Journal of the International Neuropsychological Society (2009), **15**, 217–224.

doi:10.1017/S1355617709090250

Construct validity of the Neuropsychological Screening Battery for Hispanics (NeSBHIS) in a neurological sample

H. ALLISON BENDER,[1] JEFFREY R. COLE,[2] MYRELIS APONTE-SAMALOT,[1,3] DANIEL CRUZ-LAUREANO,[1,3] LORNA MYERS,[4] BLANCA R. VAZQUEZ,[1] AND WILLIAM B. BARR[1]

[1]Department of Neurology, Comprehensive Epilepsy Center, New York University Langone Medical Center, New York, New York
[2]Department of Neurology, Comprehensive Epilepsy Center, Columbia University Medical Center, New York, New York
[3]Ponce School of Medicine, Ponce, Puerto Rico
[4]Northeast Regional Epilepsy Group, New York, New York

(Received April 21, 2008; Final Revision October 17, 2008; Accepted October 20, 2008)

Abstract

Epidemiological studies suggest that the Hispanic population is at increased risk for neurological disorders. Yet, few assessment measures have been developed for, adapted to, or normalized with Spanish-speakers. The Neuropsychological Screening Battery for Hispanics (NeSBHIS) was developed to address the lack of resources available to this underserved community. Although the NeSBHIS possesses robust construct validity and clinical utility in a community-based sample, these properties remain largely untested in neurological populations. One hundred and twenty-seven Spanish-speaking Hispanic patients with confirmed epilepsy (mean age = 37.8, *SD* = 13.3) were evaluated using the NeSBHIS. All participants self-identified as "Hispanic" and immigrated from Spanish-speaking countries. Data were analyzed using confirmatory factor analysis with the *a priori* assumption that variables would load according to theoretical expectations reported by Pontón and colleagues (2000). The overall model fit indices were in the desired range: Comparative Fit Index = 0.936, Tucker Lewis Index = 0.915, RMSEA = 0.090, and SRMR = 0.069. All NeSBHIS subtests loaded significantly ($p < .001$) on their respective factors; the standardized loadings were high, ranging from 0.562 to 0.995, with the exception of Block Design (–0.308). Overall, findings suggest that the NeSBHIS has robust construct validity in a neurological sample. (*JINS*, 2009, *15*, 217–224.)

Keywords: Hispanics, Spanish, Latinos, Immigrants, Epilepsy, Psychometrics

INTRODUCTION

In an attempt to improve the standard of care provided to an increasingly diverse patient base, multiple investigators have evaluated differences in neuropsychological test performance among cultural and ethnic minorities (Ardila, 1995; Manly et al., 1998; Roberts & Hamsher, 1984; Wong, 2000). Yet, to date, comparatively few assessment measures have been developed for, adapted to, or normalized with historically underrepresented populations. The relative dearth of culturally fair tests is of particular concern for non-United States (U.S.) born persons of Hispanic heritage,[a] as this ethnic group is the most rapidly growing segment of the country's population, representing over half of all immigrants to the U.S. (Census, 2000). This population trend shows no signs of slowing; rather, the U.S. Census Bureau (2000) estimates that by 2050, nearly one in four Americans will be of Hispanic descent and over 50% of that population will be Spanish-speaking and non-U.S. born.

The Neuropsychological Screening Battery for Hispanics (NeSBHIS; Pontón et al., 1996) was specifically developed to address the fundamental lack of resources available to Spanish-speaking Hispanic individuals. Modeled after a battery first used by the World Health Organization (Maj et al., 1994b), the NeSBHIS was designed to assess multiple cognitive domains, including: language, memory, mental control, psychomotor speed, visuospatial functioning, and nonverbal reasoning. One of the primary advantages of this battery is that it is among the few to provide normative data stratified by age, gender, and education using a moderately large ($N = 300$) standardization sample of community-referred, Spanish-speaking Hispanics. Further investigation suggests that the NeSBHIS has a stable

[a]The term "Hispanic" does not refer to a homogenous ethnic group. Rather, the U.S. Census Bureau defines Hispanic as "a person of Mexican, Puerto Rican, Cuban, Central or South American, or other Spansh-speaking culture or origin, regardless of race."

Correspondence and reprint requests to: H. Allison Bender, Ph.D., New York University Langone Medical Center, Comprehensive Epilepsy Center, Neuropsychology Service, 223 East 34th Street, New York, NY 10016. E-mail: Heidi.Bender@nyumc.org

factor structure, indicating robust construct validity for use with Hispanic populations (Pontón et al., 2000). That is, the NeSBHIS adequately measures the putative neuropsychological domains that it was designed to assess within this population, yielding five distinct factors: (a) Language (as measured by the Escala de Inteligencia Wechsler para Adultos [EIWA] Digit Span subtest, Pontón-Satz Boston Naming Test, and the Controlled Oral Word Association Test); (b) Verbal Learning (World Health Organization – UCLA Auditory Verbal Learning Test final learning, short-delay free recall following a distracter, and 20-minute delayed recall trials); (c) Attention-Mental Control (EIWA Digit Symbol and Block Design subtests, as well as Color Trails 1 & 2); (d) Visuospatial (Rey-Osterrieth Complex Figure Test – Copy and Delayed Recall scores, as well as the Raven's Standard Progressive Matrices); and (e) Psychomotor (Pin Test).

