
- Padilla, A. M., & Borsato, G. N. (2008). Issues in culturally appropriate psychoeducational assessment. In L. A. Suzuki, J. G. Ponterotto, L. A. Suzuki & J. G. Ponterotto (Eds.), *Handbook of multicultural assessment: Clinical, psychological, and educational applications*. (pp. 5-21). San Francisco, CA US: Jossey-Bass.
- Pena-Casanova, J., Quinones-Ubeda, S., Quintana-Aparicio, M., Aguilar, M., Badenes, D., Molinuevo, J. L., . . . NEURONORMA Study Team. (2009). Spanish multicenter normative studies (NEURONORMA project): Norms for verbal span, visuospatial span, letter and number sequencing, trail making test, and symbol digit modalities test. *Arch Clin Neuropsychol, 24*(4), 321-341.
- Perianez, J. A., Rios-Lago, M., Rodriguez-Sanchez, J. M., Adrover-Roig, D., Sanchez-Cubillo, I., Crespo-Facorro, B., . . . Barcelo, F. (2007). Trail making test in

- traumatic brain injury, schizophrenia, and normal ageing: Sample comparisons and normative data. *Arch Clin Neuropsychol*, 22(4), 433-447.
- Pew Hispanic Center. (2008). *Statistical portrait of hispanics in the united states, 2006*.
<http://pewhispanic.org/factsheets/factsheet.php?FactsheetID=58>
- Pontón, M. O., Satz, P., Herrera, L., & Ortiz, F. (1996). Normative data stratified by age and education for the neuropsychological screening battery for hispanics (NeSBHIS): Initial report. *Journal of the International Neuropsychological Society*, 2(2), 96-104. doi:10.1017/S1355617700000941
- Ponton, M. O., Gonzalez, J. J., Hernandez, I., Herrera, L., & Higareda, I. (2000). Factor analysis of the neuropsychological screening battery for hispanics (NeSBHIS). *Applied Neuropsychology*, 7(1), 32-39. doi:10.1207/S15324826AN0701_5
- Ponton, M. O., Satz, P., Herrera, L., Ortiz, F., Urrutia, C. P., Young, R., . . . Namerow, N. (1996). Normative data stratified by age and education for the neuropsychological screening battery for hispanics (NeSBHIS): Initial report. *Journal of the International Neuropsychological Society: JINS*, 2(2), 96-104.
- Prifitera, A., & Saklofske, D. H. (1998). *WISC-III clinical use and interpretation : Scientist-practitioner perspectives*. San Diego: Academic Press.
- Puente, A. E., & Ardila, A. (2000). Neuropsychological assessment of hispanics. In E. Fletcher-Janzen, T. L. Strickland & C. R. Reynolds (Eds.), *Handbook of cross-cultural neuropsychology* (pp. 87). New York: KluwerAcademic/Plenum Publishers.
- Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills*, 8, 271-276. doi:10.2466/PMS.8.7.271-276

- Reitan, R. M. (2008). *Manual for administration and scoring: Trail making test*. Dallas, TX: The Neuropsychology Center.
- Reitan, R. M., & Wolfson, D. (1985). *The halstead-reitan neuropsychological test battery: Theory and clinical interpretation*. Tucson, AZ: Neuropsychology Press.
- Reitan, R. M., & Wolfson, D. (1986). *Traumatic brain injury*. Tucson, AZ: Neuropsychology Press.
- Reitan, R. M., & Wolfson, D. (1993). *The halstead-reitan neuropsychological test battery : Theory and clinical interpretation, 2nd ed.* Tucson, AZ: Neuropsychology Press.
- Reitan, R. M., & Wolfson, D. (1995). Influence of age and education on neuropsychological test results. *Clinical Neuropsychologist*, 9(2), 151-158.
doi:10.1080/13854049508401597
- Rey, A. (1958). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.
- Rey, G. J. (1990). *Multilingual aphasia examination-spanish development and normative data*. ProQuest Information & Learning. *Dissertation Abstracts International*, 50(12-)
- Rey, G. J., Feldman, E., Hernandez, D., Levin, B. E., Rivas-Vazquez, R., Nedd, K. J., & Benton, A. L. (2001). Application of the multilingual aphasia examination-spanish in the evaluation of hispanic patients post closed-head trauma. *The Clinical Neuropsychologist*, 15(1), 13-18. doi:10.1076/clin.15.1.13.1911
- Reynolds, C. R. (2000). Methods for detecting and evaluating cultural bias in neuropsychological tests. In E. Fletcher-Janzen, T. L. Strickland, C. R. Reynolds, E. Fletcher-Janzen, T. L. Strickland & C. R. Reynolds (Eds.), *Handbook of cross-*

