



Published in final edited form as:

*J Int Neuropsychol Soc.* 2005 September ; 11(5): 620–630. doi:10.1017/S1355617705050745.

## Criterion-referenced validity of a neuropsychological test battery: Equivalent performance in elderly Hispanics and Non-Hispanic Whites

Dan Mungas<sup>1,2</sup>, Bruce R. Reed<sup>1,2</sup>, Sarah Tomaszewski Farias<sup>1</sup>, and Charles Decarli<sup>1,2</sup>

<sup>1</sup>Department of Neurology, School of Medicine, University of California, Davis, California

<sup>2</sup>Veterans Administration Northern California Health Care System, Martinez, California

### Abstract

This study examined the validity of the Spanish and English Neuropsychological Assessment Scales (SENAS) in comparison with clinical diagnosis of normal cognition *versus* cognitive impairment, not demented (CIND) *versus* demented in elderly Hispanics and Whites. Relationships between SENAS scales and diagnosis were essentially the same in Hispanics and Whites. Verbal memory measures were most strongly related, with more than 35% of the variance in these measures accounted for by diagnosis independent of effects of education, age, gender, and language. Diagnosis accounted for more than 10% of the variance (19% on average) in 11 of the 17 measures examined in this study. Logistic regressions showed that verbal memory was important both for distinguishing normal from CIND and CIND from demented. Object naming improved discrimination of CIND from demented beyond that of verbal memory alone. These results provide evidence of equivalent validity across Hispanics and Whites.

### Keywords

Neuropsychological assessment; Ethnic groups; Cognitive impairment; Dementia; Early diagnosis; Cognition

## INTRODUCTION

Cognitive impairment and dementia are important public health concerns that are amplified by rapidly increasing older populations, especially ethnic minorities. Neuropsychological tests play an important role in clinical diagnosis of these disorders (American Academy of Neurology Therapeutics and Technology Assessment Subcommittee, 1996; Petersen et al., 2001) and are critical research tools in understanding cognitive disorders of aging. However, existing methods for minority populations have not been well studied and validated, and consequently, have important limitations. In particular, factors associated with minority ethnicity, such as low education, language, and cultural differences influence test scores and may lead to mistaken diagnostic decisions (e.g., Gasquoine, 1999; Manly et al., 1998; Ramírez et al., 2001; Stern et al., 1992).

The Spanish and English Neuropsychological Assessment Scales (SENAS) were created to provide psychometrically equivalent measures of multiple cognitive abilities in older

English- and Spanish-speakers. Extensive, large-sample test development and validation work (Mungas et al., 2000, 2004, in press), guided by item response theory (Hambleton & Swaminathan, 1985; Hambleton et al., 1991) underlies the SENAS.

Consistent with previous literature (see Gasquoine, 1999, for review of studies of demographic effects on neuropsychological test results in Hispanics), our previous work (Mungas et al., in press) showed that education and language influence SENAS scores, though effects varied across scales. Education was most strongly related to semantic memory, and was least related to episodic memory. Education effects were essentially the same in Whites and Hispanics. English proficiency was positively correlated with test results while Spanish proficiency had negative correlations. These effects were strongest for verbal scales and the nonverbal semantic memory test, were moderate for nonverbal scales, and weak for episodic memory. Acculturation effects were significant in Hispanics, but acculturation effects independent of education and language utilization were small. After controlling for education and language, mean ethnicity effects were small and acculturation was unrelated to test scores.

The goal of this study was to further validate SENAS measures against the important criterion of clinical diagnosis. As potential disease-modifying treatments are explored, there is considerable interest in early detection of cognitive impairment that might progress to dementia. Consequently, the ability of neuropsychological tests to distinguish among normal cognition, cognitive impairment, not dementia (CIND; Di Carlo et al., 2000; Graham et al., 1997; Unverzagt et al., 2001) and dementia is important. For tests used in multi-ethnic settings, an important component of validation addresses the extent to which results are equally valid in different groups. Thus, we examined the extent to which SENAS scores were related to clinical diagnosis within each ethnic group, and whether relationships of test scores to diagnosis were the same in Whites and Hispanics. A secondary goal was to identify which cognitive domains are important for discriminating the three levels of cognitive impairment.

## METHODS

### Research Participants

Participants were 154 persons with cognitive syndrome diagnoses established through the UC Davis Alzheimer's Disease Center (UCD-ADC). Recruitment was designed to target ethnic minorities, to maximize heterogeneity of demographic characteristics, and to emphasize normal cognition and mild impairment. Consequently, 98 (62 Hispanics, H, 36 Whites, C) participants were recruited through direct community outreach *via* a community hospital lobby, a community survey, health fairs, or word of mouth. There were 49 normals, 29 diagnosed with CIND, and 20 diagnosed as demented. The remaining 56 participants (6 H 50 C) were patients at the UCD ADC (13 normal, 29 CIND, 14 demented). Regardless of recruitment source, inclusion criteria were over age 60, White or Hispanic ethnicity, and cognitive function of mild dementia or better. Exclusion criteria included unstable major medical illness, major primary psychiatric disorder, and substance abuse or dependence within 5 years. All participants signed informed consent under protocols approved by institutional review boards at UC Davis, the Veterans Administration Northern California Health Care System, and San Joaquin General Hospital in Stockton, California.

Participants self-identified ethnic group membership. Approximately 80% of the Hispanics were of Mexican origin. Language of test administration was the participants' own preferred language unless their non-preferred language was used for more daily activities. Forty-one Hispanics were monolingual Spanish speakers. Seven were monolingual English speakers, and 20 were bilingual. All Whites spoke English as their primary language.

## SENAS Measures

Table 1 shows the domains measured by the SENAS and the specific measures of each domain. Scales and psychometric characteristics are described in more detail elsewhere (Mungas et al., 2004). The episodic memory measures were composite measures created using item response theory methods, and included scores from learning trials as well as delayed free recall.

Measures of two aspects of executive function, fluency and working memory, were also used. Fluency measures were Category Fluency (number of animals named in 60s), Phonemic Fluency (words beginning with the /f/ sound, words beginning with the /l/ sound), and number of total items and number of categories from the Supermarket Test (Mattis, 1988). Working memory measures included Digit Span Backward and Visual Span Backward, as well as a new List Sorting task. In Part 1 of List Sorting, participants are presented with a list of either fruits or animals and are asked to repeat all of the items on the list in order from smallest to largest. In Part 2, the lists include both fruits and animals and the task is to repeat fruits first, sorted from smallest to largest, and then animals in order from smallest to largest.

These measures were combined into homogeneous composite scales using item response theory methods. Confirmatory factor analyses based on a multi-ethnic sample ( $N=542$ ) showed good model fit for conceptually derived subscales of Category Fluency, Phonemic Fluency, and Working Memory (see Table 1). The homogeneous subscales were highly correlated and were well accounted for by a second order executive function factor. Both the subscales and an Executive Composite based on all executive function measures were used in this study. All SENAS scores were presented in *z*-score like units where a score of zero corresponded to the mean of a demographically diverse, non-demented normative sample composed primarily of Hispanics and Whites and differences from the mean were expressed in standard deviation units.