While the NeSBHIS represents a significant advancement in Spanish-language neuropsychological assessment, concerns exist regarding the generalizability of this instrument to a more heterogeneous sample. To this end, Pontón and colleagues (1996; 2000) stress the need for revalidation among a more diverse population of Hispanics (e.g., monolingual *vs.* bilingual, elderly *vs.* young, higher *vs.* lesser educated). These authors further suggest validation of the NeSBHIS for use with clinical populations. The latter is of particular importance to clinical neuropsychologists and researchers, given the stringent exclusionary criteria (e.g., history of a neurological or psychiatric disorder, head trauma, and substance use) applied to the NeSBHIS normative sample. Although preliminary investigation suggests adequate clinical utility of the NeSBHIS in a neurological sample (Myers et al., 2002; 2003), its psychometric properties may lack applicability to the majority of patients seeking neuropsychological evaluation in hospital or clinic settings (e.g., patients with known neurological and/or psychiatric illness). Without comparable empirical validation, it is unclear as to whether NeSBHIS data collected in a neurological sample are reflective of a patient's true level of cognitive functioning or an artifact of a measure whose validity is unknown.

The aim of the present study was to examine the construct validity of the NeSBHIS in a neurological sample. To this end, we evaluated the extent to which the NeSBHIS factor structure, originally described by Pontón and colleagues (2000), remained stable in a large, demographically heterogeneous sample of Spanish-speaking patients with a confirmed diagnosis of epilepsy. To assure comparability, we administered the NeSBHIS in its entirety, with the exception of the Pin Test. Instead, the Grooved Pegboard (GP; Matthews, 1964), a more commonly administered measure of motor speed and visuomotor coordination (Rabin et al., 2005), was administered for its ability to evaluate lateralized brain damage in adults (Hanna-Pladdy et al., 2002).

METHODS

Participants

One hundred and twenty-seven Hispanic participants with epilepsy (32% male and 68% female) were recruited from the patient population receiving a neuropsychological evaluation at New York University Comprehensive Epilepsy Center, New York, between 2000 and 2006. The gender distribution of the studied sample is comparable to Pontón et al's (1996) standardization sample (e.g., 40% male and 60% female). All participants self-identified as "Hispanic," immigrated from a Spanish-speaking country, and requested Spanish-language assessment. A diagnosis of epilepsy, as defined by the International Classification of Epilepsies (1989), was required for study inclusion. Individuals with nonepileptic seizures, severe psychiatric disturbance, dementia, and/or probable developmental delay were excluded.

Participants ranged in age from 16 to 79 years, with a mean age of 37.8 ($SD = 13.3$). The average level of education was 11.9 years ($SD = 4.4$). Approximately 90% of the sample was right-handed. Regarding seizure-related characteristics, the mean age at seizure onset was nearly 20 years of age ($SD = 15.9$). The majority of the sample (82%) exhibited seizures of temporal lobe origin; 58% of the subjects had a left-sided, 29% a right-sided, and 13% a bitemporal seizure focus. The participants with seizures arising outside of the temporal lobe were classified as having either generalized (14%) or partial (17%) epilepsies. Of the participants who were receiving antiepileptic drug therapy, nearly 85% of the sample received treatment with polytherapy and 14.5% were prescribed monotherapy. One participant was unmedicated at the time of assessment. Most of the studied sample can be categorized as originating from Puerto Rico (38%), Central or South America (30%), The Dominican Republic (12%), and Mexico (5.5%). This distribution is consistent with the population trends among Hispanic immigrants residing in the Northeastern U.S.

Procedures

The present study was performed according to the policies of the Institutional Review Board (IRB) of New York University School of Medicine. Participants were tested individually in 1–2 sessions totaling approximately 1.5–2.0 hours. As previously stated, the NeSBHIS was administered in the manner described in Pontón et al.'s initial study (Pontón et al., 1996); however, the Grooved Pegboard was administered in place of the Pin Test.

Measures

The NeSBHIS was designed to evaluate the domains of language, memory functioning, visuospatial functioning, mental control, psychomotor functioning, and reasoning (Pontón et al., 1996). Most measures within this battery were adapted from versions used internationally by the World Health Organization (Maj et al., 1994). The NeSBHIS subtests below are arranged according to Pontón and colleagues' (2000) obtained factor structure in a community-referred sample.

Language

Pontón-Satz Boston Naming Test. The P-S BNT (Pontón et al., 1992) was administered as a measure of confrontation naming. The P-S BNT is the 30-item Spanish-language equivalent of the 60-item English-language Boston Naming Test (BNT; Kaplan et al., 1983). Culturally-loaded items (e.g., "pretzel" and "beaver") were excluded from this measure to ensure cultural relevance and appropriateness in Hispanic populations.

