Journal of the International Neuropsychological Society (2001), **7**, 544–555.

A new verbal learning and memory test for English- and Spanish-speaking older people

HECTOR M. GONZÁLEZ,[1] DAN MUNGAS,[2,3] BRUCE R. REED,[2,3] SARAH MARSHALL,[4] AND MARY N. HAAN[1]

[1]University of Michigan, School of Public Health, Department of Epidemiology, Ann Arbor, Michigan
[2]University of California, Davis, School of Medicine, Department of Neurology, Sacramento
[3]Veterans Administration Northern California Health Care System, Martinez
[4]Berkshire Medical Center, Department of Psychiatry, Pittsfield, Massachusetts

(Received September 3, 1999; Revised May 3, 2000; Accepted June 19, 2000)

Abstract

Word-list verbal learning and memory tests with appropriate normative data can be highly sensitive to cognitive decline, but there are significant limitations of such tests available for use with older Hispanic and non-Hispanic people living in the US. The purpose of this study was to (1) create a new word-list learning and memory test in both English and Spanish and, (2) validate it with respect to sensitivity to cognitive impairment, and (3) develop statistical corrections for the effects of significant demographic variables, including ethnicity, language of administration, age, education, and gender. A community dwelling sample of 801 English- and Spanish-speaking older people was employed. Recall on learning trials and the delayed recall trial of the word-list learning test were strongly related to the Mini-Mental State Examination (MMSE), moderately related to age, and weakly related to gender and education. The relationship of word-list variables and the MMSE did not significantly differ across ethnicity/language groups. Regression coefficients for demographic variables were used in a statistical correction formula to adjust raw word-list scores, and then to develop specific percentile cut-off values. (*JINS*, 2001, *7*, 544–555.)

Keywords: Hispanic, Verbal learning, Memory, Aging

INTRODUCTION

Over the past decade, there has been a dramatic increase in the number of older persons living in the United States so that now persons above age 65 make up 12% of the population. The population of Hispanics aged 65 and over is growing especially quickly and is projected to double between 1995 and 2005 (US Bureau of the Census, 1993a). With this rapid increase in non-Hispanic and Hispanic geriatric populations, there comes an urgent need for neuropsychological instruments with favorable psychometric properties and relevant normative data.

A critical part of a comprehensive neuropsychological evaluation is the assessment of verbal learning and memory. Serial word-list learning and memory tests have been shown to be highly sensitive to episodic memory dysfunction (Delis et al., 1994) and to dementia (Cahn et al., 1995; Knopman & Ryberg, 1989; Mungas et al., 1998; Welsh et al., 1991). In addition, they are popular because a single test can offer measures of learning and forgetting, as well as other facets of memory performance (e.g., California Verbal Learning Test; Delis et al., 1987). There are several verbal learning tests available to clinicians for assessing cognitive functioning, but most rely on normative samples of convenience and few are appropriate for use with ethnic minorities. The purpose of this study was to attempt to fill these voids by developing a verbal learning and memory test with matched English and Spanish versions and to provide normative data from a large community dwelling sample.

Recently, there have been improvements in the normative data available for use with older persons. However, the problem with many existing instruments is the lack of appropriate normative data, and even major normative efforts for popular tests may fail to include significant numbers of minority participants. Ivnik and colleagues (Ivnik et al.,

Reprint requests to: Hector M. González, School of Public Health, Department of Epidemiology, University of Michigan, School of Public Health, Department of Epidemiology, 109 South Observatory Street, Ann Arbor, MI 48109. E-mail: hmgonz@umich.edu

https://doi.org/10.1017/S1355617701755026 Published online by Cambridge University Press

1992) reported normative data for the Rey Auditory Verbal Learning Test (RAVLT) based on 530 healthy community dwelling Mayo clinic patients and family members between ages 56 and 97. Of the sample reported, only 1 person was African American and one was Hispanic. Another normative study by Paolo and colleagues (Paolo et al., 1997) developed norms for the California Verbal Learning Test (CVLT) based on 212 healthy older volunteers from midwestern community and retirement centers, but only 2 African Americans and 2 Hispanics were included in the sample. Further, validation of the sensitivity of the CVLT with respect to cognitive impairment was not reported in either study.

There have been several attempts to develop and norm verbal learning and memory tests for Hispanics, but there are major limitations to these attempts. For example, Ardila and colleagues (Ardila et al., 1994) developed the Serial Learning Test, a 10-item, 10 trial memory test. They report data from a normative sample of 326 older Colombians between the ages of 56 and 80, stratified by education (0–5, 5–12, and 12 years and over). However, their study was conducted in Columbia, so that applicability to U.S. populations is unknown. Pontón and colleagues (Pontón et al., 1996) provided normative data based on 300 healthy Hispanics (ages 16–75) for the World Health Organization-University of California, Los Angeles Auditory Verbal Learning Test (WHO–UCLA AVLT), which tests learning and recall of 15 Spanish words that are presented over five trials. The WHO–UCLA AVLT learning and memory measures were significantly influenced by demographic variables and consequently the norms were stratified by age, education, and gender. However, this sample had only 69 participants between the ages of 50–75 and cell sizes for strata within that range were as small as 6. The external validity of the Serial Learning Test and the WHO–UCLA AVLT to cognitive impairment was not assessed.

Artiola i Fortuny and colleagues (Artiola i Fortuny et al., 1998) reported normative data for the Spanish Verbal Learning Test (SVLT), which consists of two lists of 16 words. Spanish-speaking samples from Madrid, Spain ($n = 205$) and regions along the US-Mexican border ($n = 185$) were recruited through flyers, social groups, and word-of-mouth specifying paid participation. However, of the total sample of 390 participants, only 24 (6%) were over age 65. Probably the best effort to develop and norm a Spanish-language verbal learning measure is that of Stricks and colleagues (Stricks et al., 1998). They reported normative data for English and Spanish versions of the Selective Reminding Test (Buschke & Fuld, 1974) from a community-based sample of 969 participants that included 557 English- and 412 Spanish-speaking people 65 years of age or older. Participants lived in New York City and the Spanish-speaking group was largely comprised of people of Caribbean decent. While excellent in terms of representativeness of that population, the applicability of these norms to Hispanics who are culturally and linguistically different, for example Hispanics of Mexican heritage living in California, is open to question.