## Language Usage

Each participant rated his or her ability to speak English and Spanish on a 4-point scale and the ratings were combined into a single language usage variable. A score of 3 corresponded to monolingual English proficiency, a score of -3 to monolingual Spanish, and zero to bilingual with equal proficiency in English and Spanish.

## Clinical Evaluation

All participants received a multidisciplinary clinical evaluation at the UCD ADC including a detailed medical history, physical exam, and neurological exam. A bilingual physician examined Spanish-speaking patients. A family member or informant with close contact with the participant was interviewed to obtain information about level of independent functioning. Diagnostic neuroimaging and routine dementia work-up laboratory tests were a standard part of the protocol.

All participants received a clinical neuropsychological evaluation using standard neuropsychological tests. This battery was comprised of the CERAD neuropsychological battery (Welsh et al., 1992, 1994) supplemented by WAIS-R Digit Symbol (Wechsler, 1981) and the Trail Making Test. Norms from Fillenbaum et al. (2001) were used for the tests from the CERAD battery. These norms are for African Americans and Whites and incorporate adjustments for education and age. The African American norms were used for Hispanics in this study. This is not optimal, but acceptable norms for older Hispanics are limited. To help compensate for this limitation, local norms based upon non-demented individuals recruited by the UCD ADC were used in addition. These norms included

adjustments for age, education, and language of test administration and are based on samples of about 30 non-demented Whites and 50 Hispanics recruited from community settings. The decision about whether there was significant cognitive impairment was a clinical judgment that was guided by but not algorithmically linked to either set of norms. Informant report was also considered in evaluating cognitive functioning, and was particularly important when formal test results were equivocal or when there was disagreement depending on which norms were used.

Independent functioning was evaluated using the Blessed-Roth Dementia Rating Scale (BRDRS; Blessed et al., 1968, 1988) based upon an interview with an informant. Spanish speaking informants were interviewed by bilingual staff.

Cognitive syndrome (*normal, CIND, demented*) and, in the instance of dementia, underlying etiology was diagnosed according to standardized criteria and methods. Each case was initially diagnosed by the clinical team at a consensus conference. Those appearing likely to be eligible for this study were then reviewed at a second, research case adjudication conference with broader participation. Diagnosis was based upon all available clinical information (excluding SENAS results). Dementia was diagnosed based upon DSM–III–R (American Psychiatric Association, 1987) criteria for dementia and the dementia criteria in the California ADDTC diagnostic criteria for ischemic vascular dementia (Chui et al., 1992). DSM–III–R criteria require impairment of memory plus one other cognitive domain, while the ADDTC criteria do not require memory impairment if there is impairment of two or more cognitive domains. CIND was diagnosed if the person did not meet diagnostic criteria for dementia, but had clinically significant impairment in at least one cognitive domain.

## Data Analysis

Demographic characteristics of ethnic and diagnostic groups were compared using analysis of variance for continuous variables and logistic regression for categorical variables. Two different types of analyses were used to evaluate the relationship of SENAS variables with clinical syndrome. The first was analyses of variance with SENAS scales as dependent variables and clinical syndrome as the primary independent variable. The second utilized logistic regression analyses in which clinical syndrome was the dependent variable, and demographic and language covariates, ethnicity, and SENAS scales were independent variables. The logistic regression analyses were performed to address which tests were most important for discriminating the specific diagnostic categories.

For the analysis of variance approach, an initial multivariate analysis of variance (MANOVA) was conducted with 11 SENAS measures that had complete data for all 154 cases in this sample (Picture Association, Object Naming, Pattern Recognition, Verbal Attention Span, Verbal Conceptual Thinking, Word List Learning I, Word List Learning II, Executive Composite, Category Fluency, Phonemic Fluency, Working Memory). These 11 SENAS measures were first entered as dependent variables into a multivariate general linear model with clinical syndrome diagnosis (Clinical Syndrome) as the primary independent variable of interest. Education, gender, age, language usage, and ethnicity were included as covariates to control for confounding effects of these variables. A term was also included to account for the interaction of Ethnicity  $\times$  Clinical Syndrome. The Clinical Syndrome main effect was a particularly important test of the concurrent validity of the SENAS scales. The Ethnicity  $\times$  Clinical Syndrome interaction assessed differential validity across groups and so was an important index of measurement bias. Ideally, the scales should relate strongly to Clinical Syndrome, and this relationship should not differ across ethnic groups.

Univariate ANOVAS for each individual SENAS scale were performed to estimate effect sizes. Incremental Clinical Syndrome effect sizes were estimated by adding this variable to a

baseline model that included demographics, language, and ethnicity, and subtracting the baseline model  $R^2$  from the  $R^2$  associated with the baseline model plus Clinical Syndrome. Incremental effects of Ethnicity  $\times$  Clinical Syndrome were defined as the variance explained by this interaction effect beyond that accounted for by all other effects excluding the interaction. Ethnicity effects were estimated as the increase in  $R^2$  associated with adding ethnicity to a model with other demographics and Clinical Syndrome.

Secondary univariate analyses were performed for the six scales that did not have complete data for all cases using the same methods to estimate effect sizes. Sample sizes for these analyses varied: Non-Verbal Conceptual Thinking:  $n=100$ ; Spatial Localization:  $n=148$ ; Verbal Comprehension:  $n=101$ ; Verbal Expression:  $n=97$ ; Visual Attention:  $n=88$ ; Spatial Configuration Learning:  $n=93$ . Statistical significance of effects for individual SENAS scales was determined using a Bonferroni-corrected  $p$  value ( $.0029 = .05017$ ).

Polytomous logistic regression was used in analyses with Clinical Syndrome as the dependent variable; CIND was the reference group against which both normal and demented were compared. The 11 SENAS measures with complete data were the primary independent variables. Demographic and language variables were included as covariates. Each SENAS scale was first entered alone into a separate model along with covariates that also included an Ethnicity  $\times$  SENAS variable interaction term. A Bonferroni corrected  $p$  value of .0023 (.05022; two comparisons for each of 11 scales) was used. Then, individual SENAS measures that were significantly associated with Clinical Syndrome were entered jointly along with Word List Learning I to evaluate which measures made incremental contributions to Clinical Syndrome beyond effects of verbal memory. Finally, logistic regressions were performed using SENAS variables that performed well in previous analyses to discriminate dichotomous categories of normal *versus* CIND and CIND *versus* demented. Receiver operator curve (ROC) analyses were performed, and the area under the ROC curve was used as a metric to compare various models. Diagnostic sensitivity associated with 80% specificity was used as another metric of clinical sensitivity.

