Published in final edited form as:

NeuroRehabilitation. 2009 ; 24(1): 75–85. doi:10.3233/NRE-2009-0456.

ITEM ANALYSIS OF THREE SPANISH NAMING TESTS: A CROSS-CULTURAL INVESTIGATION

Carlos Marquez de la Plata, Ph.D.^{1,2}, Juan Carlos Arango-Lasprilla, Ph.D.³, Montse Alegret, Ph.D.⁴, Alexander Moreno, M.S.⁵, Luis Tárraga, M.S.⁴, Mar Lara, M.S.⁴, Margaret Hewlitt, B.S.², Linda Hynan, Ph.D.², and C. Munro Cullum, Ph.D.²

¹ University of Texas at Dallas, Dallas, Texas

² University of Texas Southwestern Medical Center, Dallas, Texas

³ Virginia Commonwealth University, Richmond, Virginia

⁴ Fundació ACE. Institut Català de Neurociències Aplicades, Barcelona, Spain

⁵ Central Police Hospital, Bogotá, Colombia

Abstract

Neuropsychological evaluations conducted in the United States and abroad commonly include the use of tests translated from English to Spanish. The use of translated naming tests for evaluating predominately Spanish-speakers has recently been challenged on the grounds that translating test items may compromise a test's construct validity. The Texas Spanish Naming Test (TNT) has been developed in Spanish specifically for use with Spanish-speakers; however, it is unlikely patients from diverse Spanish-speaking geographical regions will perform uniformly on a naming test. The present study evaluated and compared the internal consistency and patterns of item-difficulty and -discrimination for the TNT and two commonly used translated naming tests in three countries (i.e., United States, Colombia, Spain). Two hundred fifty two subjects (126 demented, 116 nondemented) across three countries were administered the TNT, Modified Boston Naming Test-Spanish, and the naming subtest from the CERAD. The TNT demonstrated superior internal consistency to its counterparts, a superior item difficulty pattern than the CERAD naming test, and a superior item discrimination pattern than the MBNT-S across countries. Overall, all three Spanish naming tests differentiated nondemented and moderately demented individuals, but the results suggest the items of the TNT are most appropriate to use with Spanish-speakers. Preliminary normative data for the three tests examined in each country are provided.

INTRODUCTION

Despite the dramatic growth in the population of primarily Spanish-speaking individuals in the United States, there are very few neuropsychological instruments developed in Spanish, as most instruments are literal translations of English tests. Not surprisingly, the construct validity of such tests have been called into question,[1–2] as items of a test may not measure the same construct due to differences in cultural and/or linguistic salience of the items used.[1–3] For example, some English words are comparatively used more frequently in English than Spanish and vice versa, and certain words are acquired at an earlier age in one language, and later in another.[4–6] These linguistic differences render certain words more or less salient than others, and may bias language oriented neuropsychological instruments if translated from one language to another. Furthermore, cultural and related differences between English- and Spanish-speaking individuals may also affect performance on neuropsychological instruments. [7–9]

Language-laden measures, such as tests used to assess confrontation naming ability, may be especially susceptible to decreases in construct validity in cross-cultural applications. While careful translation-backtranslation procedures are utilized to preserve a test's content validity, [10–11] they may not adequately preserve its construct validity.[1] For example, in cross-cultural test development, it is common for items of a test of language ability to undergo a process that translates a word/concept from one language to another while attempting to conserve the meaning of the word/concept;[11] however, it is possible for the resulting test to measure something other than the intended language ability due to differences in the salience of the words/concepts used between the two languages.[3,12–13] For example, Lowenstein and colleagues (1995) evaluated English- and Spanish-speaking individuals with Alzheimer's disease using various English- and Spanish-translated neuropsychological measures and found fewer Spanish-translated neuropsychological tests were related to functional ability than English tests, suggesting that translated neuropsychological tests may be less able to measure the construct they intend to measure (i.e. functional ability).

The Texas Spanish Naming Test (TNT), a new confrontation naming test for use with Spanish-speakers as an alternative to English naming tests that have been translated into Spanish, was developed to address the need of neuropsychological instruments written in Spanish for Spanish-speakers. The development of this test utilized theory from the experimental psycholinguistic literature and linguistic variables including age of word acquisition, and word frequency and familiarity to select culturally salient words for Spanish-speakers to reduce error and linguistic/cultural bias that may be introduced during translation of test items. Furthermore, TNT items were ordered according to item difficulty in a sample of nondemented Spanish-speaking individuals. While preliminary studies of the TNT have been promising,[2] additional information regarding the psychometric properties of the instrument in different Spanish-speaking populations is warranted.

