Normative data for a brief neuropsychological battery administered to English- and Spanish-speaking community-dwelling elders

LAURIE STRICKS, JOHN PITTMAN, DIANE M. JACOBS, MARY SANO, AND YAAKOV STERN
Gertrude H. Sergievsky Center, Departments of Neurology and Psychiatry, and the Taub Center for Alzheimer's Disease Research, Columbia University College of Physicians and Surgeons

Abstract

The use of neuropsychological tests in non-English-speaking populations and among those with less education has been limited because most tests have been standardized for English-speaking populations with relatively high levels of education. In effort to establish norms, a battery of neuropsychological tests was administered, in either English or Spanish, to 995 normal elders with a wide range of educational attainment, residing in the community of Washington Heights–Inwood in northern Manhattan. Results indicate that age, education, and language all influence test performance and should be considered when evaluating neuropsychological measures. (*JINS*, 1998, *4*, 311–318.)

Keywords: Neuropsychological tests, Spanish, Normative data, Education

INTRODUCTION

The value of neuropsychological testing for diagnosing dementia in the elderly has been well established. Evaluation of test performance for an individual requires norms drawn from a population of similar characteristics. The use of neuropsychological tests in Spanish-speaking populations and among those with lower levels of education has been limited because the majority of these tests have been standardized only for English-speaking populations with relatively high levels of educational attainment. In response to this need, some normative data for Spanish versions of mental status and neuropsychological tests have been published (Ardila et al., 1994; Pontón et al., 1996; Taussig et al., 1996). While this represents an admirable beginning, these data have been primarily for Hispanics of Mexican or Columbian heritage, and may not be applicable to Hispanics from other countries. Most studies to date have used convenience as opposed to population samples, limiting the generalizability of their results. In addition, norms for many commonly used tests have not yet been developed.

We attempted to address these concerns for the population of Spanish-speaking elders typically evaluated in the New York City area, who are predominantly from the Dominican Republic, Puerto Rico, and Cuba. Our population is also unique in the relatively large number of English-speaking elders with limited educational backgrounds. A battery of neuropsychological tests, developed to evaluate dementia and cognitive change, was administered to a random sample of elderly individuals living in northern Manhattan. The purpose of this report is to present normative data for the performance of these nondemented elders as a function of education, age, and language.

METHODS

Participant Recruitment and Screening

Participants in the normative sample were selected from those participating in the Washington Heights–Inwood Columbia Aging Project (WHICAP), a community-based, epidemiologic study of dementia in Northern Manhattan. The Health Care Finance Agency (HCFA) provided a 50% random sample of names and addresses of all Medicare recipients age 65 or older who resided in the three ZIP codes in North Manhattan.

An attempt was made to contact each person and to administer the North Manhattan Dementia Screen (Gurland et al., 1992). A total of 2128 participants received the screen, representing approximately a 70% response rate. Based upon the experience of its authors, individuals who make zero to two errors on the North Manhattan Dementia Screen have a 7% probability of being diagnosed as demented on a fuller medical and neuropsychological evaluation. Since the primary aim of WHICAP was to identify individuals with dementia, only a subset of individuals with screen scores in the zero-to-two range were referred for a subsequent diagnostic evaluation, while almost all with three or more errors were referred.

Diagnostic Evaluation

Ethnic classification

Based on U.S. Census methodology, participants classified themselves as either *Hispanic* or *non-Hispanic*, and separately as either *Black*, *White*, or *Other*.

Physician's examination

The physician's examination consisted of a semistructured medical and psychiatric history, and a standardized physical and neurological examination. Bilingual physicians completed the evaluations in English or in Spanish based on the primary language of each participant, or the language in which the participant felt that he or she would perform best. In order to assist in arriving at a clinical diagnosis, the physician administered the short Blessed Information–Memory-Concentration Test (BIMC; Blessed et al., 1968; Katzman et al., 1983) and measures of activities of daily living including the Blessed Dementia Rating Scale (Blessed et al., 1968) and the Schwab and England Activities of Daily Living Scale (Schwab et al., 1969). Based on the information elicited, and independent of the neuropsychological examination, the physician recorded all neurological and medical diagnoses, as well as a determination of whether the subject was or was not demented, based on DSM–III–R criteria (American Psychiatric Association, 1987).

Selection of Normative Sample

Participants in the present study sample met the following inclusion criteria:

1. The individual had to speak either English or Spanish.

2. The individual had to have received a physician's evaluation and to have been diagnosed by the physician as nondemented. Although the neuropsychological evaluation was used as part of the diagnostic process for dementia in the parent WHICAP study, it was not used for diagnosis in the present analyses because of the danger of circularity.