## RESULTS

### Demographic Variables and MMSE

Table 2 shows demographic variables, global cognitive status (MMSE), and functional status (BRDRS) by ethnic group and Clinical Syndrome. Gender did not differ according to Clinical Syndrome, ethnicity, or their interaction ( $p$ 's  $> .07$ ). Education differed substantially by ethnicity [ $F(1,148)=131.7, p < .0001$ ;  $M$  education = 6.3 years for Hispanics vs. 14.3 years for Whites], but did not differ by Clinical Syndrome ( $p = .73$ ) or the Clinical Syndrome  $\times$  Ethnicity interaction ( $p = .62$ ). Mean age was older in Whites [ $F(1,148)=12.2, p < .0006$ ;  $M$  H = 72.8, C = 77.2], and the Clinical Syndrome effect for age was significant [ $F(2,148) = 5.0, p < .008$ ;  $M$  normal = 72.7, CIND=74.5, demented = 77.7]. The interaction was not significant ( $p = .97$ ). MMSE differed by ethnicity [ $F(1,148)=58.2, p < .0001$ ], Clinical Syndrome [ $F(2,148)=66.8, p < .0001$ ], and the interaction [ $F(2,148) = 5.5, p < .005$ ]. MMSE significantly differed across Clinical Syndrome groups in Hispanics [ $F(2,65) = 30.8, p < .0001$ ] and Whites [ $F(2,83)=45.5, p < .0001$ ]. BRDRS differed by ethnicity [ $F(1,144)=4.7, p < .04$ ] and Clinical Syndrome [ $F(2,144)=49.4, p < .0001$ ], but the interaction was not significant ( $p > .08$ ). BRDRS significantly differed across Clinical Syndrome groups in Hispanics [ $F(2,61) = 22.0, p < .0001$ ] and Whites [ $F(2,83)=28.2, p < .0001$ ]. Of note, demented Hispanics scored significantly higher than demented Whites ( $p = .015$ ) but Hispanic–White differences for CIND and normals did not differ ( $p$ 's = .65 and .12, respectively).

## Clinical Diagnosis and SENAS Scores

The MANOVA used to evaluate independent effects of Clinical Syndrome, demographics, language, and ethnicity yielded a highly significant Clinical Syndrome main effect, averaged across scales [ $F(2,144) = 54.3, p < .0001$ ] that accounted for approximately 43% of the overall variance in the 11 SENAS measures included in the primary analysis. The ethnicity main effect was significant [ $F(1,144)=7.4, p < .008$ ] and accounted for about 5% of the SENAS variance. The Ethnicity  $\times$  Clinical Syndrome interaction was not significant ( $F < 1.0$ ). The Scales  $\times$  Clinical Syndrome interaction was significant [approximate  $F(20,270)=5.6, p < .0001$ ], indicating that the Clinical Syndrome effect differed across scales. The three-way Scales  $\times$  Ethnicity  $\times$  Syndrome Diagnosis interaction was not significant [approximate  $F(20,270) = 1.3, p > .19$ ], indicating that Clinical Diagnosis effect did not differ for Hispanics and Whites for any SENAS scale.

Effect sizes derived from univariate ANOVAS are presented in Table 3. Clinical Syndrome was related to 12 of the 17 SENAS scales using a Bonferroni-corrected  $p$  value ( $p = .05017 = .0029$ ), and these effects were independent of demographic and language variables. Clinical Syndrome incrementally explained at least 10% of the variance of 11 of the 17 SENAS measures (19%, on average), explained about 20% for Category Fluency and Spatial Configuration Learning, and accounted for more than 35% of the verbal memory measures. The Ethnicity  $\times$  Clinical Syndrome interaction was significant at an uncorrected  $p$  value for two scales, Word List Learning I [ $F(2,144)=4.1, p < .02$ ] and Word List Learning II [ $F(2,144)=3.1, p < .05$ ], although the amount of variance explained, 2.4% and 2.0%, was small. The Clinical Syndrome effect size was about 13 times the combined effects of ethnicity and the Ethnicity  $\times$  Clinical Syndrome interaction for both verbal memory measures. Ethnicity effects were 2% or less with the exception of Verbal Comprehension, the Executive Composite, and Phonemic Fluency, and ethnicity independently accounted for less than 5% of the variance of these variables.

*Post-hoc* comparisons were performed to assess differences in each SENAS measure between Normal and CIND, and between CIND and Demented for the 11 variables included in the primary analysis. Bonferroni-corrected  $p$  values ( $.05/22 = .0023$ ; two comparisons/11 measures) were used to determine statistical significance. Both comparisons were significant for six SENAS measures (Object Naming, Verbal Conceptual Thinking, Word List Learning I, Word List Learning II, Executive Composite, Category Fluency). Figure 1 shows Clinical Syndrome group differences in average Word List Learning I and Object Naming scores for Hispanics and Whites. The CIND *versus* demented comparison was significant for Picture Association and Working Memory, and neither comparison was significant for Pattern Recognition, Verbal Attention Span, and Phonemic Fluency.

Table 4 presents raw and covariate adjusted means by ethnic group and Clinical Syndrome for the six scales that discriminated both normal from CIND and CIND from demented for both ethnic groups. Table 5 shows effects of demographic and language variables used as covariates for these six scales. Covariates accounted for significant variability in raw scores, but the pattern of effects of specific demographic and language variables differed across scales.

## SENAS Predictors of Clinical Syndrome

Results from polytomous logistic regression in which Clinical Syndrome was the dependent variable were essentially the same as for the analyses in which SENAS measures were dependent variables. The six measures from the previous analysis (Object Naming, Verbal Conceptual Thinking, Word List Learning II, Word List Learning I, Executive Composite, Category Fluency) and Picture Association significantly discriminated CIND from both

normal and demented, and Working Memory discriminated CIND from demented. None of the SENAS by ethnicity interaction terms were significant using an uncorrected  $p$  value of .05.

The two verbal memory measures were by far most strongly related to Clinical Diagnosis. Additional analyses were performed in which Word List Learning I was included in a baseline logistic regression model along with demographic and language variables, and other SENAS scales were added individually to assess incremental effects beyond effects of verbal memory. The six non-memory variables with strong effects in the previous analyses were used, and a Bonferroni corrected  $p$  value of .0042 (.05012) was used. Object Naming improved discrimination of CIND from Demented, but none of the other measures significantly improved discrimination beyond that provided by Word List Learning I.

The joint contributions to diagnostic sensitivity of Word List Learning I and Object Naming were explored in a final analysis. Two dichotomous comparisons were made, normal *versus* CIND and CIND *versus* demented. Sequential logistic regression models were used to account for these two types of diagnostic discriminations. The area under the ROC curve for each comparison and the diagnostic sensitivity associated with specificity of .80 were the primary outcomes of interest. In the first model, demographic and language variables were included as independent variables. Word List Learning I was added in Model 2, and Object Naming was added along with Word List Learning I in Model 3. Figure 2 shows areas under the ROC curves associated with these Models. Figure 3 shows diagnostic sensitivity associated with 80% specificity for these analyses. Verbal memory markedly improved both types of discrimination over that provided by demographic and language variables, which included age. Object Naming improved discrimination beyond that obtained using demographics and verbal memory, especially for the CIND from Demented distinction. Word List Learning I and Object Naming combined yielded better than 80% sensitivity for 80% specificity for both comparisons.