Thus, despite an abundance of verbal list learning measures, there are few good choices for use with older persons, in particular older Hispanics. The purpose of this study was to develop a verbal learning and memory test to fill that void. We sought to create a test with psychometrically matched English and Spanish versions and to provide normative data from a large sample of older persons encompassing a wide range of educational achievement. Secondarily we sought to define the effects of major demographic variables on test performance and to assess its sensitivity to detecting cognitive impairment.

METHODS

Research Participants

There were three sources of participants for this study: (1) Sacramento Community Survey, (2) Winters Community Survey, and (3) Sacramento Area Latino Study on Aging (SALSA). Participants were eligible for the Sacramento and Winters Community Survey if (1) they were over the age of 60, (2) English or Spanish was their primary language, and (3) they did not have physical or sensory deficits that precluded test taking. For the SALSA project, participants had to meet the above criteria and self-identify as being of Hispanic or Latino ancestry. A broad definition of Hispanic/Latino ancestry was used for all groups, and was defined by any one of the following: (1) self-selected Hispanic/Latino as ethnic group, (2) first language was Spanish, or (3) participant or parent or grandparent was born in Mexico or other Latin American country. Bilingual participants, who were fluent in both languages and actively used both languages, were allowed to self-select their preferred language of administration. (In subsequent work, language of test administration is assigned using an objective algorithm based upon language usage in a variety of situations. Preferred language and test language identified by the algorithm are the same in 96% of cases). Bilingual/bicultural technicians fluent in both languages conducted testing and participants were paid for their participation.

Sacramento County is a largely urban area with a total population of more than 1,000,000 people. Winters is a nearby rural town with a population of approximately 6000. The Sacramento and Winters recruitment was accomplished using a door-to-door survey. Technicians went to all houses in Winters and in targeted Census tracts in Sacramento, and determined if there was an eligible participant. A survey interview was conducted to obtain information about demographic variables, health status, and functional status. In addition, the Mini Mental State Exam (MMSE; Folstein et al., 1975) was administered to assess global cognitive functioning. The Spanish translation of the MMSE and results pertaining to demographic differences in this instrument have been reported previously (Marshall et al., 1997; Mungas et al., 1996).

For the Sacramento community survey recruitment, 58% of surveyed English participants who were invited to participate in test administration completed testing. Response rate for Sacramento Spanish participants was 55%. Response rates for Winters were 84% - English, and 81% - Spanish. Winters participants were scheduled for test administration at the time of the interview in their home. The Sacramento survey started while test item construction was in progress, and there was a several month delay for many participants between the survey interview and contact regarding participation in the test administration component of the study.

SALSA recruitment targeted Census tracts of Sacramento County with proportional densities of Hispanics greater than 10% based on updated 1990 U.S. Census information (US Bureau of the Census, 1991). The recruitment method was designed to enumerate all Hispanic households within the targeted Census tracts. The multi-tier approach involved mailing Hispanic households information regarding SALSA, which was followed by a telephone call. Households that could not be contacted by telephone or mail were visited by technicians. Community outreach methods were used to increase community awareness of the study, and to create a receptive environment to SALSA. Of the 579 SALSA participants, 373 (64.4%) responded to mailings and outreach recruitment methods, 119 (20.6%) responded to telephone contact, 60 (10.4%) responded when they were visited by a technician, and 27 (4.7%) were referrals from other contacts. These efforts resulted in 82.2% of eligible older Hispanics participating in the study. The verbal learning test and MMSE/Modified Mini Mental State Examination (Teng & Chui, 1987) were administered to all SALSA participants by a bilingual/bicultural technician as part of their initial survey interview. The numbers of participants from each recruitment source and demographic summaries are presented in Table 1. For purposes of this study, participants were divided into three groups defined by language of test administration and ethnicity: (1) English administration, non-Hispanic (n = 112), (2) English administration, Hispanic (n = 319), and (3) Spanish administration, Hispanic (n = 370). The demographic characteristics of these three groups are presented in Table 2.

Bilinguality in English and Spanish, defined by self report of speaking both English and Spanish well or often, was present in 2.7% of the English–non-Hispanic group (97.3% monolingual English), 61.4% of the English–Hispanic group (38.6% monolingual English), and 32.4% of the Spanish–Hispanic group (67.6% monolingual Spanish). Hispanic participants were predominantly of Mexican ancestry; 88.2% were born in Mexico or had parents or grandparents from Mexico. Length of residence in the United States was collected from the Sacramento and Winters groups (but was not available for the SALSA sub-sample). Average residence for the English–Hispanic group was 67.6 years (SD = 10.6, range = 31–86), and was 37.6 years (SD = 17.6, range = 1–69) for the Spanish–Hispanic group.

Spanish English Verbal Learning Test (SEVLT)

The SEVLT consisted of a list of 15 items that could be purchased in a store, and had an English and Spanish version that were direct translations of one another. Translation from English to Spanish occurred first, and the resulting Spanish version was then independently translated back to English by a different translator. A committee then compared the original English version with the back-translated English version, and where discrepancies were present, arrived at a consensus about the most appropriate translation. The list was composed of five semantic categories with five exemplars of vegetables, four drinks, three kitchen utensils, two reading materials, and one fruit. List items were selected based upon English language prototypicality norms for semantic categories reported by Uyeda and Mandler (1980). The testing procedure followed a standard word-list learning test format. Words were presented at a rate of 1 word/s with an immediate recall trial after each complete list presentation. Order of administration was fixed across trials, and was arranged so that two words from the same category were not presented consecutively (see Appendix,

Table 1. Summary of participant numbers and demographics by recruitment source

Ethnicity/language	Recruitment source	n	Percentage female	Age M	Age (SD)	Education M	Education (SD)	MMSE M	MMSE (SD)
Non-Hispanic	Sacramento	77	59.7	69.4	(6.5)	12.2	(3.6)	28.0	(2.5)
English	Winters	35	54.3	70.1	(6.4)	12.9	(2.7)	28.0	(1.8)
Hispanic	Sacramento	30	56.7	68.3	(5.7)	8.4	(4.2)	25.8	(3.5)
English	Winters	12	58.3	77.0	(8.3)	8.3	(3.1)	27.2	(2.4)
Hispanic	Sacramento	34	61.8	69.2	(7.9)	5.9	(4.0)	25.3	(3.7)
Spanish	Winters	34	61.8	69.0	(6.4)	3.4	(3.5)	24.4	(4.5)
English	SALSA	277	56.0	68.6	(6.6)	10.8	(4.3)	27.2	(3.6)
Spanish	SALSA	302	60.9	70.4	(7.8)	5.5	(4.5)	24.7	(5.1)
Totals		801	58.7	69.7	(7.2)	8.4	(5.1)	26.1	(4.3)

Note. Sacramento/Winters = Sacramento and Winters Community Survey, SALSA = Sacramento Area Latino Study on Aging, MMSE = Mini-Mental State.