Lack of normative data from which clinicians can base their interpretation of individual patient performance is another challenge facing neuropsychologists evaluating Spanish-speakers. Clinicians commonly use normative data from English-speaking Caucasians for interpreting performance on translated instruments. Given the linguistic and/or cultural biases that may occur during translation of language tests and possible educational differences between English-speaking Caucasians and predominantly Spanish-speakers, this practice may underestimate abilities on certain tasks or cause clinicians to interpret scores too leniently by over-adjusting for demographic/cultural/linguistic differences.

Translated naming tests, such as the Modified Boston Naming Test – Spanish (MBNT-S), [14] are commonly used to assess confrontation naming ability in Spanish-speakers, but have recently been shown to be less sensitive to dementia-related naming impairment than a naming test developed for Spanish-speakers (i.e., TNT) in the United States.[2] However, the item level psychometrics for any of the Spanish naming are largely unknown. Furthermore, it is unlikely the normative data available for the TNT (or other naming tests for that matter) can generalize to Spanish-speakers beyond the United States. This study aims to compare the internal consistency, and patterns of item difficulty and discriminability of a novel Spanish naming test developed in Spanish and two commonly used translated Spanish naming tests in three countries. A second aim is to provide cross-cultural normative data for these tests to aid their clinical interpretation when used with Spanish-speakers of different origins.

METHOD

Subjects

A sample of older participants with and without dementia was recruited from three locations including: 1) Dallas, Texas, USA, 2) Bogota, Colombia, and 3) Barcelona, Spain. All

participants reported Spanish as either their primary or only language and were at least 55 years of age. Subjects with dementia were identified as demonstrating evidence of cognitive decline during a medical evaluation, and were diagnosed with dementia using DSM-IV criteria by their physician. Subjects provided informed consent to participate in testing as part of a research study, and institutional review policies at each respective institution were adhered to.

Dallas, Texas, USA—Spanish-speaking participants from the United States were obtained from two Primary Care Clinics for older adult individuals in Dallas, Texas. Thirty Hispanic individuals with dementia and 56 nondemented controls were recruited from this location. Participants hailed from various Latin regions, including Mexico (64.7%), Central America (11%), South America (3%), and the Caribbean islands (3%). Additionally, 16% of the sample was born in the United States, and over 50% of the total group had lived in the United States over one-fourth of their lives.

Bogota, Colombia—Subjects from Colombia were recruited from the Central Police Hospital Memory Clinic in Bogota. Fifty six older participants were recruited from Colombia (36 with dementia, 20 without dementia).

Barcelona, Spain—Participants from Spain were recruited from Fundació ACE. Institut Català de Neurociències Aplicades in Barcelona. One hundred eight older Spanish-speaking participants (68 demented and 40 nondemented) underwent the research protocol.

Measure

Texas Spanish Naming Test (TNT)—The TNT is a Spanish naming test developed using depicted objects that are relevant and familiar to Spanish-speakers. The TNT was created from the 260 words studied in English and Spanish by Snodgrass and Vanderwart (1980)[6] and Cuetos, Ellis, and Alvarez (1999),[5] respectively. The latter investigated the linguistic qualities of the Snodgrass and Vanderwart pictures among a sample of Spanish-speaking college students from the University of Oviedo in Spain, and these data were used in the development of the TNT. The TNT was designed to be administered and scored in the same manner as the commonly used Boston Naming Test.[15] The 30 most salient and psychometrically reliable items were selected using standard item selection techniques from responses of demented and non-demented adult Spanish-speakers. The order of presentation for the final set of 30 items was based upon the level of item difficulty (percent correct) in the nondemented group (i.e., easier items presented before more difficult items). Additional detail regarding the development of the TNT can be found in Marquez de la Plata et al., (2008).[2] The TNT was administered in standard fashion to participants in all three countries.