## DISCUSSION

This study examined the association of SENAS scores with independently diagnosed cognitive syndrome in Hispanics and Whites. Twelve of the 17 scales significantly differed across Clinical Syndrome groups, and for 11 of these scales Clinical Syndrome accounted for at least 10% of the variance after controlling for demographic (including ethnicity) and linguistic effects. Effects of Clinical Syndrome differed across Hispanics and Whites for the two verbal learning measures, but effect sizes were small and group differences in Clinical Syndrome effects were minor in comparison with common effects of Clinical Syndrome. A majority of measures discriminated normal from CIND and CIND from demented, demonstrating criterion related validity.

A potential limitation of this study is that the primary criterion for external validation, clinical diagnosis, might be subject to demographically related measurement biases. For example, low education and minority ethnicity would be expected to affect the clinical neuropsychological tests used in diagnosis (as they do SENAS scores), and this might spuriously inflate the relationship of SENAS scores with clinical diagnosis. There are compelling reasons to believe that this sort of confounding cannot explain the results of this study.

First, a previous study showed that SENAS scores were strongly related to independent measures of global cognition and independent function after controlling for education, language, and other demographic covariates (Mungas et al., in press). That study was based on a different sample, and used different criterion measures of cognition and independent

function. Results were very similar to those of the present study. These studies together provide converging evidence using different samples and outcome criteria that SENAS measures are comparably sensitive to clinically important differences in cognitive ability and functional status in both Hispanics and Whites.

A second point is that the validity of the clinical diagnosis in this study is supported by results of an informant-based measure of independent function, the BRDRS. Hispanic CIND cases had a mean BRDRS score that did not differ from that of White CIND cases, and demented Hispanics showed greater functional impairment than demented Whites. This shows that Hispanics with cognitive impairment and dementia had at least as much functional impairment as similarly diagnosed Whites, and argues against the hypothesis that Hispanics were diagnosed as impaired because of spuriously low neuropsychological test scores.

A second potential limitation to this study is the sampling, and specifically that the vast majority of Hispanics were recruited from the community while more than half of the Whites were referrals from a dementia specialty clinic. This could result in differences between Hispanics and Whites that could reflect sampling methods as opposed to real group differences. The similarity of results in these two ethnic groups argues against differential selection bias, which would be expected to enhance group differences.

The sample size of this study also presents limitations. The number of participants within cells defined by ethnicity and Clinical Syndrome was not large. This could decrease statistical power for detecting effects, especially for the measures that did not have complete data. Statistical power would particularly be an issue for the Ethnicity 3 Clinical Syndrome interaction effects. This study examined effect size in addition to statistical significance of results, and showed that the interaction effects were small, even in the few cases where there was a significant interaction effect. Small samples also raise concerns about reliability or replicability of results, and there is a need for replication of these findings and further validation with additional samples.

Results of this study showed differences among SENAS scales in sensitivity to diagnosis of Clinical Syndrome. Verbal memory measures were most strongly related, explaining about 35% of the variance. Nonverbal memory and category fluency showed the next strongest relationships with Clinical Syndrome, sharing about 20% of the variance. Several measures were in the 10–15% range including verbal and nonverbal measures of semantic memory, verbal abstraction, spatial perception, the executive function composite measure, and working memory. It should be noted that these figures are for incremental variance explained after the effects of demographic variables are subtracted; this is a result that is not often reported, and it is obtained here in the context of a sample that has great demographic diversity.

That memory scores were especially sensitive to Clinical Syndrome is hardly surprising given the predominant role that memory plays in defining dementia, Alzheimer's disease and its prodrome. The results among the SENAS scales broadly mirrors the commonly reported findings for mild to moderate AD. Measures of episodic memory are strongly related to diagnosis, language, spatial, and executive measures are related moderately, and simple attentional tasks such as digit span are related weakly at most.

The sensitivity of the scales to cognitive impairment is most directly addressed in the ROC analyses. Memory alone was 80% sensitive with specificity in the 70% range for two clinically important comparisons: normal *versus* CIND, and CIND *versus* demented. When other SENAS measures were added Object Naming made the strongest incremental contributions (in contrast to a previous study (Testa et al., 2004)) and was especially



effective for differentiating CIND from demented. The difference between these results and those of the Testa et al. study could reflect sample differences or psychometric differences. Alzheimer's disease was the predominant clinical disorder in Testa et al. while this study had more varied diagnoses. The Testa et al. sample also was relatively homogeneous demographically, which may have resulted in limited variability in object naming ability in comparison with memory measures. Finally, the measure of object naming used in that study, the Boston Naming Test (Kaplan et al., 1983), may be less sensitive to mild changes than the SENAS Object Naming scale, which was specifically designed and constructed to have high-end sensitivity.

The combination in the present study of Word List Learning I and Object Naming had sensitivity of better than 80% associated with 80% specificity for both CIND *versus* demented and normal *versus* CIND. These sensitivity and specificity values compare well with those previously reported for other, well recognized neuropsychological tests. For example, compared to the values reported by De Jager et al. (2003), the two SENAS scales performed nearly as well as the best tests (memory measures) they studied in differentiating normal from AD, and the SENAS scales performed substantially better than any of their measures in differentiating normal from mild cognitive impairment.

This study is unique in comprehensively addressing the validity of neuropsychological tests both within and between ethnic groups. Previous studies have examined average Hispanic-White or English-Spanish differences in neuropsychological tests in community based (La Rue et al., 1999; Rey et al., 1999) and demented samples (Hohl et al., 1999; Loewenstein et al., 1993), and a few studies have compared normal with impaired or demented Spanish speakers (Arnold et al., 1998; Campo et al., 2003; Taussig et al., 1996). One study (Mulgrew et al., 1999) examined the relative validity of the Mini-Mental State Examination (Folstein et al., 1975) for detecting cognitive impairment in Hispanics and non-Hispanic whites. The present study looked at differences between Hispanics and Whites not only in terms of average scale scores for each group, but also directly compared the effects of Clinical Syndrome across Hispanics and Whites. Results generally showed that Clinical Syndrome effects were large in comparison to ethnic differences in these effects, and this is an important criterion for determining utility of tests in cross-cultural applications.

Another important feature of this study was the inclusive sampling of heterogeneity in cognitive function in a clinically important range. Demented patients were compared not only with healthy high functioning control subjects, as in previous studies (Campo et al., 2003; Taussig et al., 1996), but were compared with CIND and with demographically heterogeneous normals. CIND is by definition intermediate to the poles of normal and dementia, and thus creates a challenge to differentiate. Further, most of the sample was recruited *via* community outreach, and diagnosis in this group is unfiltered by the process of referrals to a university dementia center. Thus, the validity test in this study was particularly stringent, but also particularly relevant to many potential applications of neuropsychological testing.

A final strength of this study is that the demographic heterogeneity was exceptional. Education ranged from no formal education to doctoral degrees. About 25% of the sample were monolingual Spanish speakers, and there was a 50-year age span. An important consequence of this broad variability is that the full range of demographic and linguistic effects on cognition could be observed. This increases statistical power to detect effects, and also enhances generalizability of results for diverse populations in comparison with studies with relatively homogeneous non-minority samples. The sample heterogeneity in this study also presents methodological challenges. Demographic heterogeneity introduces confounding effects on cognitive test scores that potentially could distort results. However,

the design of this study enabled estimation of Clinical Syndrome effects independent of confounding variables that have been shown to be important in previous research (Mungas et al., in press). Another potential limitation is that there may be interactive effects of demographic and language variables, and the sample size in this study was not sufficiently large to examine these effects. Interactive effects of age and education with ethnicity were examined in a previous study with a larger sample (Mungas et al., in press) and were found to be small, but further research is needed.