3. The individual had to have received the neuropsychological battery. In some cases, participants were incapable of completing certain tests because of auditory or visual impairment. In other cases, portions of the battery were not completed due to participant refusal. We included all available data from each participant. Exclusion criteria included the presence of Parkinson's disease, history and/or clinical signs of stroke, or the presence of another neurological disease that might affect cognitive function.

Neuropsychological Battery

The neuropsychological battery was originally selected from subsets of items from standardized neuropsychological tests in order to assess intellectual functions typically affected in dementia (Stern et al., 1992). For each test, all items and instructions were translated by a committee of native Spanish speakers and back-translated to insure accuracy. A manual for the administration and scoring of the tests in the battery was developed, and is available by request from the authors. The full neuropsychological battery was administered in English or Spanish, based on the primary language of each participant, or in the language in which the participant felt that he or she would perform best. The neuropsychological testers were fluent in the language of administration.

1. *Orientation*: The 10 orientation items from the Mini-Mental State Examination (Folstein et al., 1975) were used to assess orientation to time and place.

2. *Verbal reasoning*: The Similarities subtest of the Wechsler Adult Intelligence Scale–Revised (Wechsler, 1981), which requires persons to identify relevant similarities or superordinate categories for paired items, was administered.

3. *Nonverbal reasoning*: The Identities and Oddities subtest of the Mattis Dementia Rating Scale (Mattis, 1976) was used to assess nonverbal reasoning.

4. *Naming*: A 15-item version of the Boston Naming Test (Kaplan et al., 1983), a test of visual confrontation naming, was used to assess word-finding ability.

5. *Letter fluency*: We administered the Controlled Word Association Test from the Multilingual Aphasia Examination (Benton & Hamsher, 1976). Different letters were used for Spanish- and English-speaking participants to control for word-frequency differences across the two languages as reported in the manual for the Multilingual Aphasia Examination (Benton & Hamsher, 1976).

6. *Category fluency*: All participants generated exemplars in the categories *animals*, *foods*, and *clothing*; 60 s were allowed for each category.

7. *Repetition*: Participants were asked to repeat the high-frequency phrases from the Boston Diagnostic Aphasia Examination Repetition of Phrases subtest (Goodglass & Kaplan, 1983).

8. *Auditory comprehension*: The first six items of the Complex Ideational Material subtest of the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983) were used to assess comprehension of spoken language.

9. *Word list learning and memory*: The Selective Reminding Test (SRT; Buschke & Fuld, 1974) was administered. Participants were given six trials to learn a list of 12 unrelated words. After each recall attempt, they were reminded only of those words that had not been successfully recalled. Performance measures include total recall, long-term retrieval, long-term storage, consistent long-term retrieval, and intrusions. To assess long-term retention of the word list, 15-min delayed free recall was assessed, followed by a multiple-choice recognition task.

10. *Visuoperceptual skills*: Participants matched a target design to the same design presented simultaneously in a four-choice multiple-choice array containing the target along with three distractors. Target stimuli corresponded to Form C of the original Benton Visual Retention Test (BVRT; Benton, 1955).

11. *Nonverbal memory*: A multiple-choice version of the BVRT (Benton, 1955) was used to assess nonverbal memory. Participants viewed a geometric design for 10 s. It was then removed from view, and the participant was asked to recognize the design in a four-choice array. Stimuli corresponded to Form D of the original BVRT.

12. *Visuoconstructional skills*: Participants copied five designs from the Rosen Drawing Test (Rosen, 1981) ranging in difficulty from simple geometric shapes to overlapping, parallel, and three-dimensional figures.

13. *Attention*: Two target-detection tasks were administered. The first involves the detection of a diamond shape from an array of similar shape distractors and the second involves similar detection of a consonant trigram (TMX) with phonemic distractors. Performance measures are time to completion (in seconds) and summations of errors of omission and commission.

Statistical Analyses

Weighting procedure

While the North Manhattan screen was administered to all consenting participants in the random sample, the subsequent medical and neuropsychological evaluations were not. In particular, for the reasons explained above subjects who made zero to two errors on the North Manhattan screen were undersampled. Since these individuals would be likely to perform better on the neuropsychological test battery, we established a weighting procedure to adjust the representation of participants actually tested in three screen ranges: 0 to 2, 3 to 7, and 8 and above, to the number of participants who were actually screened and fell in these ranges. To do this, we began with the participants who received neuropsychological evaluations. We then weighted each person

evaluated so that the final *N* was proportional to the distribution of screen scores in all participants who were screened. This procedure increased the number of persons evaluated to the number of persons screened, as opposed to the number that actually received the diagnostic medical and neuropsychological evaluations. To prevent inflating significance, we then divided each weight by the mean weight (i.e., the sum of all weights divided by *N*). This brought the *N* back to the actual number who received the neuropsychological evaluation while retaining the original proportional distribution of screen scores.