Table 2. Summary of participant numbers and demographics by ethnicity/language groups

Ethnicity/language	*n*	Percentage female	Age *M*	Age (*SD*)	Education *M*	Education (*SD*)	MMSE *M*	MMSE (*SD*)
Non-Hispanic/English	112	58.0	70.2	(6.5)	12.4	(3.6)	28.0	(2.3)
Hispanic/English	319	56.1	68.9†	(6.7)	10.4***	(4.3)	27.1	(3.6)
Hispanic/Spanish	370	61.0	70.2	(7.7)	5.3***	(4.4)	24.8***	(5.0)

†$p < 0.10$;*** $p < 0.001$ (Bonferroni corrected).

List A). After the five learning trials, an interference procedure was introduced in which participants were asked to repeat aloud words from a separate 15-word list (see Appendix, List B). Immediately following the interference procedure, delayed free-recall was tested for List A. Dependent measures examined in this study included the total number of items correctly recalled for each of the five learning trials and the memory or delayed recall trial.

Simulation Study of Dementia Effects on Statistical Adjustment and Diagnostic Accuracy

In order to develop test norms, groups of data analyses were performed to statistically adjust trial scores for demographic variables and to then derive normative values for these adjusted scores. The likelihood that some of the community dwelling participants in this study were demented raises an important issue with respect to derivation of norms. Older age and lower levels of education have been reported as risk factors for dementia, and consequently, part of their effect on learning variables in this study might be attributable to presence of dementia in some of the participants. Similarly, the presence of demented individuals in a normative sample might affect diagnostic sensitivity and specificity. A simulated dataset was created to examine effects of having demented individuals in the sample used for derivation of statistical adjustments and norms, and specifically, was used to model the effects of education on a screening measure of dementia.

Actual education scores from 400 participants in a previous study associated with this project (Mungas et al., 2000) were the basis for this dataset, and reflected a broad range of variability ($M = 8.82$, $SD = 5.25$, range = 0–21). From these education scores, two simulated cognitive test scores were randomly generated, each normally distributed with a mean of 100 and standard deviation of 15, one having a correlation of .30 with education and the other having a correlation of .60 with education. Effects of dementia were simulated as follows. First, it was assumed that the risk of dementia would be two times higher in low education cases ($\leq$ 8 years of education) and that prevalence would be 10% in high education cases ($>$ 8 years of education). Based upon these assumptions, 20% of low education cases and 10% of high education cases were randomly selected as demented cases. Effects of dementia were simulated by randomly generating a uniform distribution of scores ranging from 10 (two-thirds of a standard deviation of the simulated test score) to 30 (2 standard deviations). For *demented* cases, the simulated dementia effect was subtracted from the simulated test score yielding two final simulated variables incorporating effects for education (r = .30 for one and r = .60 for the other) and incorporating effects of dementia in each simulated score.

The regression coefficient for education as a predictor of the simulated score, based upon a weak education effect, was approximately 40% larger in the full sample than in the sample in which the demented cases were excluded. An increase of approximately 15% was observed for the simulated score based upon a strong education effect. Thus, inclusion of demented cases resulted in greater bias when the adjustment variable had a weak relationship to the cognitive measure and resulted in substantial overcorrection of the test score.

Diagnostic accuracy in identifying dementia was also tested for adjusted scores created and normed on the full sample *versus* those derived from the non-demented sample. The adjusted scores based upon the full sample used the regression coefficients from the full sample as the basis for statistical adjustment, and then employed either the 10th or 20th percentile of the full sample as the cutoff for identifying dementia. Results were compared with adjusted scores based upon regression coefficients and the 10th or 20th percentile derived from the nondemented sample. Results were essentially identical for weak and strong education effects, as would be expected. When the 10th percentile was used as a cut-off, sensitivity of the full sample adjusted score was 49% in contrast to sensitivity of 75% for the nondemented sample score. However, specificity was greater for the full sample (97% *vs*. 91%), as was total accuracy (91% *vs*. 88%). Using the 20th percentile as a cut-off, sensitivity was 77% (full) and 82% (nondemented), specificity was 90% (full) and 80% (nondemented), and total correct was 88% (full) and 81% (nondemented).

These results indicate that the presence of dementia in the normative sample decreased sensitivity but increased specificity. The decrease in sensitivity was quite striking when using the 10th percentile, but was quite modest when using the 20th percentile as the cut-off. Overall, these results indicate that the diagnostic accuracy attained using

https://doi.org/10.1017/S1355617701755026 Published online by Cambridge University Press

the 20th percentile of the full sample scores was essentially the same as that associated with the 10th percentile of the nondemented sample scores. Of course, these results are dependent upon the assumptions involved in creating the simulated dataset, and one would expect different results with different baserates of dementia, different strength of the risk factor for dementia, and different degree of impairment associated with dementia. The values used in this simulation were selected to represent reasonable estimates of the prevalence of dementia and relative risk for dementia associated with low education in a community sample, and to model mild to moderate dementia effects, like would be encountered in a real-world clinical situation. Results are intended to illustrate some of the issues that can influence use of tests in real world situations.

Data Analysis: Identification of Factors Affecting Learning Test Performance

The relationship between SEVLT performance, demographic variables, and global cognitive functioning was evaluated. A multivariate general linear model was used in which the five learning trials and the delayed recall trial were repeated measures, dependent variables. Independent variables included *ethnicity/language group* (non-Hispanic/English test administration, Hispanic/English, and Hispanic/Spanish) and gender as between group variables, and MMSE, age and education as continuous independent variables. Terms to model the interaction of ethnicity/language group with other independent variables were also included. Thus, the following independent variables were included: ethnicity/language group (*group*), age, education, gender, MMSE, Group × Age, and Group × Education, Group × Gender, and Group × MMSE. Age, education, MMSE were centered by subtracting from each score the overall mean based upon the whole sample to avoid problems associated with collinearity of interaction and main effects (Rawlings et al., 1998). The relationship of scale scores to MMSE was considered a critical validation of the sensitivity of the SEVLT scores to cognitive impairment. To be an effective measure of cognitive impairment, test scores should be related to MMSE scores after controlling for demographic variables. Further, test scores should be equally sensitive to cognitive impairment in different ethnicity/language groups if the test provides unbiased measurement across groups. Thus, the MMSE main effect and the Group × MMSE interaction were considered to be critical in the validation of the test.