- cultural neuropsychology*. (pp. 249-285). Dordrecht Netherlands: Kluwer Academic Publishers.
- Ricker, J. H. (2004). *Differential diagnosis in adult neuropsychological assessment*. New York: Springer Publishing Co.
- Rosenthal, M., & Ricker, J. (2000). Traumatic brain injury. In R. G. Frank, T. R. Elliott, R. G. Frank & T. R. Elliott (Eds.), *Handbook of rehabilitation psychology*. (pp. 49-74). Washington, DC US: American Psychological Association. doi:10.1037/10361-003
- Ryan, J. J., Paolo, A. M., & Skrade, M. (1992). Rey auditory verbal learning test performance of a federal corrections sample with acquired immunodeficiency syndrome. *International Journal of Neuroscience*, *64*(1-4), 177-181. doi:10.3109/00207459209000543
- Ryan, J. J., Rosenberg, S. J., & Mittenberg, W. (1984). Factor analysis of the rey auditory-verbal learning test. *International Journal of Clinical Neuropsychology*, *6*(4), 239-241.
- Sanchez, G. I. (1932). Group differences and spanish-speaking children-a critical review. *Journal of Applied Psychology*, *16*(5), 549-558.
- Schmidt, M. (1996). *Rey auditory and verbal learning test: A handbook*. Los Angeles, CA: Western Psychological Services.
- Schoenberg, M. R., Dawson, K. A., Duff, K., Patton, D., Scott, J. G., & Adams, R. L. (2006). Test performance and classification statistics for the rey auditory verbal learning test in selected clinical samples. *Archives of Clinical Neuropsychology*, *21*(7), 693-703. doi:10.1016/j.acn.2006.06.010

- Segun, T. D. (2011). *Traumatic brain injury (TBI) - definition, epidemiology, pathophysiology*. <http://emedicine.medscape.com/article/326510-overview#aw2aab6b4>
- Serpell, R. (2000). Determinants of intelligence: Culture and intelligence. In A. E. Kazdin, & A. E. Kazdin (Eds.), *Encyclopedia of psychology, vol. 2*. (pp. 493-496). Washington, DC New York, NY USUS: American Psychological Association; Oxford University Press. doi:10.1037/10517-183
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379-423.
- Sharma, P., & Kerl, S. B. (2002). Suggestions for psychologists working with mexican american individuals and families in health care settings. *Rehabilitation Psychology*, 47(2), 230-239. doi:10.1037/0090-5550.47.2.230
- Shin, H. B., & Kominski, R. A. (2010). *Language use in the united states: 2007, american community survey reports, ACS-12*.
- Smith, G. E., Ivnik, R. J., & Lucas, J. (2008). Assessment techniques: Tests, test batteries, norms, and methodological approaches. In J. E. Morgan, J. H. Ricker, J. E. Morgan & J. H. Ricker (Eds.), *Textbook of clinical neuropsychology*. (pp. 38-57). New York, NY US: Psychology Press.
- Spearman, C. (1904). "General intelligence " objectively determined and measured. *American Journal of Psychology*, 15, 201-292. doi:10.2307/1412107
- Spren, O., & Strauss, E. (1998). *A compendium of neuropsychological tests :Administration, norms, and commentary* (2nd ed.). New York: Oxford University Press.