Controlled Oral Word Association Test. The COWAT (Benton & Hamsher, 1976) measures verbal fluency in English- and Spanish-speaking individuals. On this task, participants were given one minute to generate as a many words as possible beginning with a target letter (in this case, "F," "A," and "S").

Digit Span subtest from the Escala de Inteligencia Wechsler para Adultos (EIWA). This test (Wechsler et al., 1968) was administered as a measure of attention and working memory. EIWA Digit Span is identical in content to the Wechsler Adult Intelligence Scale–Revised (Wechsler, 1981). In order to yield identical data to those collected by Pontón et al. (1996; 2000), both sets of number strings were administered per trial, rather than the one set of digits suggested by the EIWA manual.

Verbal Learning

WHO-UCLA Auditory Verbal Learning Test (AVLT). This test (Maj et al., 1993) comprised three components of the NeSBHIS evaluation paradigm. Participants were read a serially presented list of 15 Spanish words and were asked to repeat as many as they could remember immediately following each presentation. Scores from the fifth and final learning trial (AVLT V), recall following a distraction condition (AVLT VII), and 20-minute delayed recall (AVLT VIII) were all evaluated.

Attention–Mental Control

EIWA Digit Symbol subtest. This test was used as a measure of attention and processing speed. Unlike its English language counterpart, the EIWA Digit Symbol task has six target symbols and a total of 110 possible responses (as compared to 93 on the WAIS-R / Wechsler Adult Intelligence Scale-III; Wechsler, 1981).

Color Trails 1 and 2. This test (D'Elia, 1994) was administered as a measure of sustained attention and set-shifting. CT1 consists of colored circles numbered 1–25; odd numbers are pink and the even numbers are yellow. Examinees were instructed to connect the circles in sequential order as quickly as possible. On CT2, all numbers are presented in both pink and yellow; examinees were asked to connect the circles in sequential order (while maintaining numerical sequence) and to simultaneously alternate colors.

Block Design. This subtest from the EIWA is based on a similarly designed WAIS-R visuoconstruction task. Although similar in structure, the EIWA Block Design has slightly different content and awards fewer time-dependent bonus points than the WAIS.

Visuospatial Functioning

Rey Osterrieth Complex Figure Test. The RCFT (Osterrieth, 1944) copy condition was administered as a measure of visuospatial reproduction. Participants were asked to copy a complex geometric figure with numerous embedded details. The Taylor scoring system was used to evaluate the accuracy of the designs produced (Taylor, 1959). The RCFT delayed recall condition was also administered as a measure of incidental visual memory.

Raven's Standard Progressive Matrices. The RSPM (Raven, 1993) was used to assess perceptual reasoning and the ability to reason by analogy, independent of language and education. The RSPM is considered one of the most widely administered measures of nonverbal reasoning and intelligence.

Psychomotor

Grooved Pegboard. This test (GP; Lafayette Instruments #32035) was administered in lieu of the Pin Test (Satz & D'Elia, 1989). The task required the participants to placed key-shaped pegs into a board consisting of 25 (5 x 5) keyhole shaped holes as quickly as possible. The procedure was completed with the participant's dominant, followed by their nondominant, hand. It is typically used as a measure of fine motor dexterity and visuomotor coordination (Strauss et al., 2006).

Data Analysis

The primary goal of the present study was to determine whether the NeSBHIS maintains its stable factor structure in a clinical sample. For this reason, the data were analyzed using confirmatory factor analysis (CFA) with the *a priori* assumption that the variables included in the analysis would load according to theoretical expectations previously outlined by Pontón and colleagues (1996; 2000). It was hypothesized that the following five factors (and their component variables) would emerge: Language (COWAT, P-S BNT, and Digit Span), Verbal Learning (AVLT V, AVLT VII, and AVLT VIII), Attention–Mental Control (CT1, CT2, Digit Symbol, and Block Design), Visuospatial (RCFT copy, RCFT delayed recall, and RSPM), and Psychomotor (GP dominant and GP nondominant) (see Figure 1).

All 15 indicators of latent factors, however, had missing data, ranging from 1.57% to 32.28% of the total sample size. Primary reasons for these missing data were interrupted testing sessions and patient inability or unwillingness to complete task demands. The latter was particularly true in cases of the Grooved Pegboard subtest, as several inpatients were unable to complete this task because of discomfort secondary to placement of an intravenous line.