Measurement bias is a critical concern in cross-cultural neuropsychological assessment. Bias essentially refers to differential validity across groups. Validity is not a generic property of a test and must be evaluated in the context of the expected purpose of the test. Consequently, bias too must be evaluated with respect to a specific, intended use. A key point is that mean group differences in raw test scores (such as are found between Whites and Hispanics on the SENAS) do not necessarily mean that a test is biased for a particular purpose, such as detecting cognitive impairment associated with diseases of aging. If the test is equally sensitive to disease effects, and if mean differences can be adequately accounted for in norms and empirical guidelines for interpretation, then a test with mean differences across groups is an effective and unbiased instrument for this specific purpose. The SENAS shows great promise in this regard.

SENAS test materials are available upon request from the authors. Normative data for the 60+ age range is currently available for a sample of approximately 700 Hispanics (500 tested in Spanish) and 350 Whites.

## Acknowledgments

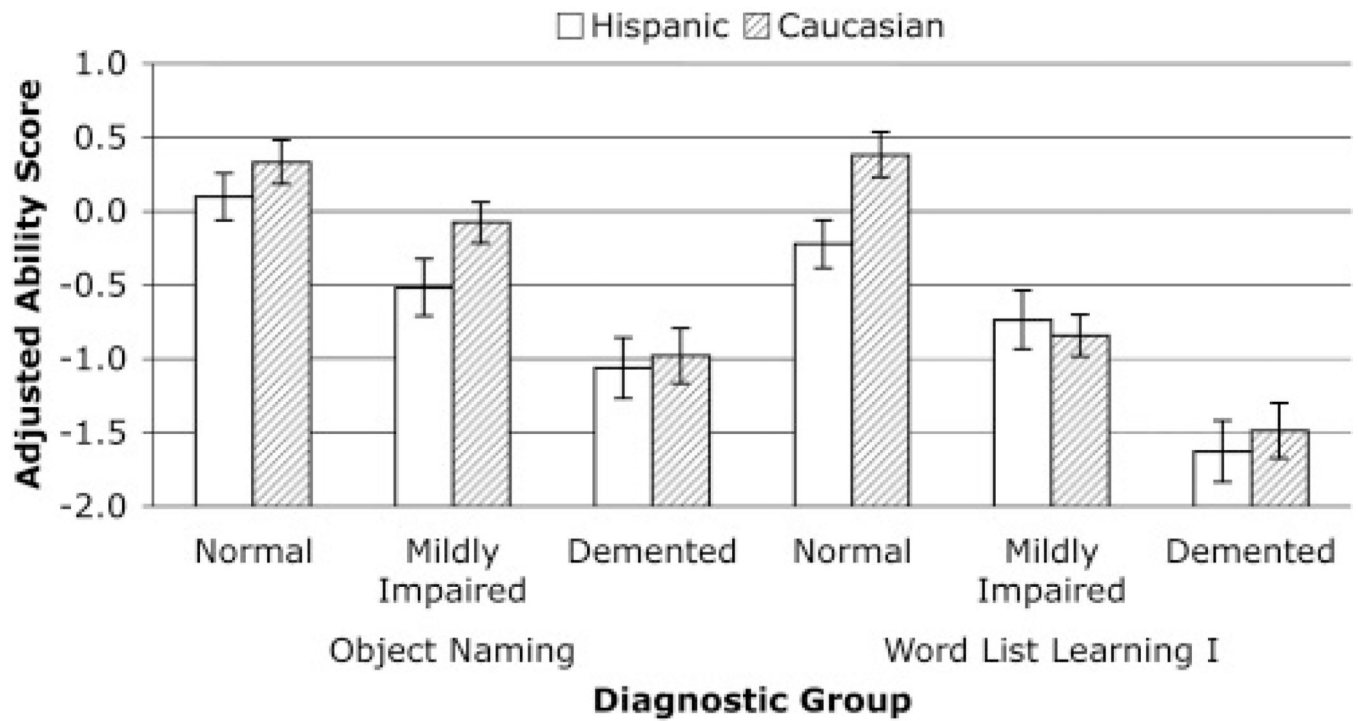
This study was supported in part by Grants AG10220, AG10129, and AG021028 from the National Institute on Aging, Bethesda, MD. Esther Lara supervised recruitment and data collection. Esther Barajas-Ochoa, Cendy Carasco, Gwen Gates, and Nancy Gubbins recruited participants and did SENAS and clinical neuropsychological testing.