Modified Boston Naming Test – Spanish (MBNT-S)—The MBNT-S is a commonly used Spanish naming test which consists of 30-items adapted from the original 60-item BNT. [14] It was developed by having expert judges select 30 items among the items on the original BNT on the basis of appropriateness for use with Spanish-speakers. They reordered the selected items according to difficulty, as they suspected different levels of difficulty for the same word across languages. The MBNT-S was normed using 300 volunteers from greater Los Angeles. The normative sample consisted primarily of monolingual Spanish-speakers (70%), but included individuals who considered themselves bilingual as well. The age of participants in the normative sample ranged from 16 to 75 years ($M = 38.4, \pm 13.5$), and their education ranged from 1 to 20 years ($M = 10.7, \pm 5.1$). Administration and scoring for the MBNT-S is similar to that of the original BNT, with one point for each correct response for a total of 30 possible points. The MBNT-S was administered to participants from all three countries using standard instructions.

Consortium to Establish A Registry for Alzheimer's Disease (CERAD)—The naming subtest of the CERAD[16] is an abridged and modified version of the BNT. Fifteen items were selected to comprise this test based on word frequency in the English language. The fifteen naming items of the CERAD have been translated to help identify word finding difficulty in Spanish-speaking patients with suspected dementia. The Spanish translation of the CERAD naming subtest is commonly used with Spanish-speakers without normative data on Spanish-speakers to base interpretive conclusions. The CERAD was administered to participants from Colombia and Spain. Participants from Dallas, Texas were not administered the CERAD.

Procedure

Clinical Examination—Each subject underwent a problem-focused clinical evaluation at their respective institutions which included an interview with the patient, a history and physical examination, a review of their medical records, and a brief screening for cognitive problems. Patients who endorsed memory complaints during their initial interview underwent a clinical mental status examination (MMSE),[17] and a more thorough dementia work-up.

Statistical Analyses

Cronbach's alpha was used to determine the internal consistency of the TNT for each sample. Mean differences in age, education, and MMSE between demented and nondemented groups were assessed using independent samples t-tests. Differences between diagnostic groups for gender (a dichotomous variable) was evaluated using a chi-squared test. A Multivariate Analysis of Covariance (MANCOVA) was used to determine naming test differences while controlling covariates.

Item difficulty (i.e., percent of sample answering each item correctly) for each country was determined for each item among demented participants, as item difficulty for nondemented participants would be less informative because very few of the early items are incorrect among nondemented participants. The greater the percentage of correct responses for an item, the easier the item. A Pearson correlation was utilized to determine the association between order of item administration and item difficulty. The coefficient of determination (r^2) was used to determine the amount of variance in naming performance accounted for by the order of the items in the three naming tests.

Item discrimination (i.e., difference between the percent of participants without dementia who answered an item correctly and the percent of participants with dementia who answered an item correctly) was determined for each item of each naming test. The greater the percentage difference, the better the item is able to discriminate demented from nondemented individuals. Item discrimination (D) was correlated with order of item presentation using a Spearman's rank correlation coefficient. Furthermore, items of each test were categorized as unacceptable ($D < 0.01$), poor ($D = 0 - 0.29$), good ($D = 0.30 - 0.39$), or excellent ($D > 0.40$) based on their item discrimination index.[18–19] Statistical Package for the Social Sciences for Windows version 11 (SPSS Inc., Chicago, IL) was used for all statistical analyses, with alpha set at the conventional 0.05 level of significance.

RESULTS

Sample Characteristics

As reported previously,[2] the demented group from the United States was older, had significantly lower MMSE scores, and was less educated than their nondemented U.S. counterparts. There were no significant difference between these groups with respect to gender

representation, country of origin (66.7% and 63.6% Mexico), and generation in the United States (86.7% and 81.8% first generation).

Colombian participants with dementia were less educated than participants without dementia and, as expected, showed poorer scores on the MMSE. There were no differences between groups with respect to age or gender.

Participants with dementia from Spain were older and showed lower MMSE scores than their nondemented counterparts. There were no differences between groups with respect to years of education or gender. Table 1 provides the subject characteristics for the two diagnostic groups.

Psychometric Properties of the Naming Tests

Inter-item Reliability—The internal consistency of the 30 items that comprise the TNT was determined by obtaining Cronbach's alpha with samples of demented and nondemented individuals from each respective site. This analysis resulted in a high alpha coefficients for all three samples (i.e., USA = 0.923; Colombia = 0.933; Spain = 0.927). Cronbach's alpha for the MBNT-S were not as high (i.e., USA = 0.854; Colombia = 0.898; Spain = 0.898). Likewise alpha for the CERAD was lower than that of the TNT (i.e., Colombia = 0.860; Spain = 0.776).