The results of this weighting procedure and subsequent participant selection were as follows. We attempted to administer the test battery to 1184 individuals out of 2128 screened participants. The weighting scheme was applied to these 1184 people to correct for the participant selection based on the screen score. From these 1184 individuals, nondemented, disease-free participants were selected based on the physician's evaluation, resulting in an *N* of 995. Fourteen people were eliminated because the physician's examination was not completed.

Calculation of normative data

Normative data for the neuropsychological tests battery were calculated separately for participants tested in English and Spanish. Participants were subdivided into three age ranges: 65 to 69, 70 to 79, and 80 years and above; and within these ranges into two education ranges: less than 9 and 9 years or more. Means and standard deviations were calculated for data in each age and education cell. Although we attempted to administer the full battery to each participant, because of various circumstances, not every test was administered to every individual. The actual number of participants in each cell for each test is presented in Tables 3 and 4.

Within each language group, performance across age and education was compared using analysis of variance (ANOVA). While there may be differences in neuropsychological performance between English and Spanish speakers, it was not our purpose to address these differences, since the two language groups were not matched for demographic variables. For an analysis of language-related issues in this cohort, the reader is referred to Jacobs et al. (1997).

RESULTS

Demographics of the groups tested in English and Spanish are summarized in Tables 1 and 2 according to the age and education stratifications included in the normative tables. Note that for participants with less than 9 years of education the average was 6.1 years in the English *versus* 4.5 in the Spanish-tested group.

Tables 3 and 4 present the means, standard deviations, and number of participants in each cell for the weighted English-tested and Spanish-tested samples respectively.

A 2 (education) $\times$ 3 (age) ANOVA was separately calculated for each test in the Spanish- and English-tested groups.

Table 1. Demographic features of elders tested in English and Spanish, stratified by age ranges

Variable	Tested in English				Tested in Spanish			
Age range (years); total	65–69	70–79	80+	Total	65–69	70–79	80+	Total
N	114	295	171	580	97	227	92	416
Age [years; M (SD)]	67.8 (1.4)	73.9 (2.7)	84.5 (4.2)	75.8 (6.8)	67.9 (1.1)	73.8 (2.8)	84.2 (3.6)	74.7 (6.2)
Education [years; M (SD)]	11.8 (3.9)	10.4 (3.8)	10.3 (3.9)	10.7 (3.9)	6.0 (3.9)	6.1 (4.2)	6.3 (4.5)	6.1 (4.2)
% Hispanic	8.4	3.7	2.8	4.4	100	98.5	97.8	98.7
Race (% Black)	53.4	63.9	57.5	59.9	5.0	8.2	8.3	7.5
Sex (% female)	64.7	67.4	70.3	67.7	72.4	69.8	70.0	70.4

In the group tested in English, there was one significant Education × Age interaction, for delayed recognition on the SRT. The main effects for education and age were significant at $p < .05$ for all but five tests; the age effect was not significant for intrusions on the SRT, and omission errors on the shape and TMX cancellation tasks. The education effect was not significant for time on the shape and commission errors on the TMX cancellation tasks.

For the group tested in Spanish, there was one significant Education × Age interaction, for omission errors on the shape cancellation task. The main effects for education and age were significant at $p < .05$ for all but the following tests. The age effect was not significant for orientation, naming, CFL, repetition, comprehension, Benton Matching and recognition, and the Rosen Drawing test. The education effect was not significant for Identities and Oddities, and commission errors on the shape and TMX cancellation tests.

DISCUSSION

The multicultural community of Washington Heights provides an opportunity to gather normative neuropsychological data from elderly people with a wide range of educational attainment who are speakers of either English or Spanish. Because many neuropsychological tests have been standardized with small groups, there is a need for standardization of neuropsychological measures in diverse populations in order to broaden applicability. The data reported here, collected from 995 elders, represent an attempt at such a standardization.

The major strength of this study is the size and ethnic diversity of the groups drawn from a random sample. There are some potential considerations in applying these data, however. In order to avoid circularity, we relied on a physician's diagnosis of dementia. While this diagnosis was based on an extensive evaluation that included formal mental status and functional assessments, it is likely that some of the elders would have been diagnosed as demented if the neuropsychological battery itself had been included in the diagnostic process. The data reported here may therefore slightly underestimate the performance to be expected of nondemented individuals.