Statistical Adjustment for Demographic Variables and Norms

The general approach used to adjust for demographic variables and create norms was as follows. Trial total raw scores were entered as dependent variables into a multivariate, general linear model with demographic variables group, gender, age, and education as independent variables. In addition, MMSE was added as an independent variable so that effects related to demographic variables would be independent of degree of global cognitive performance, and dementia. Normative data was calculated based upon two different samples. The first included all 801 participants, while in the second ($n = 714$), participants were excluded if they had MMSE scores, adjusted for effects of age and education (Mungas et al., 1996), that fell below the 10th percentile of the total sample.

Unstandardized regression coefficients for demographic variables entered into this analysis were used to adjust raw scale scores for effects of demographic variables. The general form for this adjustment was Adjusted Trial Score = Raw Trial Score $- b_{\text{Group}} - b_{\text{Gender}} - (b_{\text{Age}} \times (\text{Age} - 70)) - (b_{\text{Education}} \times (\text{Education} - 12))$. In this equation, b_{Group} is the coefficient corresponding to the ethnic/language group of the subject, b_{Gender} is the coefficient corresponding to the gender of the subject, and b_{Age} and $b_{\text{Education}}$ are the unstandardized coefficients for age and education. Age of 70 and education of 12 are arbitrary values selected as standard reference points. The resulting adjusted score is uncorrelated with the demographic variables, but maintains a scale of measurement similar to that of the raw scores. This is the same approach previously used to adjust the MMSE for effects of age and education in older persons tested in English and Spanish (Mungas et al., 1996).

RESULTS

Demographic Characteristics of Groups

Differences in age, education and MMSE score among participants from the three recruitment sources (Table 1) were evaluated using an analysis of variance in which recruitment source (Sacramento, Winters, SALSA) and language of test administration (English, Spanish) were crossed, between groups factors. For age, the effect for the recruitment source [$F(2,795) = 2.16$, $p < .115$] and language group [$F(1,795) = 1.13$, $p < .288$] did not reach statistical significance, but the interaction term was statistically significant [$F(2,795) = 6.58$, $p < .001$]. For education, the recruitment source effect was not significant [$F(2,795) = 1.07$, $p < .342$], but language group [$F(1,795) = 204.04$, $p < .001$] and their interaction [$F(2,795) = 4.40$, $p < .013$] were statistically significant. The MMSE by recruitment source [$F(2,795) = 2.16$, $p < .342$] and interaction term [$F(2,795) = .49$, $p < .611$] did not reach significant levels, but the language group effect for the MMSE was significant [$F(1,795) = 37.86$, $p < .001$], with Spanish administration participants having lower scores. There were no significant gender differences corresponding to recruitment source, language group, or their interaction.

Recruitment source differences within language groups were examined using *post-hoc* pairwise comparisons of the three recruitment sources, with Bonferroni correction of the p value required for statistical significance (a value of $p = .05/3 = .017$ was required). None of the compari-

https://doi.org/10.1017/S1355617701755026 Published online by Cambridge University Press

sons between recruitment sources for age, education, MMSE, and gender reached statistically significant levels either before and after the Bonferroni criteria were applied. Further, the age ($p = .288$) and gender composition between the language groups did not reach significance. There were, however, statistically significant differences between language groups for education ($p < .001$) and MMSE ($p < .001$). Spanish-speakers, on the average, had fewer years (6.28 years) of education than English-speakers. Similarly, they had lower MMSE scores (2.65 points) than did English-speakers.

Comparisons of the ethnicity/language groups demographics (Table 2) with ANOVA indicated that there were significant age [$F(2,798) = 3.11$, $p < .045$), education [$F(2,798) = 186.49$, $p < .001$], and MMSE [$F(2,798) = 40.31$, $p < .001$] differences between the three groups, but not gender [$\chi^2(2) = 1.77$, $p = .41$]. *Post-hoc* analyses with Bonferroni corrections showed a statistical trend for Hispanic English-speakers to be younger. Non-Hispanics had significantly more years of education than both Hispanic groups and Hispanic English-speakers had significantly more education than Spanish-speakers. For MMSE scores, there was not a significant difference between the two English speaking groups, and Spanish speakers scored lower than both of these groups.

The population and normative samples differed according to whether participants were included who scored below the 10th percentile on the age and education adjusted MMSE. Consequently, those scoring below and above the 10th percentile on this measure were compared with respect to demographic characteristics. These two groups significantly differed with respect to distribution across ethnicity/language group [$\chi^2(2) = 15.6$, $p < .001$]. Low-MMSE participants comprised 15.4% of the Hispanic–Spanish group, 7.8% of the Hispanic–English group, and 4.5% of the non-Hispanic–English group. Significant differences were also present for age [$F(1,799) = 10.0$, $p < .002$; M (SEM) low-MMSE = 72.0 (0.77), high-MMSE = 69.4 (0.27)] and education [$F(1,799) = 9.5$, $p < .003$; M (SEM) low-MMSE = 6.8 (0.54), high-MMSE = 8.5 (0.19)]. Groups did not significantly differ in gender distribution. These results show that low-MMSE participants were more likely to be in the Hispanic-Spanish group, were older, and had lower average education.