Item Difficulty—The TNT had a mean level of item difficulty of 0.51 (SD=0.23) in the United States, 0.56 (SD=0.25) in Colombia, and 0.75 (SD=0.16) in Spain. The MBNT-S showed a mean item difficulty of 0.39 (SD=0.34) in the United States, 0.40 (SD=0.33) in Colombia, and 0.60 (SD=0.31) in Spain. The CERAD showed a mean item difficulty of 0.53 (SD=0.32) in Colombia, and 0.78 (SD=0.20) in Spain. Note that the TNT had the smallest degree of variability around the mean across countries.

Spearman's rank correlation coefficients show order of item presentation is significantly associated with item difficulty for all three tests in each country (Figures 1–3). Note that item difficulty increases as the TNT and MBNT-S administration progresses; however the association is much weaker for the CERAD. Figures 1–3 demonstrates the relationship between order of item presentation and item difficulty for the naming tests in all three countries.

Item Discrimination

United States Sample: Item discrimination classification for the naming tests among participants from the United States shows that 10% of the items of the TNT were poor, 43.3% were mediocre, 20% were good, and 26.7% were excellent. Item classification for the MBNT-S showed that 56.7% of the items were poor, 16.7% were mediocre, 23.3% were good, and 3.3% were excellent. Furthermore, item discrimination indices for the items of the TNT correlated significantly with order of item presentation ($r = 0.621$, $p < 0.001$), but not with MBNT-S ($r = -0.007$, $p = 0.969$; see Figure 4).

Colombia Sample: Item discrimination classification for the naming tests among participants from Colombia showed that 26.7% of the items on the TNT were poor, 10% were mediocre, 36.7% were good, and 26.7% were excellent. Item classification for the MBNT-S showed that 40% of the items were poor, 13.3% were mediocre, 13.3% were good, and 33.3% were excellent. Item discrimination classification for the CERAD showed that 33.3% of the items fell in the poor range, 20% of the items were mediocre, 13.3% were good, and 33.3% were excellent. Additionally, item discrimination indices for the TNT, MBNT-S, and CERAD correlated significantly with order of item presentation ($r = 0.758$, $p < 0.001$; $r = 0.376$, $p = 0.041$; $r = 0.732$, $p = 0.002$, respectively). Note the difference in magnitude of correlation between the MBNT-S and its counterparts (see Figure 5).

Spain Sample: Item discrimination classification for the naming tests among Spaniards showed 50% of TNT items fell in the poor range, 20% were mediocre, 30% were good, and none of them were excellent. For the BNT, 53.3 % of the items were poor, 20% were mediocre, 16.7% were good, and 10% were excellent. Of the 15 items of the CERAD, item discrimination classification showed 33.3% of the items fell in the poor range, 13.3% were mediocre, 3.3% were good, and none of them were excellent. Furthermore, item discrimination indices for the TNT, MBNT-S, and CERAD correlated significantly with order of item presentation ($r = 0.842$, $p < 0.001$; $r = 0.507$, $p = 0.004$; $r = 0.757$, $p = 0.001$, respectively; see Figure 6). Note the difference in magnitude of correlation between the MBNT-S and its counterparts.

Discriminant Validity

Multivariate analysis of covariance (MANCOVA) was conducted to determine naming differences between nondemented and demented individuals from each country, using age and education as covariates. For the sample from the United States, both the TNT and the MBNT-S were significantly lower among patients than controls even after controlling for age and education [$F(1, 82) = 28.87$, $p < 0.001$; $F(1, 82) = 25.77$, $p < 0.001$, respectively]. Among Colombians, scores from all three naming tests (i.e., TNT, MBNT-S, and CERAD) were significantly lower among patients than controls after controlling for age and education [$F(1, 52) = 28.98$, $p < 0.001$; $F(1, 52) = 44.86$, $p < 0.001$; $F(1, 52) = 26.24$, $p < 0.001$, respectively]. Likewise, among participants from Spain, all three naming tests showed significant differences between controls and patients after controlling for age and education [$F(1, 106) = 18.26$, $p < 0.001$; $F(1, 106) = 26.26$, $p < 0.001$; $F(1, 106) = 11.01$, $p = 0.001$, respectively]. Covariate adjusted mean scores for the three Spanish naming tests are presented in Table 2.

DISCUSSION

Three Spanish naming tests were studied using demented and nondemented individuals from three countries (United States, Colombia, and Spain) to determine differences in psychometric properties of a naming test developed for Spanish-speakers and two commonly used Spanish naming tests that were translated from their original English language.