In order to address the missing data, the multiple imputation (MI) technique was used to preserve the overall power, as well as to retain the variability among individuals (Graham et al., 2003; Graham & Schafer, 1999; Little & Rubin, 1986; Rubin, 1987). While simpler approaches are available

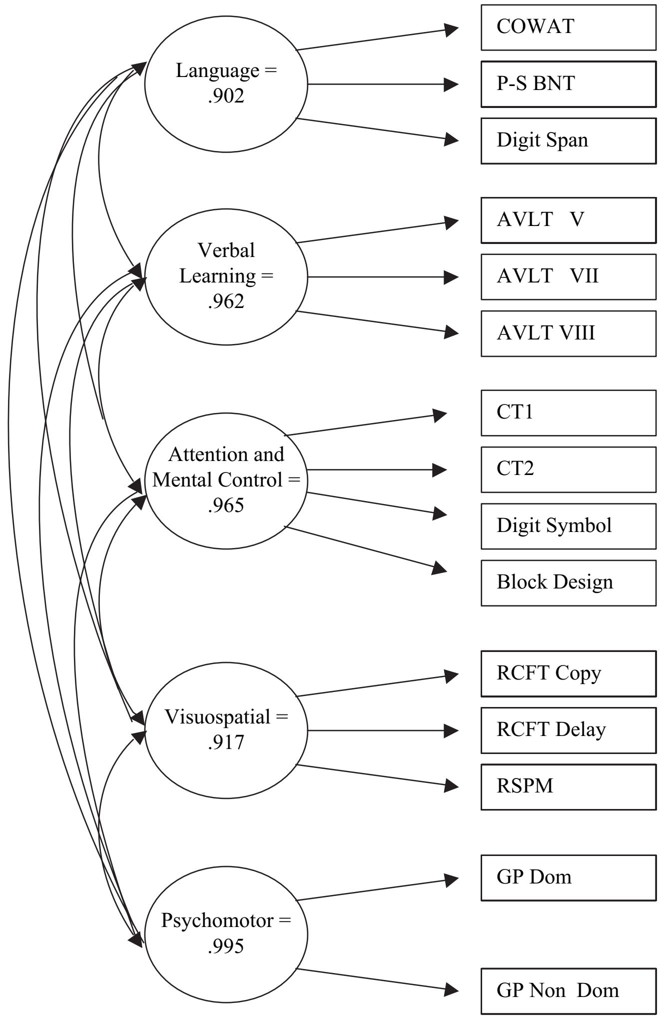

Fig. 1. Factor structure as suggested by Pontón et al (2000) and obtained factor determinacies (by domain).

(e.g., mean substitution and single imputation), they are known to alter the naturally occurring relationships between existing variables, and may bias the results (Graham et al., 1994; Wayman, 2003). Conceptually, in the MI approach, missing values are regressed on all possible explanatory variables; variables that later may be part of the main analysis or variables that assist in explaining the missingness. For the current data, 15 factor indicators were regressed on each other, in addition to demographic variables (e.g., education and age) and seizure-related variables (e.g., age at seizure onset, duration of seizures, lesion on MRI, and language dominance as determined by Wada testing). Beyond the prediction from existing explanatory variables, MI adds a random error term to each imputed value to account for imprecision of prediction and sample variability. This process is repeated several times, yielding several data sets where missing values are substituted with imputed values; five imputations are traditionally accepted as sufficient (Allison, 2002).

MI analysis was carried out in SAS 9.1 (SAS Institute, Cary, NC) with PROC MI. Descriptive statistics such as means, standard deviations, and correlations were obtained by averaging the corresponding values across five computations.

The CFA was subsequently performed on the imputed correlation matrix (see Table 1) using the specialized statistical software for latent variable analyses MPlus 4.2 (Muthén & Muthén, 2006). By program default, five proposed factors were set to correlate with each other and all error terms were independent and normally distributed.

RESULTS

As shown in Table 1, the correlation matrix reveals that many of the subtests were highly correlated amongst themselves; scales tapping on common underlying constructs (e.g., CT1 & CT2) had large correlation values.

Table 2 describes group performance on the NeSBHIS subtests within the studied sample. The highest mean scores were obtained on measures of nonverbal abstract reasoning, simple attention, and verbal fluency. The lowest scores were yielded on measures of mental scanning and tracking, confrontation naming, and memory.

Results of the CFA demonstrated that the proposed model (see Figure 1) fits the data well, providing support for the theoretical framework proposed by Pontón and colleagues

https://doi.org/10.1017/S1355617709090250 Published online by Cambridge University Press

Table 1. Correlation matrix for 15 NeSBHIS subtests based on MI data *(N* = 127)

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	***1.00***														
2	0.65	***1.00***													
3	0.40	0.39	***1.00***												
4	0.25	0.27	0.29	***1.00***											
5	0.27	0.26	0.29	0.80	***1.00***										
6	0.25	0.28	0.29	0.80	0.79	***1.00***									
7	–0.42	–0.46	–0.45	–0.42	–0.35	–0.36	***1.00***								
8	–0.41	–0.44	–0.46	–0.52	–0.45	–0.53	0.79	***1.00***							
9	0.40	0.42	0.54	0.40	0.35	0.33	–0.66	–0.68	***1.00***						
10	0.07	0.07	0.19	0.17	0.06	0.10	–0.30	–0.27	0.27	***1.00***					
11	0.24	0.14	0.29	0.33	0.24	0.26	–0.43	–0.50	0.56	0.21	***1.00***				
12	0.35	0.18	0.20	0.42	0.31	0.35	–0.37	–0.42	0.49	0.11	0.58	***1.00***			
13	0.42	0.28	0.40	0.36	0.26	0.28	–0.55	–0.58	0.63	0.45	0.58	0.49	***1.00***		
14	–0.24	–0.27	–0.33	–0.48	–0.39	–0.52	0.56	0.76	–0.54	–0.21	–0.57	–0.41	–0.52	***1.00***	
15	–0.21	–0.20	–0.35	–0.46	–0.40	–0.47	0.50	0.68	–0.55	–0.20	–0.54	–0.41	–0.49	0.79	***1.00***