Factors Affecting Learning Test Performance

Multivariate tests for effects involving trials were used since a test for sphericity indicated that this assumption for univariate repeated measures tests was violated. Results of the general linear model are presented in Table 3. Significant main effects, indicating a relationship between the independent variable and the sum of scores across trials, were observed for age, education, gender, and MMSE, but not for group. The Group × Age and Group × Gender interactions were significant, and there was a statistically significant trend for the Group × MMSE interaction. Even though this effect approached significance, it was a very weak effect. The canonical correlation for this effect, which adjusts for effects of all other independent variables, was .084, which corresponds to about 0.7% of the variance. The canonical correlation of the MMSE with the sum of scores across trials was .30, with higher MMSE scores associated with

Table 3. Summary of multivariate test interactions and main effects between the Spanish English Verbal Learning Test, demographic variables, and Mini-Mental State Exam (MMSE)

Effect	Wilk’s	F	df	p
Ethnicity/language group	.995	1.72	2, 786	0.18
Age (years)	.927	62.22	1, 786	0.0001
Education (years)	.994	4.44	1, 786	0.035
Gender	.983	13.35	1, 786	0.0003
MMSE	.912	76.14	1, 786	0.0001
Group × Age	.999	0.30	2, 786	0.74
Group × Education	.997	1.23	2, 786	0.29
Group × Gender	.983	6.99	2, 786	0.001
Group × MMSE	.993	2.82	2, 786	0.06
Trials	.394	240.39	5, 782	0.0001
Trials × Group	.990	0.79	10,1564	0.637
Trials × Age	.978	3.53	5, 782	0.0037
Trials × Education	.995	0.82	5, 782	0.536
Trials × Gender	.996	0.61	5, 782	0.689
Trials × MMSE	.917	14.11	5, 782	0.0001
Trials × Group × Age	.972	2.25	10,1564	0.013
Trials × Group × Education	.992	0.62	10,1564	0.79
Trials × Group × Gender	.988	0.94	10,1564	0.494
Trials × Group × MMSE	.971	2.34	10,1564	0.0096

better overall performance. The canonical correlation of gender with learning performance was .13, with significantly better performance for females. However, the Group × Gender interaction was also significant ($p < .001$, with a canonical correlation of .13). *Post-hoc* pairwise comparisons of males and females within each group were performed, using Bonferroni correction for the number of comparisons (a value of $p = .05/3 = .017$ was required for statistical significance). Non-Hispanic English females had significantly better overall performance than non-Hispanic English males ($p < .0001$), while males and females from the other two groups did not significantly differ. Younger age (canonical $r = .27$) and higher education (canonical $r = .07$) were associated with better overall performance. It is noteworthy that the magnitude of the education effect was quite small, accounting for less than 1% of the variance, even though it was statistically significant.

The trials main effect was highly significant. Significant interactions with trials were observed for Age and MMSE, which indicates that the effects of these independent variables differed across trials (Table 3). Three way interactions involving Group × Age × Trials and Group × MMSE × Trials were also statistically significant, but relationships were weak with canonical correlations less than .15. To better characterize these interaction effects, univariate analyses were performed using each trial as a dependent variable and the independent variables from the multivariate analysis (Table 4). Bonferroni correction was used to adjust the criterion p value for statistical significance ($p = .05/6 = .0083$). MMSE was not significantly related to Trial 1 performance, but was significantly related to scores on all other trials with increasing strength of relationship across trials. Age was associated with performance on all trials.

Results indicate that the MMSE was clearly the strongest predictor of test performance, and that later trials and delayed recall were most strongly associated with the MMSE. Age showed a moderately strong relationship to test performance, while there was a weak but statistically significant effect of gender (poorer performance for males), especially in the non-Hispanic English administration group. The association of education with test performance was very weak. Test performance did not differ across ethnicity/language group, and a very similar relationship of test performance and the MMSE was observed across the three groups. The results were essentially unchanged when this analysis was repeated excluding participants who scored below the 10th percentile on the age and education adjusted MMSE.

An additional, secondary analysis was performed to address effects related to monolingual *versus* bilingual language status. Participants were grouped according to whether they were monolingual English, monolingual Spanish, or bilingual. A MANOVA was performed in which trials were dependent variables and gender, age, education, and linguistic group were independent variables. In this analysis, linguistic group was significantly related to the average score across trials [$F(2,795) = 5.20$, $p < .006$], but the strength of relationship was relatively weak (canonical $r = .11$), accounting for 1.2% of total variance. The effect of linguistic group on test performance did not significantly differ across trials ($F < 1.0$). For all trials, the two monolingual groups were nearly identical, while the bilingual group showed slightly poorer performance, with a maximum difference with either other group on any trial of 0.6 words recalled. An additional analysis was performed in which MMSE was added to the model. In this analysis, neither the linguistic group main effect ($p > .13$) nor the Group × Trial interaction ($F < 1.0$) were significant. These results indicate that bilingual subjects had slightly poorer performance than their monolingual counterparts, but these differences were not present after adjusting for MMSE score.

Table 4. Univariate analyses results of learning and memory trials, age, and Mini-Mental State Exam (MMSE)

	Standardized beta	
Trial	Age	MMSE
1	−.21**	.16
2	−.23**	.39**
3	−.25**	.44**
4	−.25**	.46**
5	−.21**	.53**
Delayed Recall	−.24**	.55**

*$p < .0083$.
**$p < .0017$.

Statistical Adjustment of Scores and Norms

Results of the regression analyses used to create adjusted scores are presented in Table 5. Residuals from each analysis were examined to test viability of statistical assumptions underlying multiple regression. The Shapiro-Wilk W test was used to evaluate normality of the distributions of residuals for each analysis. The distribution for the delayed recall trial was significantly nonnormal ($W = .98$, $p < .04$), but distributions for the other five trials were not significantly nonnormal (p's $> .15$). The distribution of residuals for delayed recall was negatively skewed (skewness = −.32). Residuals were plotted against predicted values and did not show any significant linear or nonlinear relationships. In the overall sample, the numbers of participants falling within specific age and education ranges are shown in Table 6.

The derived regression coefficients were entered into the general correction formula previously described and used to derive adjusted scores for each of the six dependent variables. Results are presented for the complete sample (Table 7), and for the normative sample in which cases with adjusted MMSE scores falling below the 10th percentile were eliminated (Table 8). The number of participants within specific age and education groups above and below the 10th percentile are shown in Table 6.

https://doi.org/10.1017/S1355617701755026 Published online by Cambridge University Press

Table 5. Results of the regression analyses used to create adjusted scores for the Spanish English Verbal Learning Test

	Unstandardized Beta coefficients					
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Delayed Recall
Non-Hispanic English*	.6483	.6660	.7850	.9301	.6655	.7441
Hispanic English*	−.3180	−.2911	−.4420	−.3675	−.2682	−.2201
Hispanic Spanish*	−.3303	−.3749	−.3431	−.5627	−.3973	−.5239
Male**	−.0963	−.1869	−.1424	−.1362	−.1916	−.1268
Female**	.0963	.1869	.1424	.1362	.1916	.1268
Age	−.0737	−.1000	−.1193	−.1444	−.1372	−.1498
Education	.0781	.1055	.1444	.1157	.1343	.1235

*Coefficient values are added or subtracted from the equation depending on the ethnicity and language of the individual.
**Coefficient values are added or subtracted from the equation depending on the individual's gender.