To our knowledge, this investigation is the first to compare internal consistencies of three Spanish naming tests. Internal consistency was generally rated as good for all three naming tests using conventional standards[20] (except for the CERAD in Spain); however, it is noteworthy that Cronbach's alpha was greater for the TNT than its counterparts in the countries studied. These findings suggest the items of the TNT reliably measure the same construct to a greater degree than its counterparts. The degree of item interrelatedness among these tests is not surprising. The items of the TNT were expected to have greater inter-item reliability, as they were selected according to their published normative data with respect to naming accuracy among nondemented Spanish-speaking adults.[5] This is a significant departure from the manner in which the MBNT-S and CERAD were adapted for use with Spanish-speakers. As discussed previously, only unambiguous words were included in the development of the TNT by selecting words from Cuetos et al. (1999)[5] in which at least 85% of participants in this study correctly used the target name. Given the stringent empirical criteria by which items were selected for inclusion in the test, it is not surprising that the TNT has such high internal consistency.

Cronbach's alpha for the TNT was also higher than those found for short forms of the BNT in primarily Caucasian samples. For example, Fastenau et al. (1998)[21] and Graves et al. (2004) [22] found 15-item short forms of the BNT demonstrated an alpha of between 0.37 to 0.84 in healthy Caucasians. Similarly, Cheung et al. (2004)[23] reported an internal consistency of 0.83 in a Cantonese version of the BNT.

While alpha for the TNT was greater than the above mentioned short forms, it was most similar to a 30-item short form of the BNT (alpha = 0.90). This short form of the BNT was derived using item response theory to identify items that maintain high reliability.[22] Though there are many sample differences between Graves et al. (2004)[22] and the present study, their respective naming tests may have comparable internal consistency because of the theory based manner in which their test items were selected, whereas the items in the MBNT-S, CERAD, and other 30- and 15-item versions of the BNT were selected by less rigorous means. These results support the notion that a test's construct validity is influenced by the manner in which the test items are selected; therefore, literal translations of test items may detract from their ability to assess a unitary construct.

Results showing a positive correlation between the TNT and its counterparts across the three countries suggest the TNT has convergent validity in these populations. As expected, the TNT had a relatively greater correlation with the MBNT-S across countries than with the CERAD, perhaps due to similarities in test length and scoring. There is greater similarity with regards to scoring between the TNT and the MBNT-S than with the CERAD, as a correct response for the two former tests consists of a response matching the target word spontaneously or with a semantic cue. However, the CERAD awards credit for a correct response after phonemic cues. This may influence test results by artificially inflating naming scores, as the task is simplified by giving credit to a correct response after a phonemic cue. This scoring scheme renders the CERAD slightly easier than its counterparts, and dilutes its ability to assess naming ability by confounding semantic and phonemic access to words.

With respect to item difficulty, most test items should have an item difficulty index of approximately 0.5 to best distinguish those who know and do not know correct responses. [24] The TNT showed an adequate level of item difficulty in the United States and Colombia; however, its item difficulty index in Spain suggests the test may be too easy for highly educated individuals. Likewise, the item difficulty index appeared to be highest for all three tests in Spain. This finding may be the result of the higher education level among the Spaniards in this study, as education is known to influence performance on naming tests.[25–26]

In light of previous investigations calling into question the graded nature of the Boston Naming Test when used among Spanish-speakers, we examined the relationship between item difficulty and the order in which the items from each naming test are presented.[27–28] The association between item difficulty and item number was best for the TNT in the United States, as items increase in difficulty in a much more linear fashion as test administration progresses compared to the MBNT-S. Item difficulty increases in a similar linear fashion for the TNT and MBNT-S in both Colombia and Spain; however, the CERAD item difficulty does not share as linear a relationship with item number as its counterparts in these countries. These results suggest the CERAD is too easy to accurately assess naming ability among Spanish-speakers with dementia, as even the most difficult CERAD items can be correctly identified by 72% and 86% of individuals with dementia in Colombia and Spain, respectively. The relative lack of increasing item difficulty seems to render the CERAD too easy, and may negatively impact its ability to assess confrontation naming among mildly impaired individuals. This result is consistent with the report by Franzen et al. (1995)[29] which found the average item difficulty index for the CERAD naming test was unfavorable among a large sample of older adults with various acquired brain injuries, and consistent with Fillenbaum, Huber, and Taussig, (1997) [30] who found item difficulty did not correlate well with order of item presentation.