## REFERENCES

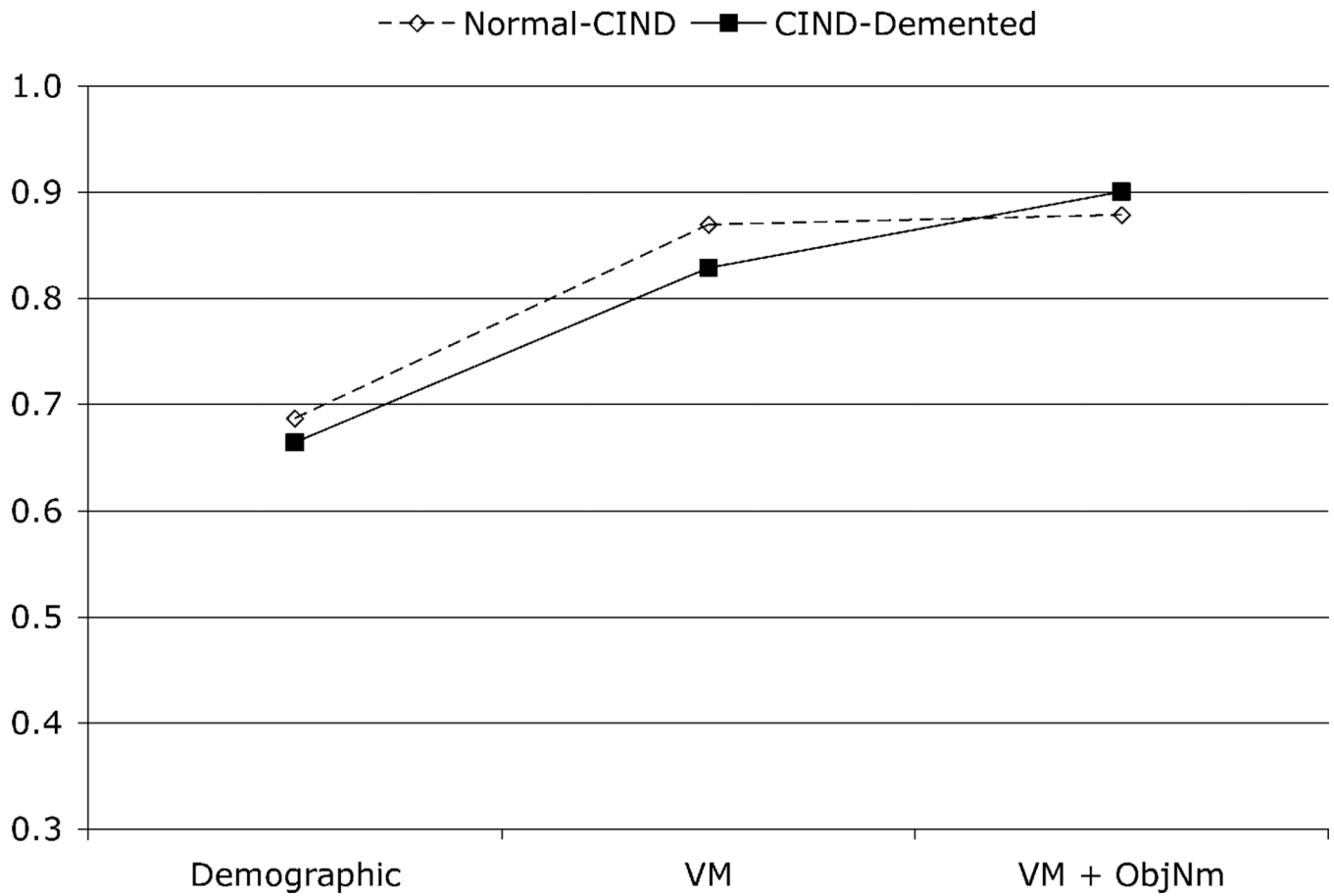
- American Academy of Neurology Therapeutics and Technology Assessment Subcommittee. Assessment: Neuropsychological testing of adults. Considerations for neurologists. *Neurology*. 1996; 47:592–599. [PubMed: 8757049]
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Rev. 3rd ed. Washington, DC: Author; 1987.
- Arnold BR, Cuellar I, Guzman N. Statistical and clinical evaluation of the Mattis Dementia Rating Scale- Spanish adaptation: An initial investigation. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 1998; 53:P364–369.
- Blessed GT, Roth BE, Tomlinson M. The association between quantitative measures of dementia and of senile changes in the cerebral grey matter of elderly subjects. *British Journal of Psychiatry*. 1968; 114:797–811. [PubMed: 5662937]
- Blessed G, Tomlinson BE, Roth M. Blessed-Roth Dementia Scale (DS). *Psychopharmacology Bulletin*. 1988; 24:705–708. [PubMed: 3249772]
- Campo P, Morales M, Martinez-Castillo E. Discrimination of normal from demented elderly on a Spanish version of the verbal selective reminding test. *Journal of Clinical and Experimental Neuropsychology*. 2003; 25:991–999. [PubMed: 13680445]
- Chui HC, Victoroff JI, Margolin D, Jagust W, Shankle R, Katzman R. Criteria for the diagnosis of ischemic vascular dementia proposed by the state of California Alzheimer's disease diagnostic and treatment centers. *Neurology*. 1992; 42:473–480. [PubMed: 1549205]
- De Jager CA, Hogervorst E, Combrinck M, Budge MM. Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychological Medicine*. 2003; 33:1039–1050. [PubMed: 12946088]

- Di Carlo A, Baldereschi MLA, Maggi S, Grigeletto F, Scarlato G, et al. Cognitive impairment without dementia in older people: Prevalence, vascular risk factors, impact on disability. The Italian longitudinal study on aging. *Journal of American Geriatric Society*. 2000; 48:775–782.
- Fillenbaum GG, Heyman A, Huber MS, Ganguli M, Unverzagt FW. Performance of elderly African American and White community residents on the CERAD Neuropsychological Battery. *Journal of the International Neuropsychological Society*. 2001; 7:502–509. [PubMed: 11396552]
- Folstein M, Folstein S, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12:189–198. [PubMed: 1202204]
- Gasquoine PG. Variables moderating cultural and ethnic differences in neuropsychological assessment: The case of Hispanic Americans. *Clinical Neuropsychology*. 1999; 13:376–383.
- Graham JE, Rockwood K, Beattie BL, Eastwood R, Gauthier S, Tuokko H, et al. Prevalence and severity of cognitive impairment with and without dementia in an elderly population. *Lancet*. 1997; 349:1793–1796. [PubMed: 9269213]
- Hambleton, RK.; Swaminathan, H. Item response theory. Principles and applications. Boston: Kluwer-Nijhoff Publishing; 1985.
- Hambleton, RK.; Swaminathan, H.; Rogers, HJ. Fundamentals of item response theory. Newbury Park, CA: Sage Publications; 1991.
- Hohl U, Grundman M, Salmon DP, Thomas RG, Thal LJ. Mini-mental state examination and Mattis Dementia Rating Scale performance differs in Hispanic and non-Hispanic Alzheimer's disease patients. *Journal of the International Neuropsychological Society*. 1999; 5:301–307. [PubMed: 10349293]
- Kaplan, E.; Goodglass, H.; Weintraub, S. Boston Naming Test. (Rev. 60-item version). Philadelphia: Lea & Febiger; 1983.
- La Rue A, Romero LJ, Ortiz IE, Liang HC, Lindeman RD. Neuropsychological performance of Hispanic and non-Hispanic older adults: An epidemiologic survey. *Clinical Neuropsychologist*. 1999; 13:474–486. [PubMed: 10806461]
- Loewenstein DA, Arguelles T, Barker WW, Duara R. A comparative analysis of neuropsychological test performance of Spanish-speaking and English-speaking patients with Alzheimer's disease. *Journal of Gerontology: Psychological Science*. 1993; 48:P142–P149.
- Manly JJ, Jacobs DM, Sano M, Bell K, Merchant CA, Small SA, et al. Cognitive test performance among nondemented elderly African Americans and whites. *Neurology*. 1998; 50:1238–1245. [PubMed: 9595969]
- Mattis, S. Dementia Rating Scale. Odessa, FL: Psychological Assessment Resources; 1988.
- Mulgrew CL, Morgenstern N, Shetterly SM, Baxter J, Baron AE, Hamman RF. Cognitive functioning and impairment among rural elderly Hispanics and non-Hispanic whites as assessed by the mini-mental state examination. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 1999; 54:P223–P230.
- Mungas D, Reed BR, Crane PK, Haan MN, González H. Spanish and English Neuropsychological Assessment Scales (SENAS): Further development and psychometric characteristics. *Psychological Assessment*. 2004; 16:347–359. [PubMed: 15584794]
- Mungas D, Reed BR, Haan MN, González H. Spanish and English Neuropsychological Assessment Scales (SENAS): Relationship to demographics, language, cognition, and independent function. *Neuropsychology*. (in press).
- Mungas D, Reed BR, Marshall SC, González HM. Development of psychometrically matched English and Spanish neuropsychological tests for older persons. *Neuropsychology*. 2000; 14:209–223. [PubMed: 10791861]
- Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review). Report of the quality standards subcommittee of the American Academy of Neurology. *Neurology*. 2001; 56:1133–1142. [PubMed: 11342677]
- Ramírez M, Teresi JE, Silver S, Holmes D, Gurland B, Lantigua R. Cognitive assessment among minority elderly: Possible test bias. *Journal of Mental Health and Aging*. 2001; 7:91–118.