DISCUSSION

This study describes a newly developed verbal learning test with English and Spanish language versions, and provides favorable evidence of psychometric equivalence of the two versions. The effects of important demographic variables are reported, and sensitivity to an independent measure of global cognitive functioning is described. In addition, normative data are reported for a large sample of English- and Spanish-speaking older persons and statistical adjustments for the effects of demographic variables are described.

The sample used in this study is larger than previously reported for this type of test. This is especially true for the combined sample of 689 Hispanics, which included 319 participants tested in English and 370 tested in Spanish. This study is relatively unique in that it provides normative data for English- and Spanish-speaking older persons of predominantly Mexican-American decent, a group which comprises nearly two-thirds of the Hispanics in the country (US Bureau of the Census, 1993a). There are important clinical and research advantages of having carefully constructed, well-normed English and Spanish versions of the same test since Hispanics have varying degrees of English and Spanish proficiency. Clinically, this provides flexibility for testing in either language depending upon an evaluation of the most appropriate language for assessment. Having equivalent English and Spanish language versions also is critical for research with Hispanic samples that include individuals with varying degrees of English and Spanish proficiency. The non-Hispanic subsample in this study was smaller than the Hispanic subsamples, but compares favorably with others in the literature. Others (Ivnik et al., 1992; Paolo et al., 1997) have reported larger samples, but their samples have significant limitations. An important consideration in normative studies is in the degree to which the samples are representative of the population at large. Education levels of the samples of the previous two studies were considerably higher than the general US population of older Americans (US Bureau of the Census, 1993b), raising an important question about generality of their samples.

Table 6. Summary of participant numbers by age and education groups

Age groups	Education groups (years)	Total sample *n*	Normative sample *n*
60–69	0–8	196	176
	9–12	124	117
	13+	112	107
70–79	0–8	166	151
	9–12	85	79
	13+	44	43
80+	0–8	51	42
	9–12	16	13
	13+	7	6

The representativeness of the sample in this study is an equally important issue. The Sacramento and Winters participants were recruited using door-to-door survey methods designed to select a sample representative of the communities being surveyed. The response rate in Sacramento was lower than would be desired, and non-responders were older, had less education, and had lower MMSE scores (Mungas et al., 2000). Consequently, mean scores from Sacramento participants are likely to be somewhat higher than they would be in the population at large. The majority of SALSA participants were volunteers responding to outreach efforts, and consequently might differ from those who did not volunteer. While these are factors that could potentially limit the representativeness of the sample used in this study, several considerations argue that this should be a very effective sample for normative purposes. First, the three different recruitment sources were relatively closely matched in age, education, and gender, and the higher age and lower education of Winters participants may well represent real demographic differences between Winters and Sacramento.

https://doi.org/10.1017/S1355617701755026 Published online by Cambridge University Press

Table 7. Population means for the total sample ($N = 801$) with specific cutoffs for adjusted scores corresponding to different percentile values for Spanish English Verbal Learning Test Measures

			Percentiles														
Measure	*M*	(*SD*)	1	5	10	15	20	30	40	50	60	70	80	85	90	95	99
Trial 1	4.95	(1.72)	0.9	2.2	2.7	3.2	3.5	4.0	4.5	4.9	5.4	5.8	6.3	6.6	7.1	7.9	9.2
Trial 2	7.07	(2.08)	2.0	3.8	4.6	5.0	5.3	6.0	6.5	7.0	7.6	8.1	8.8	9.2	9.7	10.6	12.0
Trial 3	8.44	(2.49)	2.5	4.2	5.2	5.8	6.3	7.3	7.8	8.4	9.2	9.9	10.5	11.1	11.6	12.6	13.7
Trial 4	9.25	(2.65)	2.6	4.5	5.7	6.6	7.1	8.0	8.7	9.3	10.0	10.7	11.5	12.0	12.6	13.3	14.8
Trial 5	9.83	(2.75)	2.2	4.9	6.2	7.0	7.7	8.6	9.3	10.0	10.7	11.5	12.3	12.7	13.2	13.9	15.1
Delayed Recall	9.24	(2.90)	1.2	4.0	5.4	6.4	7.0	8.0	8.7	9.4	10.2	10.9	11.6	12.2	12.8	13.6	14.7

Table 8. Normative sample of cases ($N = 734$) scoring above the 10th percentile of the adjusted Mini-Mental Status Examination specific cutoffs for adjusted scores corresponding to different percentile values for Spanish English Verbal Learning Test Measures.

			Percentiles														
Measure	*M*	(*SD*)	1	5	10	15	20	30	40	50	60	70	80	85	90	95	99
Trial 1	5.05	(1.68)	1.0	2.4	2.9	3.3	3.7	4.1	4.6	5.0	5.5	5.9	6.4	6.8	7.2	8.0	9.3
Trial 2	7.23	(1.99)	2.5	4.1	4.8	5.2	5.5	6.2	6.7	7.2	7.7	8.2	8.9	9.3	9.8	10.6	12.1
Trial 3	8.65	(2.37)	2.7	4.8	5.6	6.1	6.6	7.5	8.0	8.6	9.3	10.0	10.7	11.1	11.7	12.7	13.8
Trial 4	9.50	(2.49)	3.2	5.4	6.2	6.9	7.5	8.3	8.9	9.5	10.2	10.9	11.6	12.2	12.7	13.4	15.1
Trial 5	10.11	(2.53)	3.6	5.8	6.8	7.4	8.1	8.9	9.5	10.2	11.0	11.6	12.4	12.8	13.3	13.9	15.1
Delayed Recall	9.53	(2.68)	1.3	4.8	6.1	6.9	7.4	8.3	9.0	9.7	10.3	11.0	11.8	12.4	12.9	13.7	14.8

In addition, the recruitment sources were closely matched in MMSE scores after accounting for differences associated with language groups. Second, the statistical adjustments take effects of age and education into account and so would substantially decrease any bias associated with nonrepresentativeness in sampling related to these variables. Regression coefficients, used in the statistical corrections, are not dependent upon strictly representative sampling if a reasonable range of variation is included in the sample (Hambleton et al., 1991) and there was very broad variability in participants included in this study. Finally, the sample size, sampling methods, and overall response rate in this study compare very favorably to previous normative studies for neuropsychological tests.