The population of elders participating in WHICAP is ethnically diverse, and approximately 40% primarily or exclusively Spanish speaking. Participants identifying themselves as Hispanic varied in country of origin, but were predominantly from the Dominican Republic, Puerto Rico, and Cuba. Thus, our cohort of Spanish-speaking participants is largely limited to Hispanics of Caribbean origin and is not a representative sample of all Spanish speakers. The extent to which our results can be generalized to Hispanics from other geographic regions (e.g., Mexico, Central and South America, Spain) will need to be determined empirically. Similarly, 15% of the participants tested in Spanish reported that they speak English well or very well. Preliminary analyses suggest that these 15% perform better on many of the neuropsychological tests. This may reflect a degree of acculturation of our participants, which may not be comparable to other Spanish-speaking populations. Finally, approximately 60% of the population tested in English identified themselves as

Table 2. Demographic features of elders tested in English and Spanish, stratified by education ranges

Variable	Tested in English		Tested in Spanish	
Years of education	<9	9+	<9	9+
N	179	401	322	94
Age [years; M(SD)]	77.5 (6.8)	75.1 (6.6)	74.5 (6.0)	75.4 (7.0)
Education [years; M (SD)]	6.1 (2.1)	12.7 (2.6)	4.5 (2.7)	12.0 (3.0)
% Hispanic	3.5	4.7	99.3	96.3
Race (% Black)	75.8	52.8	8.0	5.9
Sex (% female)	66.1	68.4	71.7	66.9

Table 3. Normative data for English speakers

Age range (years)	60–69						70–79						80+						Grand mean		
Education range (years)	< 9			9+			< 9			9+			< 9			9+					
	M	(SD)	N	M	(SD)	N	M	(SD)	N	M	(SD)	N	M	(SD)	N	M	(SD)	N	M	(SD)	N
Orientation	9.6	(0.5)	21	9.6	(0.7)	93	9.4	(1.1)	92	9.6	(0.7)	200	9.0	(1.2)	63	9.2	(1.2)	102	9.4	(1.0)	571
Similarities (raw score)	6.3	(6.3)	21	15.0	(6.5)	93	6.3	(5.4)	92	12.6	(7.1)	199	4.3	(3.8)	60	10.1	(6.9)	102	10.4	(7.3)	566
Identities and Oddities	13.8	(2.0)	21	14.9	(1.3)	92	13.7	(1.7)	90	14.5	(1.7)	197	13.1	(2.7)	57	14.6	(1.4)	99	14.3	(1.8)	555
Boston Naming	12.8	(1.6)	21	14.2	(1.2)	93	12.8	(1.9)	91	13.7	(1.6)	199	12.3	(2.4)	61	13.0	(1.8)	100	13.3	(1.8)	564
CFL mean	7.8	(4.5)	21	10.6	(4.3)	93	6.1	(2.8)	91	10.1	(4.4)	200	5.6	(2.9)	58	9.1	(4.6)	102	8.8	(4.4)	566