Note. 1 = P-S BNT; 2 = COWAT; 3 = Digit Span; 4 = AVLT V; 5 = AVLT VII; 6 = AVLT VIII; 7 = CT1; 8 = CT2; 9 = Dig. Symbol; 10 = Blocks; 11 = RCFT copy; 12 = RCFT delayed recall; 13 = RSPM; 14 = GP dominant; 15 = GP nondominant.

(1996; 2000). The overall model fit indices were in the desired range: Comparative Fit Index (CFI) = 0.936, Tucker Lewis Index (TLI) = 0.915, Root Mean Square Error of Approximation (RMSEA) = 0.090, and Standardized Root Mean Square Residual (SRMR) = 0.069. For the CFI, a value closer to .95 is desired, as this statistic represents relative improvement when compared to an alternative model; values above .90 suggest an adequate fit (Hulland et al., 1996) and are regularly seen throughout the literature (Bollen, 1989; Hoyle & Panter, 1995). The TLI, also referred to as the Non-Normed Fit Index, compares the proposed model's fit to a null model; a TLI > .90 is recommended when evaluating model fit (Hulland et al., 1996). The RMSEA value of .09, while high compared to the desired values of .06 to .08 suggested by Hu & Bentler (1999), still indicates a passable fit (MacCallum et al., 1996). RMSEA represents the difference between observed and predicted values within the model. Lastly, a SRMR less than .08 is considered a good fit, as it suggests little differences in the matrices (Hu & Bentler, 1995, 1999).

Furthermore, all observed indicators loaded significantly ($p < .001$) on their respective corresponding factors (see Table 3).

Table 2. Descriptive statistics for 15 NeSBHIS subtests based on MI data *(N* = 127)

Test Name	Percent Missing	Minimum	Maximum	Mean	SD
P-S BNT	2.4	8	34	18.49	5.19
COWAT	1.57	2	56	22.81	10.5
Digit Span	5.51	3	17	8.26	2.24
AVLTV	2.4	3	15	9.95	2.83
AVLTVII	3.94	0	15	7.67	3.35
AVLTVIII	2.4	0	14	7.73	3.34
CT1	19.68	22	197	76.11	36.92
CT2	21.26	50	460	160.09	83.37
Dig. Symbol	5.57	2	104	40.74	19.06
Blocks	6.3	4	48	26.53	9.75
RCFT copy	6.3	5	36	25.06	8.39
RCFT delayed recall	8.66	0	26	10.02	7.06
RSPM	4.72	8	54	30.88	13.94
GP dominant	31.49	53	491	102.28	51.21
GP nondominant	32.28	54	522	107.88	48.61

Note. P-S BNT = Pontón-Satz Boston Naming Test; COWAT = Controlled Oral Word Association Test; Digit Span = EIWA Digit Span subtest; AVLT V, VII, & VIII = WHO-UCLA Auditory Verbal Learning Test, Trials V, VII, and VIII; CT1 & CT2 = Color Trails 1 & 2, respectively; Dig. Symbol = EIWA Digit Symbol Coding subtest; Blocks = EIWA Block Design Subtest; RCFT copy & delayed recall = Rey Complex Figure Test copy trial and delayed recall conditions; RSPM = Raven's Standard Progressive Matrices; GP dominant and nondominant = Grooved Pegboard Test, dominant hand and nondominant hand, respectively.

Table 3. Obtained factor loadings

	Factor Loadings	*z*-test	*p*-value
Language By			
BNT	0.787	9.3	<0.001
COWAT	0.776	9.155	<0.001
Digit Span	0.562	6.226	<0.001
Verbal Learning By			
AVLT V	0.901	12.802	<0.001
AVLT VII	0.882	12.382	<0.001
AVLT VIII	0.894	12.642	<0.001
Attention and Mental Control By			
CT1	0.835	11.38	<0.001
CT2	0.932	13.58	<0.001
Dig. Symbol	–0.766	–10.004	<0.001
Blocks	–0.308	–3.433	<0.001
Visuospatial By			
RCFT copy	0.773	9.503	<0.001
RCFT delayed recall	0.663	7.777	<0.001
RSPM	0.774	9.511	<0.001
Psychomotor By			
GP dominant	0.995	15.356	<0.001
GP nondominant	0.929	13.594	<0.001

These were the same five factors predicted by Pontón's (2000) model. The standardized loadings were high (ranging from 0.562 to 0.995), with the exception of Block Design on the Attention–Mental Control factor (–0.308), which was relatively smaller, albeit still statistically significant. The *R*-square statistics, measuring the amount of variance in each factor indicator explained by a corresponding factor, ranged from 33% to 99%, with an exception of Block Design, which was left, largely, unexplained (R^2 = .095). Factor determinacies for the five factors ranged from 0.902 to 0.995 (see Figure 1), indicating strong correlations between estimated and true factor scores.