There were clear and predictable differences in the education levels and mean MMSE total scores between the language/ethnicity groups in this study. Non-Hispanic–English participants had significantly higher education than both Hispanic groups, regardless of language. The low levels of education of Hispanic samples are consistent with other reports on the educational levels of older Hispanics (US Bureau of the Census, 1993a). Not surprisingly, when the Hispanic groups were divided by language of test administration, average education for English administered participants (10.4) was nearly twice that for Spanish administered participants (5.3). A similar pattern was observed for MMSE scores, with clearly lower scores for Spanish language (24.8) compared to English-language (27.1) participants. Although it is possible that the differences in score can be attributable to translations from English to Spanish, the Spanish version of the MMSE used in this study was carefully translated and backtranslated for accuracy, thereby minimizing this possibility (Mungas et al., 1996). Further, when adjustments for demographic variables were applied to raw MMSE scores in the previous study, group differences were no longer statistically significant.

Results showed a clear association between verbal learning and memory measures and the MMSE that did not differ across language groups. The strength of the relationship increased across trials. This provides favorable evidence of sensitivity to cognitive impairment, and of lack of strong measurement bias across language groups with respect to identifying cognitive impairment. There is a need for further validation using clinical diagnosis and other, more sensitive measures of cognitive impairment (e.g., specific neuropsychological tests) as criteria. However, the MMSE is well established as an instrument that is sensitive to cognitive impairment in both English and Spanish language versions (Mungas et al., 1996; O'Connor et al., 1989) and the obtained results show strong evidence of sensitivity of the verbal learning test to cognitive impairment, with similar sensitivity for English and Spanish language versions.

The results also showed relationships with other demographic variables. Gender and age were related to performance, as was education, but to a lesser extent. There is a well-established association of age and memory performance (Spar & La Rue, 1997), and there are also previous reports of better verbal learning and memory performance for females (Paolo et al., 1997). Education effects on verbal learning tests are very prominent (Ardila et al., 1994; Artiola i Fortuny et al., 1998), and indeed, the lack of a strong education effect in this study, although also reported with another Hispanic sample (Pontón et al., 1996), is of special interest. Overall, demographic effects on test performance were substantial, which would suggest that these effects should be considered when interpreting test results.

Language of SEVLT administration grouping was based on the participant's preference for testing, which may have affected results. A previous study examining effects of bilingualism on performance on a word-list learning test showed significantly poorer performance for bilingual participants whose level of proficiency was not balanced between English and Spanish (Harris et al., 1995). The sample in that study was much younger (range 21–50 years) and was also substantially smaller ($n = 44$). Effects of bilingualism were examined in this study, and were not statistically significant after accounting for effects of other demographic variables and MMSE score. This suggests that degree of proficiency in English and Spanish was not an important determinant of test scores. Further research with more precise assessment of language proficiency is likely to have important practical implications for memory assessment of older, bilingual persons.

The selection of the words used in the SEVLT was based on prototypicality in English and not Spanish. Ideally, the words for the two lists would have similar prototypicality, frequency of occurrence, and number of phonemes in both languages. The English and Spanish versions of this list, while very similar, cannot be considered to be linguistically equivalent, nor are they necessarily culturally equivalent. However, psychometric equivalence is a critical goal, and indeed, tests that are linguistically or culturally equivalent might still have different measurement characteristics in different populations. Results of this study provide very favorable evidence that the English and Spanish versions of this test have similar measurement properties in English and Spanish and are equally sensitive to an independent measure of global cognitive functioning.

An important consideration in developing normative data from population-based samples of older persons is that dementia is likely to be present to some degree in the sample. The presence of dementia in a sample might be expected to bias regression coefficients for demographic variables that serve as risk factors for dementia. Both age and education have been reported to be risk factors for dementia (Katzman, 1993; Stern et al., 1994), and there is some evidence that gender and ethnicity might also be risk factors (Liu et al., 1998; Tang et al., 1998). Participants with risk factors would then be more likely to be demented and as a result would have lower test scores. Consequently, one might expect that the correlation between the demographic variable (age for example) and the test score might be greater in the full sample including cases with dementia than in a restricted sample of nondemented cases. If the regression co-

https://doi.org/10.1017/S1355617701755026 Published online by Cambridge University Press

efficient is amplified by the presence of dementia, then overcorrection of scores could occur, which might actually decrease the likelihood of detecting dementia.

Traditionally, one would exclude participants with a clinical diagnosis of dementia from the normative sample. However, dementia evaluation is a very costly and resource-demanding process, and is not always practical in normative studies. One might also exclude participants falling below a cut-off score on a screening measure, the MMSE for example. However, this is also problematic. Assuming that a screening test has high sensitivity and specificity values of .90 and that there is a 10% prevalence of dementia, 50% of the screened out cases would be nondemented even though 90% of demented cases would be correctly identified. This would eliminate the low functioning normals, and might consequently positively bias normative data.

The issue of potential contamination of the sample dementia in this study by participants with was addressed in two ways. First, the MMSE was incorporated into regression equations used to derive statistical corrections for demographic variables. Theoretically, this should substantially limit bias in regression coefficients associated with inclusion of participants with dementia since the MMSE provides an effective measure of cognitive impairment of dementia. Second, percentiles are reported for the full sample as well as for a sample in which cases falling below the 10th percentile on the age- and education-adjusted MMSE (Mungas et al., 1996) were eliminated. The most appropriate set of norms may well depend upon the specific purpose for which the test is used, and availability of both sets of norms should provide flexibility for use of this test by clinicians and researchers. There clearly is a need for further research to validate this instrument with other populations and for other applications, which will be critical for establishing optimal cut-offs for detecting clinically relevant conditions.

ACKNOWLEDGMENTS

The research was supported by grants from the National Institute on Aging: AG10220 (D. Mungas), AG12975 (M. Haan), AG10229 (Jagust) and Minority Investigator Research Supplement to AG12975 (H. González) and the California Department of Health Services, 98 14970 (Mungas). The authors thank Teresa Ortiz for her help preparing this manuscript and tireless work directing the Sacramento Area Latino Study on Aging.

REFERENCES

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

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

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

Cahn, D.A., Salmon, D.P., Butters, N., Wiederholt, W.C., Corey-Bloom, J., Edelstein, S.L., & Barrett-Connor, E. (1995). Detection of dementia of the Alzheimer type in a population-based sample: Neuropsychological test performance. *Journal of the International Neuropsychological Society*, *1*, 252–260.