Neuropsychological tests are commonly used to determine whether a patient's performance fits a profile of a particular diagnosis or reflects a cognitively intact individual. Item discrimination is one way to examine a test's ability to discriminate groups at the test construction/item selection level. Using the previously stated criteria for evaluating item

discrimination indices,[18–19] the TNT showed the most desirable item discrimination index distribution among its counterparts in each country sampled. Consider that approximately 47%, 63%, and 30% of TNT items were rated as either good or excellent ($D \geq 0.30$) in the United States, Colombia, and Spain, respectively, while 26%, 47%, and 27% of BNT items had good or excellent ratings in the three countries, respectively. Additionally, approximately 47% and 7% of the CERAD's item discrimination indices were rated as good or excellent in Colombia and Spain, respectively. These results suggest that in general, the items that comprise the TNT adequately discriminated patients with and without dementia in the United States and Colombia; however, the TNT had as much difficulty with this task in Spain as its counterparts. One explanation for the difficulty these naming tests have in discriminating groups in Spain may be the relatively greater amount of education this sample had compared to participants from the United States and Colombia, as it is well known that education influences confrontation naming performance.[25–26] Nonetheless, the pattern of item discrimination indices across countries supports the notion that simply translating a naming test into Spanish does not result in a test that adequately assess differences between patients with and without cognitive impairments.

Additional evidence to support this notion is the relatively weak association between the order of item administration and the item discrimination indices for the MBNT-S in all three countries. While the item discrimination index for the items that comprise the TNT approximated the ideal relationship between order of item presentation and item discrimination (i.e., item discrimination generally increases linearly as the test progresses), the distribution of item discrimination indices for the MBNT-S showed many items near the beginning of the test had the highest discrimination values and discrimination values for items near the middle and end of the test were among the lowest. This pattern is contrary to what is expected for a test that assesses a spectrum of ability in a graded fashion, and may be a function of being adapted from the English original for use in a different language. Although the CERAD's item discrimination indices displayed a linear relationship with order of item presentation, the number of inadequate items with respect to item discrimination (and item difficulty) suggests its items are generally inadequate. One might deduce its inadequacy is also related to the manner in which it was adapted from English into Spanish.

Given naming deficits are common in dementia,[31–34] the total scores for individuals with dementia from the United States, Colombia, and Spain are not surprisingly lower than nondemented counterparts on all three naming tests even after controlling for age and education. This is commensurate with prior studies examining neurocognitive differences between Spanish-speaking individuals with and without dementia;[2,35] however, this study extends the utility of Spanish naming tests (especially the Texas Naming Test) for use with Latin Americans and Spaniards by providing normative data from these populations (see Table 2). Additionally, clinicians using a naming test with highly educated Spanish-speakers should consider using normative data from Spain (see Table 2), as education appears to influence naming performance both at the total score and item level.

Furthermore, while this study demonstrated these tests are sensitive to differences between nondemented and moderately demented individuals, future studies should determine which of these Spanish naming tests can best detect mild cases of dementia. Based on the item analyses in this investigation, such a study may find the CERAD is too easy and the MBNT-S is too difficult to effectively discriminate between age-related naming difficulty and mild cases of dementia. Additionally, the TNT may also be a useful instrument for rehabilitation professionals working with Spanish-speaking patients with a variety of other neuropsychiatric conditions (i.e., aphasia secondary to stroke, or traumatic brain injury); however, normative data from younger cross-cultural cohorts are warranted.

Conclusions

This investigation examined psychometric qualities of the items that comprise two commonly used translated Spanish naming tests and a novel Spanish naming test developed for use with Spanish-speakers. We used groups from three countries that included two samples with limited education and one highly educated sample. The TNT demonstrated superior internal consistency to its counterparts, a superior item difficulty pattern than the CERAD naming test, and a superior item discrimination pattern than the MBNT-S across countries. Overall, all three Spanish naming tests differentiated nondemented and moderately demented individuals, but the results suggest the items of the TNT are most appropriate to use with Spanish-speakers. Preliminary normative reference data for each of the three tests in each country are provided to help guide clinical interpretation for whichever test one uses with Spanish-speakers. Furthermore, clinicians/rehabilitation professionals should consider using normative data from Spain when evaluating highly educated Spanish-speakers.