- Rey GJ, Feldman E, Rivas-Vazquez R, Levin BE, Benton A. Neuropsychological test development and normative data on Hispanics. *Archives of Clinical Neuropsychology*. 1999; 14:593–601. [PubMed: 14590573]
- Stern Y, Andrews H, Pittman J, Sano M, Tatemichi T, Lantigua R, et al. 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*. 1992; 49:453–460. [PubMed: 1580806]
- Taussig IM, Mack WJ, Henderson VW. 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*. 1996; 2:286–298. [PubMed: 9375177]
- Testa JA, Ivnik RJ, Boeve B, Petersen RC, Pankratz VS, Knopman D, Tangalos E, Smith GE. Confrontation naming does not add incremental diagnostic utility in MCI and Alzheimer's disease. *Journal of the International Neuropsychological Society*. 2004; 10:504–512. [PubMed: 15327729]
- Unverzagt FW, Gao S, Baiyewu O, Ogunniyi AO, Gureje O, Perkins A, et al. Prevalence of cognitive impairment: Data from the Indianapolis study of health and aging. *Neurology*. 2001; 57:1655–1662. [PubMed: 11706107]
- Wechsler, D. Wechsler Adult Intelligence Scale–Revised. San Antonio, TX: The Psychological Corporation; 1981.
- Welsh KA, Butters N, Hughes J, Mohs R, Heyman A. Detection and staging of dementia in Alzheimer's disease: Use of neuropsychological measures developed for the consortium to establish a registry for Alzheimer's disease (CERAD). *Archives of Neurology*. 1992; 49:448–452. [PubMed: 1580805]
- Welsh KA, Butters N, Mohs RC, Beekly B, Edland S, Fillenbaum G, et al. The consortium to establish a registry for Alzheimer's disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology*. 1994; 44:609–614. [PubMed: 8164812]

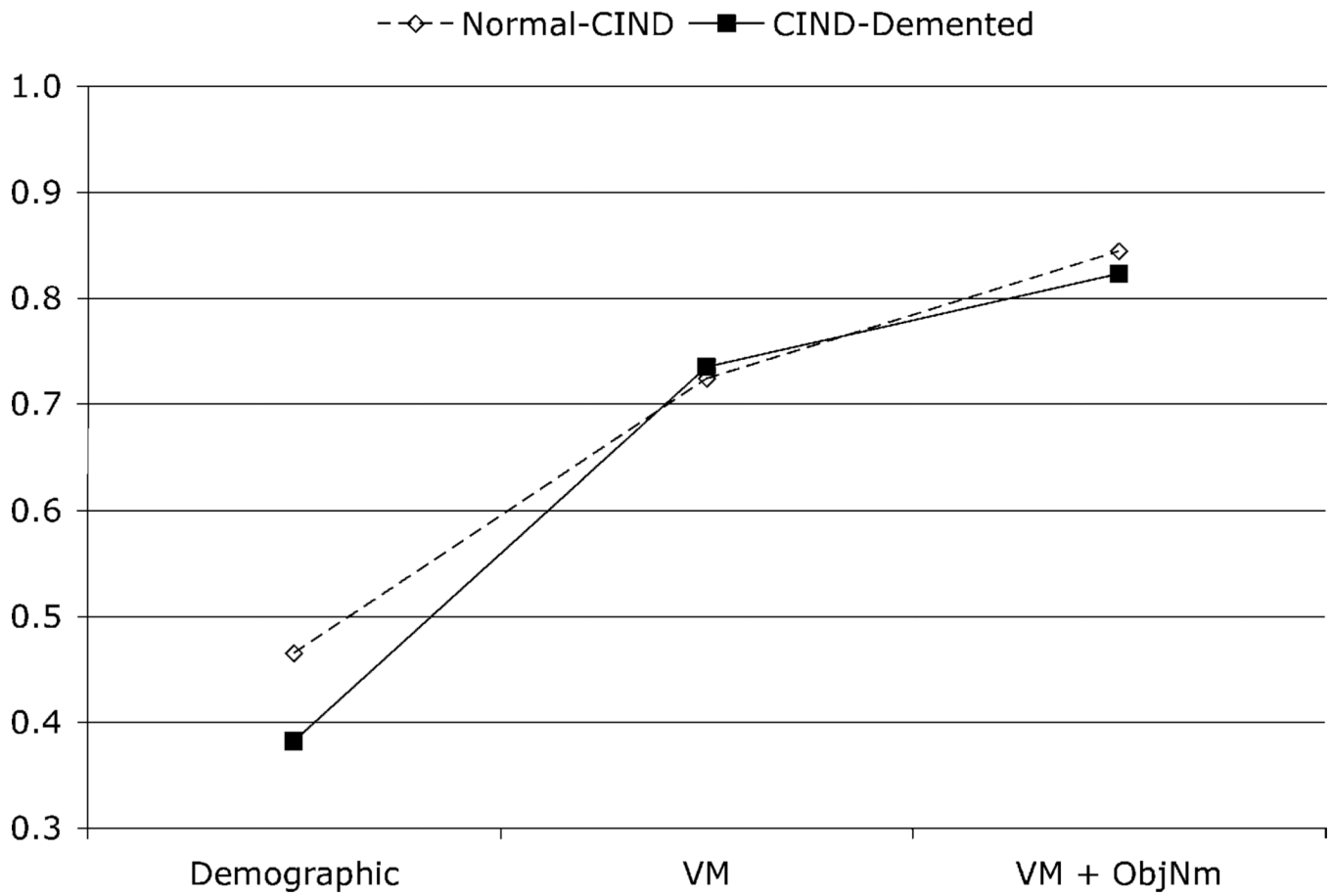


**Fig. 1.** Means and standard errors for Hispanics and Whites by normal *versus* CIND *versus* demented. Means are adjusted for effects of education, gender, age, and language usage. Ability scores are in standard deviation units based upon the distribution of the SENAS development sample.



**Fig. 2.**

Areas under the receiver operator characteristic curve associated with different logistic regression models to discriminate normal from CIND (open diamonds and dashed lines) and CIND from demented (filled squares and solid lines). Demographic = age, education, gender, language usage, VM = Verbal Memory (Word List Learning I), ObjNm = Object Naming.



**Fig. 3.** Diagnostic sensitivity (for specificity of .80) associated with different logistic regression models to discriminate Normal from CIND (open diamonds and dashed lines) and CIND from Demented (filled squares and solid lines). Demographic = age, education, gender, language usage, VM = Verbal Memory (Word List Learning I), ObjNm = Object Naming.