DISCUSSION

The evaluation of Spanish-speaking Hispanic populations presents clinicians and researchers with numerous linguistic, sociocultural, and ethical challenges. Although language-specific neuropsychological measures obviate much of these concerns, many of the existing test batteries lack appropriate psychometric validation in patient groups. The NeSBHIS' good fit with Pontón et al.'s theoretical model, as well as its highly significant factor loadings and determinacies when used with a sample of Spanish-speakers with epilepsy, suggest robust construct validity in a neurological population.

While still loading on the Attention–Mental Control factor, the EIWA Block Design subtest was not a strong indicator of this factor; 91% of subtest variance was explained by other constructs. However, this finding is not entirely surprising, as the Block Design subtest did not meet criteria for factor loading in Pontón and coauthors' (2000) original factor analytic study of the NeSBHIS, but was instead grouped onto this factor based on past research (Benton, 1984). In non-native English-speaking, non-Western cultures, Block Design may not accurately measure the putative cognitive domain that it was developed to assess. Like other nonverbal neuropsychological test measures, this subtest requires specific strategies and cognitive styles that may not be taught and cultivated across cultures (Ardila, 1995; Cohen, 1969).

Pontón et al.'s theoretical model of the NeSBHIS (2000) also dictated that the EIWA Digit Span subtest would load on the Language factor, not on the Attention–Mental Control factor. While the present findings are consistent with these *a priori* assumptions, the obtained factor structure is still somewhat unexpected, as the Digit Span subtest is a robust measure of sustained auditory attention and executive functioning in English speakers with epilepsy (Bornstein et al., 1988). The absence of the Digit Span subtest on the Attention–Mental Control factor, coupled with the comparatively weak loadings of the Block Design subtest, suggests that measures subsumed by this factor probably require multiple cognitive processes (e.g., cognitive flexibility, speed of processing, and response suppression), rather than a single unitary function. Multiple investigations of both patient and community-referred samples have reported low correlations among frontally-mediated tasks, positing that these abilities are fractionated (Burgess et al., 1998). Assessing specific aspects of patient performance, including CT2 set loss errors, COWAT number of produced words, perseverances and errors, as well as maximum span on Digits Backward, may provide further explanations of the NeSBHIS component structure, and, by extension, executive abilities in Spanish-speaking Hispanic populations. Similar attempts to define specific sets of factors among frontally-mediated tasks have been successfully undertaken throughout the neuropsychology literature (Burgess et al., 1998; Burgess & Wood, 1990; Nagahama et al., 2003; Rodriguez-Aranda & Sundet, 2006).

The present study represents an important and necessary step towards investigating the generalizability of the NeSBHIS factor structure. This battery's relatively stable structure suggests that the NeSBHIS is a valid assessment tool for patients with known neurological disorders. However, like all research, ours is not free of methodological limitations. First, as previously discussed, the GP was substituted for the Pin Test, because of the formers' established ability to lateralize cerebral dysfunction in adults (Hanna-Pladdy et al., 2002). Although the GP, like the Pin Test, loaded onto a discrete Psychomotor factor, we are unable to fully comment on the stability of this subtest and related factor in a clinical sample. Second, several of the subjects were missing data secondary to interrupted testing sessions and patient inability to complete task demands. While a Monte Carlo technique (e.g., MI) was used to simultaneously preserve statistical power and retain participant variability, maximum-likelihood estimates may have been calculated equally, if not more effectively, by an expectation-maximization algorithm (Dempster et al., 1977). Third, patients included in the present study were treated at a tertiary epilepsy center; much of the studied sample has medically refractory seizures, which often require polypharmacy. As a result, participants may have an increased prevalence of medication-related side effects (e.g., psychomotor

https://doi.org/10.1017/S1355617709090250 Published online by Cambridge University Press

slowing, attention deficits, and language difficulties) and affective disorder. To this end, selection bias may limit the generalizability of the obtained findings to a less symptomatic cohort.