References

1. Gutierrez, G. The empirical development of a neuropsychological screening instrument for Mexican-Americans. In: Ferraro, FR., editor. *Minority and cross-cultural aspects of neuropsychological assessment*. Studies on Neuropsychology, development, and cognition. Swets & Zeitlinger Publishers; Lisse, Netherlands: 2002. p. 205-224.
2. Marquez de la Plata CD, Vicioso B, Hynan L, Evans HM, Diaz-Arrastia R, Lacritz L, Cullum CM. Development of the Texas Spanish Naming Test: A test for Spanish-speakers. *The Clinical Neuropsychologist* 2008;22:288-304. [PubMed: 17853138]
3. Loewenstein DA, Rubert MP, Arguelles T, Duara R. Neuropsychological test performance and prediction of functional capacities among Spanish-speaking and English-speaking patients with dementia. *Archives of Clinical Neuropsychology* 1995;10:75-88. [PubMed: 14589730]
4. Alameda, JR.; Cuetos, F. *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo; 1995.
5. Cuetos F, Ellis A, Alvarez B. Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers* 1999;31:650-658.
6. Snodgrass JG, Vanderwart M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory* 1980;5:174-215. [PubMed: 7373248]
7. Lichtenberg PA, Ross T, Christensen B. Preliminary normative data on the Boston Naming Test for an older urban population. *The Clinical Neuropsychologist* 1994;8:109-111.
8. 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]
9. Whitfield KE, Fillenbaum GG, Pieper C, Albert MS, Berman LF, Blazer DG, et al. The effect of race and health-related factors on naming and memory: The MacArthur studies of successful aging. *Journal of Aging and Health* 2000;12:69-89. [PubMed: 10848126]
10. Brislin, RW. Translation and content analysis of oral and written materials. In: Triandis, HC.; Berry, JW., editors. *Handbook of cross-cultural psychology-methodology*. Allyn & Bacon Inc; Boston: 1980.
11. Geisinger KF. Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment* 1994;6:304-312.
12. Ellis NC, Hennelly RA. A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology* 1980;71:45-51.
13. Valencia RR, Rankin RJ. Evidence of content bias on the McCarthy Scales with Mexican American children: Implications for test translation and nonbiased assessment. *Journal of Educational Psychology* 1985;77:197-207.

14. Ponton MO, Satz P, Herrera L, Ortiz F, Urrutia CP, Young R, et al. Normative data stratified by age and education for the Neuropsychological Screening Battery for Hispanics (NeSBHIS): Initial report. *Journal of the International Neuropsychological Society* 1996;2:96–104. [PubMed: 9375194]
15. Kaplan, EF.; Goodglass, H.; Weintraub, S. The Boston Naming Test. Vol. 2. Lea & Febiger; Philadelphia: 1983.
16. Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 1989;39:1159–1165. [PubMed: 2771064]
17. 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]
18. Ebel, RL.; Frisbie, DA. Essentials of educational measurement. Prentice-Hall; Englewood Cliffs, New Jersey: 1986.
19. Crocker, L.; Algina, J. Introduction to classical and modern test theory. Holt, Rinehart & Winston; New York: 1986.
20. Robinson, JP.; Shaver, PR.; Wrightsman, LS. Criteria for scale selection and evaluation. In: Robinson, John P.; Shaver, Phillip R.; Wrightsman, Lawrence S., editors. Measures of Personality and social psychological attitudes. Academic; San Diego: 1991. p. 1-15.
21. Fastenau PS, Denburg NL, Mauer BA. Parallel short forms for the Boston Naming Test: Psychometric properties and norms for older adults. *Journal of Clinical and Experimental Neuropsychology* 1998;20:828–834. [PubMed: 10484693]
22. Graves RE, Bezeau SC, Fogarty J, Blair R. Boston Naming Test short forms: A comparison of previous forms with new item response theory based forms. *Journal of Clinical and Experimental Neuropsychology* 2004;26:891–902. [PubMed: 15742540]
23. Cheung RW, Cheung M, Chan AS. Confrontation naming in Chinese patients with left, right, or bilateral brain damage. *Journal of the Neuropsychological Society* 2004;10:46–53.
24. Jacobs, LC.; Chase, CL. Developing and using tests effectively. Jossey-Bass; San Francisco: 1992. p. 51-82.
25. Welch LW, Doineau D, Johnson S, King D. Educational and gender normative data for the Boston Naming Test in a group of older adults. *Brain and Language* 1996;53:260–266. [PubMed: 8726536]
26. Ross TP, Lichtenberg PA, Christensen BK. Normative data on the Boston Naming Test for elderly adults in a demographically diverse medical sample. *The Clinical Neuropsychologist* 1995;9:321–325.
27. Allegri RF, Mangone CA, Villavicencio AF, Rymberg S, Taragano FE, Baumann D. Spanish Boston Naming Test Norms. *The Clinical Neuropsychologist* 1997;11:416–420.
28. Kohnert K, Hernandez A, Bates E. Bilingual performance on the Boston Naming Test: Preliminary norms in Spanish and English. *Brain and Language* 1998;65:422–440. [PubMed: 9843612]
29. Franzen MD, Haut MW, Rankin E, Keefover R. Empirical comparison of alternate forms of the Boston Naming Test. *The Clinical Neuropsychologist* 1995;9:225–229.
30. Fillenbaum GG, Huber M, Taussig IM. Performance of elderly White and African American community residents on the abbreviated CERAD Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology* 1997;19:204–210. [PubMed: 9240480]
31. Bayles, KA.; Tomoeda, CK. Confrontation and generative naming abilities of dementia patients. In: Brookshire, RH., editor. Clinical Aphasiology Conference Proceedings. BRK Publishers; Minnesota: 1983. p. 304-321.
32. Frank EM, McDade HL, Scott WK. Naming in dementia secondary to Parkinson's, Huntington's, and Alzheimer's diseases. *Journal of Communication Disorders* 1996;29:183–197. [PubMed: 8799853]
33. Lichtenberg P, Vangel S, Kimbarow M, Ross T. The Boston Naming Test – Clinical utility. *The Clinical Gerontologist* 1996;16:69–72.
34. Testa JA, Ivnik RJ, Boeve B, Petersen RC, Pankratz VS, et al. 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]