Category Fluency mean	14.3	(3.1)	21	16.5	(4.7)	93	12.1	(3.8)	92	14.6	(4.2)	200	9.8	(3.4)	60	12.5	(4.1)	101	13.6	(4.5)	567
Repetition	7.3	(1.1)	21	7.7	(0.6)	93	7.3	(1.0)	92	7.7	(0.7)	200	6.8	(1.4)	61	7.3	(1.0)	102	7.5	(0.9)	569
Comprehension	5.0	(1.1)	21	5.6	(0.8)	93	4.9	(1.2)	92	5.4	(1.0)	200	4.9	(1.3)	60	5.1	(1.3)	100	5.2	(1.1)	566
SRT: Total recall	39.8	(10.3)	21	43.5	(9.3)	93	31.8	(9.5)	91	39.8	(9.8)	198	27.7	(10.7)	56	34.2	(11.9)	99	36.9	(11.3)	557
SRT: LTR	26.6	(13.3)	21	31.7	(13.3)	93	18.6	(10.7)	91	25.9	(13.3)	198	14.1	(10.6)	56	20.2	(14.1)	99	23.5	(13.8)	557
SRT: LTS	29.5	(14.2)	21	35.2	(13.6)	93	22.1	(11.9)	91	29.3	(14.0)	198	17.1	(11.5)	56	23.0	(14.7)	99	26.8	(14.5)	557
SRT: CLTR	19.1	(12.1)	21	23.4	(12.8)	93	11.3	(9.7)	91	18.2	(12.5)	198	7.1	(8.0)	56	13.8	(12.5)	99	16.1	(12.7)	557
SRT: Intrusions	1.6	(2.7)	21	0.9	(1.7)	93	1.3	(2.7)	91	0.6	(1.3)	198	1.3	(2.1)	56	1.1	(1.7)	99	1.0	(1.9)	557
SRT: Delayed recall	5.3	(2.8)	21	6.7	(2.8)	93	4.5	(2.6)	91	5.8	(2.7)	198	2.8	(2.5)	56	4.9	(2.9)	98	5.3	(2.9)	555
SRT: Delayed recognition	10.9	(1.9)	21	11.4	(1.2)	93	10.2	(2.0)	91	11.2	(1.4)	198	9.1	(2.3)	56	10.8	(2.1)	98	10.7	(1.8)	555
Benton matching	8.2	(1.7)	21	9.0	(1.4)	93	7.3	(2.0)	90	8.8	(1.4)	194	7.2	(2.1)	59	8.3	(1.5)	101	8.3	(1.7)	557
Benton recognition	6.6	(1.7)	21	7.7	(1.7)	93	5.8	(2.1)	90	7.3	(1.9)	193	5.3	(2.3)	59	6.9	(2.2)	100	6.8	(2.1)	556
Rosen drawing	2.2	(1.1)	21	3.2	(1.1)	93	2.4	(1.0)	90	3.0	(0.8)	197	2.1	(1.2)	58	2.6	(0.8)	98	2.7	(1.0)	557
Shape cancellation: Time	70.6	(24.5)	21	78.1	(37.4)	90	92.4	(39.0)	87	78.1	(28.1)	190	103.1	(52.6)	50	98.4	(46.7)	93	86.1	(39.1)	531
Shape cancellation: Omission errors	6.4	(3.7)	21	4.5	(3.1)	90	6.2	(3.9)	87	4.9	(3.7)	190	7.0	(4.9)	50	5.7	(4.3)	93	5.5	(3.9)	531
Shape cancellation: Commission errors	0.3	(0.6)	21	0.5	(2.5)	90	0.6	(1.7)	87	0.4	(2.1)	190	1.5	(3.5)	50	0.7	(2.5)	93	0.6	(2.3)	531
TMX cancellation: Time	93.4	(27.0)	20	85.5	(34.7)	89	110.0	(42.6)	86	88.6	(29.2)	189	124.5	(49.1)	53	105.2	(47.6)	94	98.3	(40.2)	531
TMX cancellation: Omission errors	2.9	(3.4)	20	1.1	(1.6)	88	2.2	(2.9)	86	1.5	(2.4)	189	3.1	(4.2)	53	1.8	(2.2)	94	1.8	(2.7)	529
TMX cancellation: Commission errors	0.7	(1.7)	20	0.5	(1.0)	89	1.2	(2.8)	86	0.4	(1.4)	189	2.4	(5.1)	53	0.8	(3.0)	94	0.8	(2.6)	531