- Steiger, J. H. (1990). Structural model evaluation and modification - an interval estimation approach. *Multivariate Behavioral Research*, 25(2), 173-180. doi:10.1207/s15327906mbr2502_4
- Steinberg, B. A., Bieliauskas, L. A., Smith, G. E., Ivnik, R. J., & Malec, J. F. (2005). Mayo's older americans normative studies: Age- and IQ-adjusted norms for the auditory verbal learning test and the visual spatial learning test. *The Clinical Neuropsychologist*, 19(3-4), 464-523. doi:10.1080/13854040590945193
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY US: Cambridge University Press.
- Sternberg, R. J. (2003). A broad view of intelligence: The theory of successful intelligence. *Consulting Psychology Journal: Practice and Research*, 55(3), 139-154. doi:10.1037/1061-4087.55.3.139
- Sternberg, R. J., & Grigorenko, E. L. (2008). Ability testing across cultures. In L. A. Suzuki, J. G. Ponterotto, L. A. Suzuki & J. G. Ponterotto (Eds.), *Handbook of multicultural assessment: Clinical, psychological, and educational applications*. (pp. 449-470). San Francisco, CA US: Jossey-Bass.
- Sternberg, R. J., Grigorenko, E. L., & Kidd, K. K. (2005). Intelligence, race, and genetics. *American Psychologist*, 60(1), 46-59. doi:10.1037/0003-066X.60.1.46
- Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist*, 54(12), 1070-1077. doi:10.1037/0003-066X.54.12.1070
- Sue, S. (2009). Ethnic minority psychology: Struggles and triumphs. *Cultural Diversity and Ethnic Minority Psychology*, 15(4), 409-415. doi:10.1037/a0017559

- Suleiman, L. P. (2003). Beyond cultural competence: Language access and latino civil rights. *Child Welfare: Journal of Policy, Practice, and Program*, 82(2), 185-200.
- Suzuki, L. A., & Ponterotto, J. G. (2008). *Handbook of multicultural assessment : Clinical, psychological, and educational applications* (3rd / Lisa A. Suzuki and Joseph G. Ponterotto, editors ed.). San Francisco: Jossey-Bass.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (5th ed.)*. Boston, MA: Allyn & Bacon/Pearson Education.
- Talley, J. L. (1986). Memory in learning disabled children: Digit span and the rey auditory verbal learning test. *Archives of Clinical Neuropsychology*, 1(4), 315-322.
doi:10.1016/0887-6177(86)90135-6
- Taylor, E. (. (1959). *Psychological appraisal of children with cerebral defects*. Cambridge: Published for the Commonwealth Fund by Harvard University Press.
- Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. *Lancet*, 2(7872), 81-84.
- Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the stanford revision and extension of the binet-simon intelligence scale*. Boston, MA US: Houghton, Mifflin and Company.
- U.S. Census Bureau. (2010). *2006-2010 american community survey*.
http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_10_5YR_B16001&prodType=table
- U.S. Census Bureau. (2011). *Newsroom: Facts for features & special editions: Facts for features — hispanic heritage month 2011: Sept. 15 - oct. 15*.

http://www.census.gov/newsroom/releases/archives/facts_for_features_special_editions/cb11-ff18.html

- Uomoto, J. M. (2004). The contribution of the neuropsychological evaluation to traumatic brain injury rehabilitation. In M. J. Ashley (Ed.), *Traumatic brain injury: Rehabilitative treatment and case management (2nd ed.)*. (pp. 581-611). Boca Raton, FL US: CRC Press.
- Vakil, E. (2005). The effect of moderate to severe traumatic brain injury (TBI) on different aspects of memory: A selective review. *Journal of Clinical and Experimental Neuropsychology*, 27(8), 977-1021. doi:10.1080/13803390490919245
- Vakil, E., & Blachstein, H. (1993). Rey auditory-verbal learning test: Structure analysis. *Journal of Clinical Psychology*, 49(6), 883-890. doi:10.1002/1097-4679(199311)49:6<883::AID-JCLP2270490616>3.0.CO;2-6
- Vakil, E., Greenstein, Y., & Blachstein, H. (2010). Normative data for composite scores for children and adults derived from the rey auditory verbal learning test. *The Clinical Neuropsychologist*, 24(4), 662-677. doi:10.1080/13854040903493522
- van den Burg, W., & Kingma, A. (1999). Performance of 225 dutch school children on rey's auditory verbal learning test (AVLT): Parallel test–retest reliabilities with an interval of 3 months and normative data. *Archives of Clinical Neuropsychology*, 14(6), 545-559. doi:10.1016/S0887-6177(98)00042-0
- Washington State Department of Health. (2009). *TBI: Prevalance, external causes and associated risk factors*. <http://www.doh.wa.gov/hsqa/ocrh/har/hcresrch.htm>