Delis, D.C., Filoteo, J.V., Massman, P.J., Kaplan, E., & Kramer, J.H. (1994). The clinical assessment of memory disorders. In L.S. Cermak (Ed.), *Neuropsychological explorations of memory and cognition: Essays in honor of Nelson Butters. Critical issues in neuropsychology* (pp. 223–239). New York: Plenum Press.

Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B. (1987). *California Verbal Learning Test–Research edition*. San Antonio, TX: The Psychological Corporation.

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.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory* (Vol. 2). Newbury Park, CA: Sage Publications.

Harris, J.G., Cullum, C.M., & Puente, A.E. (1995). Effects of bilingualism on verbal learning and memory in Hispanic adults. *Journal of the International Neuropsychological Society*, *1*, 10–16.

Ivnik, R.J., Malec, J.F., Smith, G.E., & Tangalos, E.G. (1992). Mayo's Older Americans Normative Studies: Updated AVLT norms for ages 56 to 97. *Clinical Neuropsychologist*, *6*, 83–104.

Katzman, R. (1993). Education and the prevalence of dementia and Alzheimer's disease. *Neurology*, *43*, 13–20.

Knopman, D.S. & Ryberg, S. (1989). A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. *Archives of Neurology*, *46*, 141–145.

Liu, C.K., Lai, C.L., Tai, C.T., Lin, R.T., Yen, Y.Y., & Howng, S.L. (1998). Incidence and subtypes of dementia in southern Taiwan: Impact of socio-demographic factors. *Neurology*, *50*, 1572–1579.

Marshall, S.C., Mungas, D., Weldon, M., Reed, B. & Haan, M. (1997). Differential item functioning in the Mini-Mental State Examination in English- and Spanish-speaking older adults. *Psychology & Aging*, *12*, 718–725.

Mungas, D., Marshall, S.C., Weldon, M., Haan, M., & Reed, B.R. (1996). Age and education correction of Mini-Mental State Examination for English and Spanish-speaking elderly. *Neurology*, *46*, 700–706.

Mungas, D., Reed, B.R., Marshall, S.C., & González, H.M. (2000). Development of psychometrically matched English and Spanish language neuropsychological tests for the elderly. *Neuropsychology*, *14*, 209–223.

Mungas, D., Wallace, R., & Reed, B.R. (1998). Dimensions of cognitive ability in dementia: Differential sensitivity to degree of impairment in Alzheimer's disease. *Clinical Neuropsychologist*, *12*, 129–142.

O'Connor D.W., Pollitt, P.A., Hyde, J.B., Fellows, J.L., Miller, N.D., Brook, C.P., & Reiss, B.B. (1989). The reliability and validity of the Mini-Mental State in a British community survey. *Journal of Psychiatric Research*, *23*, 87–96.

Paolo, A.M., Tröster, A.I., & Ryan, J.J. (1997). California Verbal Learning Test: Normative data for the elderly. *Journal of Clinical and Experimental Neuropsychology*, *19*, 220–234.

Pontón, M.O., Satz, P., Herrera, L., & Ortiz, F. (1996). Normative data stratified by age and education for the Neuropsychologi-

cal Screening Battery for Hispanics (NeSBHIS): Initial report. *Journal of the International Neuropsychological Society*, *2*, 96–104.
Rawlings, J.O., Pantula, S.G., Dickey, D.A. (1998). *Applied regression analysis. A research tool* (2nd ed.). New York: Springer-Verlag.
Spar, J.E. & La Rue, A. (1997). *Concise guide to geriatric psychiatry* (2nd ed.). Washington, DC: American Psychiatric Press.
Stern, Y., Gurland, B., Tatemichi, T.K., Tang, M.X., Wilder, D., & Mayeux, R. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. *Journal of the American Medical Association*, *271*, 1004–1010.
Stricks, L., Pittman, J., Jacobs, D.M., Sano, M., & Stern, Y. (1998). Normative data for a brief neuropsychological battery administered to English- and Spanish-speaking community-dwelling elders. *Journal of the International Neuropsychological Society*, *4*, 311–318.
Tang, M.X., Stern, Y., Marder, K., Bell, K., Gurland, B., Lantigua, R., Andrews, H., Feng, L., Tycko, B., & Mayeux, R. (1998). The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *Journal of the American Medical Association*, *279*, 751–755.
Teng, E.L. & Chui, H.C. (1987). The Modified Mini-Mental State (3MS) Examination. *Journal of Clinical Psychiatry*, *48*, 314–318.
US Bureau of the Census. (1991). *US census of population and housing, 1990: Summary population and housing characteristics*. Washington, DC: US Department of the Commerce Economics and Statistics Administration, Bureau of the Census.
US Bureau of the Census. (1993a). *We are the Americans . . . Hispanics*. Washington, DC: US Department of the Commerce Economics and Statistics Administration, Bureau of the Census.
US Bureau of the Census. (1993b). *We are the Americans . . . Elderly*. Washington, DC: US Department of the Commerce Economics and Statistics Administration, Bureau of the Census.
Uyeda, K.M. & Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Behavioral Research Methods and Instrumentation*, *12*, 587–595.
Welsh, K.A., Butters, N., Hughes, J., Mohs, R., & Heyman, A. (1991). Detection of abnormal memory decline in mild cases of Alzheimer's disease using CERAD neuropsychological measures. *Archives of Neurology*, *48*, 278–281.

APPENDIX

SPANISH ENGLISH VERBAL LEARNING TEST (SEVLT) ITEMS

List A		List B	
English	Spanish	English	Spanish
Cabbage	Repollo	Eggs	Huevos
Ladle	Cucharón	Pot	Olla
Coffee	Café	Milk	Leche
Beets	Remolachas	Cherries	Cerezas
Dictionary	Diccionario	Bowl	Tazón
Beans	Frijoles	Lettuce	Lechuga
Strainer	Coladera	Spoon	Cuchara
Peach	Durazno	Water	Agua
Corn	Maíz	Fish	Pescado
Newspaper	Periodico	Pen	Pluma
Juice	Jugo	Oranges	Naranjas
Asparagus	Espárrago	Cookies	Galletas
Pan	Cazuela	Notebook	Libreta
Tea	Té	Onions	Cebollas

https://doi.org/10.1017/S1355617701755026 Published online by Cambridge University Press