Note. SRT = Selective Reminding Test, LTR = long-term retrieval, LTS = long-term storage, CLTR = consistent long-term retrieval.

Table 4. Normative data for Spanish speakers

Age range (years)	60–69						70–79						80+						Grand mean		
Education range (years)	< 9			9+			< 9			9+			< 9			9+					
	M	*(SD)*	*N*	*M*	*(SD)*	*N*	*M*	*(SD)*	*N*	*M*	*(SD)*	*N*	*M*	*(SD)*	*N*	*M*	*(SD)*	*N*	*M*	*(SD)*	*N*
Orientation	9.3	(0.8)	74	9.6	(0.8)	21	9.1	(1.1)	180	9.6	(0.5)	47	8.7	(1.5)	67	9.5	(0.8)	25	9.2	(1.1)	413
Similarities (raw score)	6.6	(4.8)	74	10.2	(6.8)	21	5.6	(4.3)	180	12.4	(6.1)	47	4.8	(3.7)	67	9.4	(6.3)	25	6.9	(5.4)	413
Identities and Oddities	13.5	(2.1)	74	14.2	(2.3)	21	13.2	(2.4)	177	13.9	(1.9)	45	12.4	(2.5)	61	13.1	(2.2)	23	13.3	(2.3)	402
Boston Naming	12.8	(1.7)	74	13.8	(1.2)	21	13.1	(1.7)	176	13.5	(1.7)	45	12.2	(2.2)	64	13.4	(1.8)	23	13.0	(1.8)	404
ABS mean	6.7	(3.3)	67	8.0	(3.5)	21	6.3	(2.7)	162	9.1	(2.9)	47	6.2	(2.9)	57	9.1	(5.4)	25	7.0	(3.3)	378
Category Fluency mean	11.6	(3.2)	74	12.8	(2.5)	21	11.2	(3.2)	180	13.2	(3.5)	47	10.4	(3.2)	67	11.8	(4.3)	25	11.5	(3.3)	413
Repetition	7.5	(0.8)	74	7.8	(0.4)	21	7.4	(0.9)	180	7.5	(1.0)	47	7.0	(1.6)	66	7.7	(0.6)	25	7.4	(1.0)	412
Comprehension	4.2	(1.3)	74	4.8	(1.6)	21	3.9	(1.5)	178	4.9	(1.1)	47	3.8	(1.4)	66	4.6	(1.0)	25	4.1	(1.4)	411
SRT: Total recall	36.2	(7.7)	74	40.5	(7.7)	21	34.1	(8.3)	179	39.1	(9.8)	47	27.6	(7.4)	66	30.6	(9.2)	25	34.1	(9.0)	412
SRT: LTR	22.8	(9.5)	74	27.2	(11.4)	21	21.3	(9.5)	179	26.0	(12.1)	47	14.8	(7.4)	66	16.1	(11.1)	25	21.0	(10.3)	412
SRT: LTS	25.5	(10.6)	74	30.5	(12.0)	21	24.5	(10.2)	179	28.9	(12.6)	47	18.4	(8.6)	66	18.9	(11.7)	25	24.2	(11.0)	412
SRT: CLTR	15.6	(8.9)	74	19.6	(11.8)	21	13.5	(8.9)	179	19.0	(11.0)	47	7.8	(8.6)	66	9.7	(9.5)	25	13.7	(9.9)	412
SRT: Intrusions	1.2	(1.8)	74	0.7	(1.4)	21	1.3	(2.1)	179	1.3	(1.7)	47	1.8	(2.6)	66	0.9	(1.7)	25	1.3	(2.1)	412
SRT: Delayed recall	4.9	(2.1)	74	5.9	(2.3)	20	4.6	(1.9)	178	5.5	(2.5)	47	3.5	(2.1)	66	4.4	(2.2)	25	4.6	(2.2)	409
SRT: Delayed recognition	10.8	(1.7)	73	11.3	(1.1)	20	10.4	(1.8)	177	10.9	(1.8)	47	9.8	(2.6)	65	10.7	(1.7)	25	10.5	(1.9)	406
Benton matching	7.2	(1.8)	72	8.3	(2.0)	21	6.8	(2.1)	167	8.2	(1.4)	45	6.1	(2.3)	60	8.1	(1.4)	23	7.1	(2.1)	389
Benton recognition	5.1	(2.1)	73	6.6	(2.2)	21	5.1	(2.0)	169	6.6	(1.7)	45	4.6	(2.1)	58	5.8	(2.1)	23	5.3	(2.1)	391
Rosen drawing	2.2	(1.2)	74	2.9	(1.1)	21	2.2	(1.1)	175	3.0	(0.7)	45	1.7	(1.2)	65	2.7	(1.2)	23	2.3	(1.2)	404
Shape cancellation: Time	88.7	(39.9)	70	64.6	(25.7)	21	98.8	(40.9)	161	77.6	(34.3)	45	121.6	(50.7)	48	109.7	(35.0)	23	96.0	(42.6)	370
Shape cancellation: Omission errors	6.1	(4.0)	70	7.1	(4.0)	21	6.2	(3.9)	161	5.0	(3.6)	47	8.9	(4.1)	48	6.3	(4.7)	23	6.4	(4.1)	371
Shape cancellation: Commission errors	1.4	(4.6)	70	0.7	(1.7)	21	1.0	(2.4)	161	0.4	(1.3)	45	2.3	(4.3)	48	4.5	(14.4)	23	1.4	(4.8)	370
TMX cancellation: Time	110.3	(41.9)	64	81.6	(35.5)	21	121.9	(43.8)	150	96.2	(36.4)	45	141.8	(51.7)	46	116.3	(31.6)	23	116.3	(44.9)	349
TMX cancellation: Omission errors	2.9	(2.8)	64	2.3	(3.1)	21	3.2	(3.3)	149	1.1	(1.9)	45	4.7	(4.4)	46	2.6	(2.9)	23	3.0	(3.3)	348
TMX cancellation: Commission errors	0.3	(0.7)	64	0.1	(0.3)	21	0.4	(1.1)	150	0.1	(0.5)	45	0.7	(1.7)	46	0.6	(1.1)	23	0.4	(1.1)	349

Black. This may not be comparable to English-speaking populations tested in other settings. This is of note because more focused analyses of our battery have detected performance differences between Whites and African Americans even after careful matching for education (Manly et al., in press).