We also acknowledge the potential impact of intracultural variability between the studied sample and the NeSBHIS standardization sample. For example, although the composition of the NeSBHIS normative sample was representative of the overall Hispanic population (by country of origin) in 1992, these data are not indicative of the current cultural characteristics of the New York Hispanic population (which represented nearly all participants of the present study). Rather, the proportion of Mexican and Puerto Rican immigrants residing in the Northeastern states (e.g., Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont) varies considerably from the other geographic regions of the contiguous U.S. Census (2003). As was the case in the current study, only 5.5% of our participants immigrated from Mexico, as compared to 62% of Mexican-born Hispanics comprising the NeSBHIS standardization sample. Accordingly, the NeSBHIS reference group, comprised of a predominantly Mexican sample, may not accurately reflect the "standard performance" of the more culturally diverse population (including Puertorriqueños, Dominicanos, and Cubanos) residing in the Northeastern U.S. To this end, caution should be exercised when interpreting and generalizing conclusions for persons who represent significantly different linguistic characteristics from the NeSBHIS standardization sample (e.g., primarily of Mexican origin, largely monolingual Spanish-speaking, and lengthy periods of U.S. residence).

Future studies are needed to evaluate additional psychometric properties of the NeSBHIS, including its convergent and discriminant validity, when compared to other instruments. Formal assessment of this battery's sensitivity and specificity for differentiating between diagnoses is also highly warranted. By statistically evaluating the psychometric properties in a "real life" decision-making environment, and not exclusively in an experimental setting, study findings may be more applicable to other medical groups with existing neurological dysfunction (e.g., traumatic brain injury, stroke, and brain neoplasm) and to individuals receiving preoperative testing. Furthermore, like all investigations in the field of cross-cultural neuropsychology, the relationship between test performance and sociocultural variables with the potential for measurement bias should be thoroughly investigated.

ACKNOWLEDGMENTS

This investigation was supported by the National Institutes of Health under a Ruth L. Kirschstein National Research Service Award (No. 1F32HS016819) and Finding a Cure for Epilepsy and Seizures (FACES). We would also like to thank Mariya Shiyko, M.A., M.Phil., William MacAllister, Ph.D., ABPP-CN, and Thomas Braun, B.S. for their assistance with manuscript preparation.

REFERENCES

Allison, P.D. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Ardila, A. (1995). Directions of research in cross-cultural neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*(1), 143–150.

Benton, A.L. (1984). Constructional apraxia: An update. *Seminars in Neurology*, *4*, 220–222.

Benton, A.L. & Hamsher, K. (1976). *Multilingual Aphasia Examination*. Iowa City, IA: University of Iowa Press.

Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bornstein, R.A., Drake, M.E., Jr., & Pakalnis, A. (1988). WAIS-R factor structure in epileptic patients. *Epilepsia*, *29*(1), 14–18.

Burgess, P.W., Alderman, N., Evans, J., Emslie, H., & Wilson, B.A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, *4*, 547–558.

Burgess, P.W. & Wood, R.Ll. (1990). Neuropsychology of behaviour disorders following brain injury. In R.Ll. Wood (Ed.) *Neurobehavioural sequelae of Traumatic Brain Injury*. 110–132. London: Taylor & Francis.

Census, U.S. Bureau of the Census. (2000). *Statistical abstract of the United States: 2000*.

Cohen, R.A. (1969). Conceptual styles, culture conflict, and nonverbal tests of intelligence. *American Anthropologist*, *71*(5), 828–856.

Commission on Classification and Terminology of the International League Against Epilepsy. Proposal for revised classification of epilepsies and epileptic syndromes. (1989). *Epilepsia*, *30*, 389–399.

D'Elia, F., Satz, P., & Uchiyama, C. (1994). *Color Trails: Adult form manual*. Odessa, FL: Psychological Assessment Resources.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, *39*(1), 1–38.

Graham, J.W., Cumsille, P.E., & Elek-Fisk, E. (2003). Methods for handling missing data. In I.B. Weiner, D.K. Freedheim, J.A. Schinka, and W.F. Velicer. (Eds.), *Handbook of psychology* (pp. 13–63). New York: Wiley.

Graham, J.W., Hofer, S.M., & Piccinin, A.M. (1994). *Analysis with missing data in drug prevention research*. Rockville, MD: National Institute on Drug Abuse.

Graham, J.W. & Schafer, J.L. (1999). On the performance of multiple imputations for multivariate data with small sample size. In R.H. Hoyle (Ed.), *Statistical Strategies for Small Sample Research* (pp. 1–29). Thousand Oaks, CA: Sage.

Hanna-Pladdy, B., Mendoza, J.E., Apostolos, G.T., & Heilman, K.M. (2002). Lateralised motor control: Hemispheric damage and the loss of deftness. *Journal of Neurology, Neurosurgery and Psychiatry*, *73*, 574–577.

Hoyle, R.H. & Panter, A.T. (1995). *Writing about structural equation models*. London: Sage.

Hu, L. & Bentler, P.M. (1995). Evaluating model fit. In R.H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications* (pp. 76–99). London: Sage.

Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.

Hulland, J., Chow, Y.H., & Lam, S. (1996). Use of causal models in marketing research: A review. *International Journal of Research in Marketing*, *13*(2), 181–197.

Kaplan, E.F., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test*. Philadelphia: Lea & Febiger.

Little, R.J.A. & Rubin, D.B. (1986). *Statistical analysis with missing data*. New York: Wiley.

MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2) 130–149.

Maj, M., D'Elia, L., Satz, P., Janssen, R., Zaudig, M., Uchiyama, C., Starace, F., Galderisi, S., & Chervinsky, A. (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.

Maj, M., Satz, P., Janssen, R., Zaudig, M., Starace, F., D'Elia, L., Sughondhabirom, B., Mussa, M., Naber, D., Ndetei, D., Schulte, G., & Sartorious, N. (1994b). WHO Neuropsychiatric AIDS study, cross-sectional Phase II: Neuropsychological and neurological findings. *Archives of General Psychiatry*, *51*(1), 51–61.

Manly, J.J., Miller, S.W., Heaton, R.K., Byrd, D., Reilly, J., Velasquez, R.J., Saccuzzo, D.P., & Grant, I. and the HIV Neurobehavioral Research Center (HNRC) Group. (1998). The effect of African-American acculturation on neuropsychological test performance in normal and HIV-positive individuals. *Journal of the International Neuropsychological Society*, *4*, 291–302.

Matthews, C.G. & Klove, H. (1964). *Instruction Manual for the Adult Neuropsychology Test Battery*. Madison, WI: University of Wisconsin Medical School.

Muthén, B. & Muthén, L.K. (2006). MPlus (Version 4.2). Los Angeles, CA: Statmodel.

Myers, L., Barr, W.B., Vazquez, B., & Devinsky, O. (2003). An effective neuropsychological screening battery for Hispanic epilepsy patients. *The Clinical Neuropsychologist*, *17*, 92.

Myers, L.V., Barr, W.B., Benitez, J.L., Ritaccio, A., & Vazquez, B. (2002). The utility of a novel neuropsychological screening battery for Hispanics (NESBHIS) with Spanish-speaking epilepsy patients. *Epilepsia*, *43*, 185.

Nagahama, Y., Okina, T., Suzuki, N., Matsuzaki, S., Yamauchi, H., Nabatame, H., & Matsuda, M. (2003). Factor structure of a modified version of the wisconsin card sorting test: an analysis of executive deficit in Alzheimer's disease and mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*, *16*(2), 103–112.

Osterrieth, P.A. (1944). Le test de copie d'une figure complex: Contribution a l' etude de la perception et de la memoire. *Archives de Psychologie*, *30*, 286–356.

Pontón, 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.

Pontón, M.O., Satz, P., Herrera, L., Ortiz, F., Urrutia, C.P., Young, R., D'Elia, L.F., Furst, C.J., & 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*, *2*(2), 96–104.

Pontón, M.O., Satz, P., Herrera, L., Young, R., Ortiz, F., D'Elia, L., Furst, C., & Namerow, N. (1992). Modified Spanish version of the Boston Naming Test. *Clinical Neuropsychologist*, *3*(6), 334.

Rabin, L.A., Barr, W.B., & Burton, L.A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, *20*(1), 33–65.

Raven, J., Raven, J.C., & Court, J.H. (1993). *Manual for the Raven's Progressive Matrices and Vocabulary scales*. Oxford, UK: Oxford Psychologists Press.

Roberts, R.J. & Hamsher, K.D. (1984). Effects of minority status on facial recognition and naming performance. *Journal of Clinical Psychology*, *40*(2), 539–545.

Rodriguez-Aranda, C. & Sundet, K. (2006). The frontal hypothesis of cognitive aging: Factor structure and age effects on four frontal tests among healthy individuals. *The Journal of Genetic Psychology*, *167*(3), 269–287.

Rubin, D.B. (1987). *Multiple imputations for nonresponse in surveys*. New York: Wiley.

Satz, P. & D'Elia, L.F. (1989). *The Pin Test: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Strauss, E.M.S., Sherman, E.M.S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary* (3rd ed.). New York: Oxford University Press.

Taylor, E.M. (1959). *Psychological appraisal of children with cerebral defects*. Cambridge, MA: Harvard University Press.

U.S. Bureau of the Census (2003). The Hispanic Population in the United States: March 2002. Current Population Report. P20–545. Washington, D.C.

Wayman, J.C. (2003). Multiple imputation for missing data: What is it and how can I use it? Paper presented at the 2003 Annual Meeting of the American Educational Research Association, Chicago, IL.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale–Revised*. San Antonio, TX: Psychological Corporation.

Wechsler, D., Green, R.F., & Martinez, J.N. (1968). *Manual para la Escala de Inteligencia Wechsler para Adultos (EIWA)*. Cleveland, OH: Psychological Corporation.

Wong, T.M., Strickland, T.L., Fletcher-Janzen, E., Ardila, A., & Reynolds, C.R. (2000). Theoretical and practical issues in the neuropsychological assessment and treatment of culturally dissimilar patients. In E. Fletcher-Janzen, T.L. Strickland, & C.R. Reynolds (Eds.), *Handbook of Cross-Cultural Neuropsychology* (pp. 3–18). New York: Kluwer Academic.