- Wechsler, D. (1944). The nature of intelligence. In D. Wechsler (Ed.), *The measurement of adult intelligence (3rd ed.)* (pp. 3-12). Baltimore, MD US: Williams & Wilkins Co. doi:10.1037/11329-001
- Wechsler, D. (1945). A standardized memory scale for clinical use. *The Journal of Psychology*, 19(1), 87-95. doi:10.1080/00223980.1945.9917223
- Wechsler, D. (2008). *Escala de inteligencia de wechsler para adultos - tercera edición* (3rd ed.). San Antonio, TX: Pearson.
- Williams, R. L. (1975). The BITCH-100: A culture-specific test. *Journal of Afro-American Issues*, 3, 103-104-116.

Curriculum Vitae

Cristobal Neblina
2405 Flower Spring Street
Las Vegas, NV 89134

Education in Progress

Fall 2009 Doctor of Philosophy in Clinical Psychology
Admission University of Nevada Las Vegas
 Neuropsychology emphasis
 Expected Completion Date: Spring 2015
 4.00 GPA (current)

Higher Education

May 2008 Bachelor of Arts in Psychology
 University of Nevada Las Vegas
 Summa Cum Laude
 4.00 GPA

May 2006 Associate of Arts
 College of Southern Nevada
 High Honors
 4.00 GPA

Honors, Awards and Memberships

2011 UNLV Summer Session Scholarship Recipient
2011 Graduate Student Member, Golden Key International Honour Society
2010 Graduate Student Member, American Psychological Association
2010 Graduate Student Member, Nevada Psychological Association
2009 Graduate Student Member, National Academy of Neuropsychology
2008 UNLV Outstanding Graduate Award
2008 Recognition, College of Liberal Arts Honors Convocation
2008 Member, Phi Kappa Phi National Academic Honor Society
2007-2008 Osher Re-Entry Scholarship Recipient
2007-2008 Dee Smith Endowment Scholarship Recipient
2006-2008 MGM Mirage-Hites Foundation Scholarship Recipient
2006-2008 UNLV Dean's Honor List
2006 Member, Psi Chi National Honor Society in Psychology
2005 CSN Faculty Recognition Award for Outstanding Student Achievement

Publications and Presentations

Neblina, C., San Miguel Montes, Liza E., Allen, D.N. (2011, March). *Criterion Validity of the Rey Auditory Verbal Learning Test-Spanish in Adults with Traumatic Brain Injury*. Poster presented at the 3rd Annual Conference of the American College of Professional Neuropsychology, Las Vegas, NV.

Neblina, C., Terranova, J., Mayfield, J., & Allen, D.N. (2010, February). *Longitudinal Stability of Neurocognitive Profiles in Children with Traumatic Brain Injury*. Poster presented at the 2nd Annual Conference of the American College of Professional Neuropsychology, Las Vegas, NV.

San Miguel Montes, L. E., Allen, D. N., Puente, A. E., & Neblina, C. (2010). Validity of the WISC-IV spanish for a clinically referred sample of hispanic children. *Psychological Assessment*, 22(2), 465-469.

Professional Experience

Fall 2010-Spring 2011 Teaching Assistant: Statistical Methods in Psychology

Fall 2009-Summer 2010 Graduate Assistant: Psychology 101 and UNLV
Neuropsychology Research Program