There are some additional caveats. Although an attempt was made to obtain a random sample, not all individuals agreed to participate; thus, some biases are introduced. As this is a normative sample, equally constructed groups of English and Spanish speakers were not created, nor is the relative distribution in each cell of the tables equivalent.

Level of educational attainment significantly affected performance on the great majority of neuropsychological measures. Individuals with more education evidenced significantly and consistently higher performance than elders with lower educational attainment on all neuropsychological measures . The mediating effect of education found in English and Spanish speakers on neuropsychological test performance has been reported (Ardila et al., 1994; Bird et al., 1987; Escobar et al., 1986; Folstein et al., 1975; Magaziner et al., 1987). Although the effects of education on neuropsychological test performance have often been described, it is important in the present context to bear in mind that a given amount of education will reflect different things in different cultures. Given the strength of the established effect of education on cognitive factors (e.g., Kittner et al., 1986; Pittman et al., 1992; Stern et al., 1992; Zhang et al., 1990), it would be useful to develop normative data using more fine-grained subdivisions as opposed to the stratification presented here.

Previous studies have suggested that tests translated from English and administered in Spanish often are not equivalent to the original English version. In comparisons of Spanish and English versions of brief mental status screening instruments, such as the Mini-Mental State Examination (Bird et al., 1987; Valle et al., 1991), results indicated the Spanish-speaking participants were more likely than English speakers to miss certain items regardless of their general cognitive functioning. In a comparison of English and Spanish-translated versions of the Wechsler Adult Intelligence Scale (WAIS–R and EIWA), Lopez and Taussig (1991) found that normal elderly English-speaking participants scored considerably higher than Spanish speakers on nonverbal subtests (Block Design and Digit Span) as well as verbal subtests (Vocabulary and Similarities). There are several explanations for this observations, including the nature of the normative sample applied to the raw test scores. In addition, it is possible that the process of translation and back-translation, done in attempt to insure accuracy of the instructions and the tests in both languages, does not ensure equivalence or comparability in terms of difficulty. Lowenstein et al. (1993) report that in some instances, material translated and back-translated into Spanish may not be equivalent when compared with the English version.

Because of the paucity of normative data adjusted for the language and educational level of the subject, concern regarding the potential effects of culture, or "cultural bias," on neuropsychological test performance among the elderly has been reported (Bird et al., 1987; Valle et al., 1991). The influence of culture may be subtle and may not include aspects routinely addressed in neuropsychology, such as the readiness of members of different cultures to display their intellectual skills to strangers. Also, the actual content of tests developed in the United States may be either less salient or more unfamiliar to individuals from different cultural backgrounds (Lowenstein et al., 1993).

We have begun to address performance differences between English and Spanish speakers (Jacobs et al., 1997). These analyses, reported separately, rely on groups carefully selected to have matched distributions of salient variables including age and education. Multivariate analysis indicated significant group differences on the test battery. English and Spanish speakers scored comparably on many language-based tasks, but Spanish speakers scored significantly lower on almost all of the nonverbal measures. Significant group differences were observed on multiple-choice matching and recognition memory for stimuli from the Benton Visual Retention Test, as well as on Identities and Oddities from the Mattis Dementia Rating Scale, category fluency, and Complex Ideational Material from the BDAE. We suggested that performance differences on verbal tasks might be related to translation issues, while group differences on nonverbal tasks might be a function of cultural differences in exposure to these type of stimuli.

Our approach to these issues in the present analyses is to establish separate normative data for the Spanish- and English-speaking participants. Intensive research is required regarding the bases for performance differences across groups. Until these sources of performance differences are understood, however, we consider it practical to use observed performance in our population as a normative base. Our results will be of most direct utility to clinicians and investigators evaluating individuals similar to those in our population.

ACKNOWLEDGMENTS

REFERENCES

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (Rev. 3rd ed.). Washington, DC: Author.

Ardila, A., Rosselli, M., & Puente, A.E. (1994). *Neuropsychological evaluation of the Spanish speaker*. New York: Plenum Press.

Benton, A.L. (1955). *The Visual Retention Test*. New York: The Psychological Corporation.

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

Bird, H.R., Canino, G., Stipec, M.R., & Shrout, P. (1987). Use of the Mini-Mental State Examination in a probability sample of a Hispanic population. *Journal of Nervous and Mental Disease*, *175*, 731–737.

Blessed, G., Tomlinson, B.E., & Roth, M. (1968). The association between quantitative measures of senile change in the cerebral grey matter of elderly subjects. *British Journal of Psychology*, *114*, 797–811.

Buschke, H. & Fuld, P.A. (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology*, *24*, 1019–1025.