**Table 1**

Scales of neuropsychological test battery and abilities measured

<b>Ability domain</b>	<b>Verbal measure</b>	<b>Non-verbal measure</b>
Conceptual thinking	Verbal Conceptual Thinking	Non-Verbal Conceptual Thinking
Semantic memory	Object Naming	Picture Association
Attention span	Verbal Attention Span	Visual Attention Span
Episodic memory	Word List Learning–I Word List Learning–II	Spatial Configuration Learning
Non-verbal/spatial abilities		Pattern Recognition Spatial Localization
Verbal abilities	Verbal Comprehension Verbal Expression	
Executive function	Category Fluency (animals, supermarket test) Phonemic Fluency (/f/, /l/)  Working Memory (digit span backward, list sorting) Executive Composite	Working Memory (visual span backward) Executive Composite



**Table 2**

Demographic characteristics of sample by ethnicity and Clinical Syndrome

Ethnic group	Clinical diagnosis	N (% of Hispanic or White)	Gender N (%)	Education (years) M (SD)	Age (years) M (SD)	MMSE		BRDRS	
						Female	M (SD)	M (SD)	M (SD)
Hispanic	Normal	32 (47.1)	23 (71.9)	7.1 (5.2)	70.7 (6.9)	27.2 (2.9)	0.9 (1.7)		
	CIND	19 (27.9)	8 (42.1)	6.1 (5.8)	72.2 (6.8)	21.7 (5.9)	1.2 (1.0)		
	Demented	17 (25.0)	13 (76.5)	5.6 (4.4)	75.4 (5.8)	16.2 (6.0)	4.6 (2.7)		
White	Normal	30 (34.9)	20 (66.7)	14.2 (2.6)	74.7 (7.5)	29.2 (1.0)	0.2 (0.5)		
	CIND	39 (45.3)	21 (53.8)	14.5 (3.3)	76.9 (8.9)	27.1 (1.9)	1.4 (1.2)		
	Demented	17 (19.8)	7 (41.2)	14.2 (3.0)	80.1 (7.3)	22.7 (3.9)	3.2 (2.3)		
Total		154	92 (59.7)	10.9 (5.7)	74.8 (7.9)	25.2 (5.3)	1.6 (2.0)		

Note. MMSE = Mini-Mental State Examination (Folstein et al., 1975). BRDRS 5 Blessed Roth Dementia Rating Scale (Blessed et al., 1968; Blessed et al., 1968). See text for significant differences. CIND 5 cognitively impaired, not demented.

**Table 3**

Effect sizes for covariates (education, gender, age, language usage), ethnicity, Clinical Syndrome, and the ethnicity by Clinical Syndrome interaction

Scale	Covariates	Ethnicity	Clinical Syndrome	Ethnicity × Clinical Syndrome
Picture Association	.52	.001	.120*	.001
Object Naming	.55	.003	.146*	.003
Pattern Recognition	.45	.003	.027	.002
Verbal Attention Span	.44	.016	.013	.002
Verbal Conceptual Thinking	.52	.010	.144*	.013
Word List Learning–II	.15	.007	.355*	.020
Word List Learning–I	.19	.005	.373*	.024
Executive Composite	.40	.033*	.149*	.006
Category Fluency	.22	.010	.209*	.003
Phonemic Fluency	.30	.045*	.059*	.015
Working Memory	.43	.021	.131*	.001
Non-Verbal Conceptual Thinking	.34	.011	.063	.005
Spatial Localization	.29	.005	.137*	.013
Verbal Comprehension	.48	.033*	.086*	.015
Verbal Expression	.63	.001	.010	.008
Visual Attention Span	.31	.007	.109	.010
Spatial Configuration Learning	.29	.000	.227*	.005

*Note.* The covariate effect size is the  $R^2$  accounted for by the covariates entered jointly. Joint covariate effects were significant ( $p < .0001$ ) for all scales. Effect sizes for Ethnicity × Clinical Syndrome were calculated as the  $R^2$  for a model including covariates, ethnicity, clinical syndrome, and Ethnicity × Clinical Syndrome minus the  $R^2$  for a model without this interaction term. Effect sizes for ethnicity are the difference between the  $R^2$  value associated with a model with covariates, ethnicity, and clinical syndrome, and that model without ethnicity. Effect sizes for Clinical Syndrome represent incremental  $R^2$  beyond that accounted for by Covariates and Ethnicity.

\* statistically significant independent effect based upon the full model including interactions effects of ethnicity with Clinical Syndrome. Bonferroni adjusted  $p$  values were used ( $p = .05017 = .0029$ ).

**Table 4**

Raw and covariate-adjusted means by ethnicity and diagnostic group for six SENAS scales that significantly discriminated normal from CIND and CIND from demented

SENAS Measure	Diagnostic Syndrome	Raw score means		Covariate-adjusted means	
		Hispanic	White	Hispanic	White
Object Naming	Normal	-0.51	0.84	0.15	0.30
	CIND	-1.23	0.46	-0.48	-0.11
Verbal Conceptual Thinking	Demented	-1.92	-0.48	-1.03	-1.01
	Normal	-0.39	0.62	0.04	0.22
Word List Learning I	CIND	-1.16	0.34	-0.61	-0.08
	Demented	-2.00	-0.20	-1.41	-0.58
Word List Learning II	Normal	-0.26	0.57	-0.23	0.38
	CIND	-0.93	-0.71	-0.74	-0.84
Executive Composite	Demented	-1.76	-1.43	-1.63	-1.49
	Normal	-0.32	0.46	-0.32	0.35
Category Fluency	CIND	-0.96	-0.81	-0.85	-0.89
	Demented	-1.89	-1.50	-1.80	-1.52
Verbal Fluency	Normal	-0.15	0.88	-0.12	0.70
	CIND	-0.67	0.27	-0.50	0.15
Category Fluency	Demented	-1.53	-0.21	-1.33	-0.23
	Normal	0.12	0.76	0.15	0.60
Category Fluency	CIND	-0.42	0.09	-0.27	-0.02
	Demented	-1.17	-0.52	-1.03	-0.53

**Table 5**  
Regression analysis results for six SENAS scales that significantly discriminated normal from CIND and CIND from demented

SENAS measure	Effect	Coefficient	Standard error	p	Standardized Beta
Object Naming	Education	.046	.019	.01	.22
	Gender–Male	.058	.066	.37	.05
	Age	–.026	.008	.003	–.17
	Language	.276	.043	.0001	.57
Verbal Conceptual Thinking	Education	.073	.018	.0001	.38
	Gender–Male	–.036	.062	.56	–.03
	Age	–.015	.008	.07	–.11
	Language	.178	.041	.0001	.40
Word List Learning I	Education	.048	.021	.03	.27
	Gender–Male	–.237	.076	.002	–.23
	Age	–.03	.01	.002	–.23
	Language	.044	.05	.38	.11
Word List Learning II	Education	.057	.023	.01	.31
	Gender–Male	–.175	.081	.03	–.16
	Age	–.03	.01	.004	–.23
	Language	.014	.054	.78	.03
Executive Composite	Education	.085	.019	.0001	.47
	Gender–Male	–.12	.067	.07	–.11
	Age	–.031	.008	.0003	–.24
	Language	.084	.044	.06	.20
Category Fluency	Education	.043	.019	.03	.26
	Gender–Male	–.148	.068	.03	–.16
	Age	–.03	.009	.008	–.25
	Language	.072	.045	.11	.19

*Note.* The coefficient for female gender is –1.0 times the coefficient for males. Standardized beta squared is approximately the percent of variance independently accounted for by the independent variable.