35. Taussig IM, Henderson VW, Mack W. Spanish translation and validation of a neuropsychological battery: Performance of Spanish- and English-speaking Alzheimer's disease patients and normal comparison subjects. *Clinical Gerontology* 1992;2:95–108.

Figure 1.
Correlations between item difficulty and order of item presentation for United States sample.

Figure 2.
Correlations between item difficulty and order of item presentation for Colombia sample.

Figure 3.
Correlations between item difficulty and order of item presentation for Spain sample.

Figure 4.
Correlations between item discrimination and order of item presentation for United States sample.

Figure 5.
Correlations between item discrimination and order of item presentation for Colombia sample.

Figure 6.
Correlations between item discrimination and order of item presentation for Spain sample.

Table 1

Sample Demographics

	Dallas			Colombia			Spain		
	Nondemented	Demented		Nondemented	Demented		Nondemented	Demented	
	(n=56)	(n=30)		(n=20)	(n=36)		(n=40)	(n=70)	
Age	72.9 (6.3)	77.8 (7.5) *		69.4 (9.6)	74.0 (6.7)		71.9 (6.7)	77.9 (6.6) *	
Education (yrs)	4.8 (3.8)	1.6 (1.8) *		8.8 (4.1)	5.6 (4.7) *		15.4 (3.3)	14.3 (5.1)	
MMSE	22.5 (4.3)	12.2 (4.3) *		27.4 (1.6)	16.8 (5.2) *		29.0 (1.3)	21.5 (3.8) *	
Gender (% male)	35.7	43.3		40.0	44.4		45.0	45.7	

Note. Age, education, MMSE scores, and gender for each diagnostic group by country are provided. Values in parentheses are standard deviations.

* Denotes significant difference between diagnostic groups within each country ($p < 0.05$).

Table 2
Age and Education Corrected Spanish Naming Test Scores

	Dallas		Colombia		Spain	
	Nondemented	Demented	Nondemented	Demented	Nondemented	Demented
	(n=56)	(n=30)	(n=20)	(n=36)	(n=40)	(n=70)
TNT	23.9 (0.7)	16.8 (1.0)	26.4 (1.4)	16.7 (1.0)	27.9 (0.9)	22.9 (0.7)
MBNT-S	17.4 (0.5)	12.6 (0.7)	21.1 (1.0)	12.6 (0.7)	23.5 (0.8)	18.4 (0.6)
CERAD	-	-	12.5 (0.7)	7.9 (0.5)	13.4 (0.4)	11.8 (0.3)

Note. Age- and education- adjusted group mean scores for the Texas Naming Test (TNT), Modified Boston Naming Test-Spanish (MBNT-S), and the naming test from the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD). Values in parentheses are standard errors. Patients with dementia scored significantly lower than nondemented patients on all naming tests regardless of origin.