Escobar, J.I., Burnam, A., Karno, M., Forsythe, A., Landsverh, J., & Golding, J.M. (1986). Use of mini-mental state examination (MMSE) in a community population of mixed ethnicity. *Journal of Nervous & Mental Disease*, *174*, 607–614.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). 'Mini-mental State': A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Goodglass, H. & Kaplan, E. (1983). *The assessment of aphasia and related disorders*. Philadelphia: Lea & Febiger.

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

Jacobs, D.M., Sano, M., Albert, S., Schofield, P., Dooneief, G., & Stern, Y. (1997). Cross-cultural neuropsychological assessment: A comparison of randomly selected, demographically matched cohorts of English- and Spanish-speaking older adults. *Journal of Clinical and Experimental Neuropsychology*, *19*, 331–339.

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

Katzman, R., Brown, T., & Fuld, P. (1983). Validation of a short orientation-memory-concentration test of cognitive impairment. *American Journal of Psychiatry*, *140*, 734–738.

Kittner, S.J., White, L.R., Farmer, M.E., Wolz, M., Kaplan, E., Moes, E., Brody, J.A., & Feinleib, M. (1986). Methodological issues in screening for dementia: The problem of education adjustment. *Journal of Chronic Diseases*, *39*, 163–170.

Lopez, S.R. & Taussig, I.M. (1991). Cognitive–intellectual functioning of Spanish-speaking impaired and nonimpaired elderly: Implications for culturally sensitive assessment. *Psychological Assessment*, *3*, 448–454.

Lowenstein, D.A., Argeullas, T., Barker, W.W., & Dvara, R. (1993). A comparative analysis of neuropsychological test performance of Spanish-speaking and English-speaking patients with Alzheimer's disease. *Journal of Gerontology*, *48*, 142–149.

Magaziner, J., Basset, S.S., & Hebel, J.R. (1987). Predicting performance on the mini-mental state examination: Use of age and education-specific equations. *Journal of the American Geriatric Society*, *35*, 996–1000.

Manly, J.J., Jacobs, D.M., Sano, M., Bell, K., Merchant, C.A., Small, S.A., & Stern, Y. (in press). Cognitive test performance among nondemented elderly African Americans and Whites. *Neurology*.

Mattis, S. (1976). Mental status examination for organic mental syndrome in the elderly patient. In L. Bellak & T.B. Karasu (Eds.), *Geriatric psychiatry* (pp. 77–121). New York: Grune & Stratton.

Pittman, J., Andrews, H., Tatemichi, T., Link, B., Struening, E., Stern, Y., & Mayeux, R. (1992). Diagnosis of dementia in a heterogeneous population. A comparison of paradigm-based diagnosis and physician's diagnosis. *Archives of Neurology*, *49*, 461–467.

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*, 96–104.

Rosen, W. (1981). *The Rosen Drawing Test*. Bronx, NY: Veterans Administration Medical Center.

Schwab, J.F., England, A.C., Gillingham, F.H., & Donaldson, M.C. (1969). Projection technique for evaluating surgery in Parkinson's disease. In *Third Symposium on Parkinson's Disease* (pp. 152–157). New York: Churchill Livingstone.

Stern, Y., Andrews, H., Pittman, J., Sano, M., Tatemichi, T., Lantigua, R., & Mayeux, R. (1992). Diagnosis of dementia in a heterogeneous population: Development of a neuropsychological paradigm-based diagnosis of dementia and quantified correction for the effects of education. *Archives of Neurology*, *49*, 453–460.

Taussig, I.M., Mack, W.J., & Henderson, V.W. (1996). Concurrent validity of Spanish-language versions of the Mini-Mental State Examination, Mental Status Questionnaire, Information-Memory-Concentration Test, and Orientation–Memory–Concentration Test: Alzheimer's disease patients and nondemented elderly comparison subjects. *Journal of the International Neuropsychological Society*, *2*, 286–298.

Valle, R., Hough, R., Kolody, B., Cook-Gait, H., Figueroa, G., Jimenez, R., & Atkinson, J.H. (1991). The validation of the Blessed Mental Status Test and the Mini-Mental Status Examination with an Hispanic population. *Final report to the National Institute of Mental Health, Grant #R01 MH43390–01*.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale–Revised*. New York: The Psychological Corporation.

Zhang, M., Katzman, R., Salmon, D., Jin, H., Cai, G., Wang, Z., Qu, G., Grant, I., Yu, E., Levy, P., Klauber, M.R., & Liu, W.T. (1990). The prevalence of dementia and Alzheimer's disease in Shanghai, China: Impact of age, gender and education. *Annals of Neurology*, *27*, 428–437.