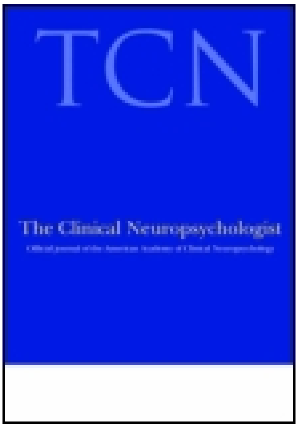

The Clinical Neuropsychologist
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ntcn20

Development of a Confrontation Naming Test For Spanish-speakers: The Cordoba Naming Test

Alberto Luis Fernández [a]

[a] Catholic University of Córdoba, Cortex Foundation , Universidad Nacional de Córdoba , Córdoba , Argentina

Published online: 01 Aug 2013.

To cite this article: Alberto Luis Fernández (2013) Development of a Confrontation Naming Test For Spanish-speakers: The Cordoba Naming Test, The Clinical Neuropsychologist, 27:7, 1179-1198, DOI: 10.1080/13854046.2013.822931

To link to this article: http://dx.doi.org/10.1080/13854046.2013.822931

The Clinical Neuropsychologist, 2013
Vol. 27, No. 7, 1179–1198, http://dx.doi.org/10.1080/13854046.2013.822931

Development of a Confrontation Naming Test For Spanish-speakers: The Cordoba Naming Test

Alberto Luis Fernández
Catholic University of Córdoba, Cortex Foundation, Universidad Nacional de Córdoba, Córdoba, Argentina

To date, a psychometrically sound standardized Spanish-language test of confrontation naming has not been developed for clinical use. Because of the shortcomings of adapting tests developed in other cultures, it was decided to develop a confrontation naming test suitable for Spanish-speakers. For the validity study the performance on the test of 26 control subjects between 70 and 87 years old and 23 subjects with a mild to moderate degree of dementia of the Alzheimer type was compared. Stability of the test was assessed with a test–retest design (n = 80). Norms were developed using a regression-based method. Four hundred and fifty-six Spanish-speaking subjects of both sexes were recruited for the normative sample. Subjects were between 14 and 94 years old, and three educational levels were represented. Mean differences between the control and dementia groups were significant, yielding a large effect size (η^2 = .25). The test–retest correlation coefficient was r = .90. Education, age, and gender significantly influenced test scores. The validity study confirmed that the test discriminates between individuals with and without anomia. The magnitude of the reliability coefficient of this test can be considered as "very high". Norms were developed considering the influence of three demographic variables: gender, age, and education.

Keywords: Anomia; Alzheimer's disease; Confrontation naming test; Cross-cultural neuropsychology; Neuropsychological assessment; Test in Spanish.

INTRODUCTION

In the course of spontaneous speech, a speaker needs to retrieve words fluently. Non-fluent speech can manifest in a variety of symptoms. One of these symptoms is *anomia* which has been defined as "…a difficulty in finding high information words, both in fluent discourse, and when called upon to identify an object or action by name" (Goodglass & Wingfield, 1997, p. 3). Confrontation naming tests are the typical method of assessing anomia. In these tests, the subject is shown a series of objects or drawings of objects (often of varying frequency and/or familiarity) and is instructed to name the object.

I would like to thank the numerous colleagues and students who participated in the data collection in different stages of these studies. Among them I would like to highlight the contributions of Ivana Bollati, María Lanza Castelli, Paula Ferrrochio, Paula Bianconi, Laura Manoiloff, Gustavo Godoy-Garcete, Pablo Marchetti, and Aiby Pérez Quinteros. I would also like to thank the Private Institute of Neurosciences for their patient data provided for the validity study and Richard Fulbright, PhD for his help with the English revision. This research was partially founded by a grant of the Catholic University of Córdoba.

Address correspondence to: Cortex Foundation, Chile 279, PB, CP 5000, Córdoba, Argentina. E-mail: neurorehab@onenet.com.ar

Accepted for publication: July 2, 2013. First publish online: August 1, 2013

The assessment of confrontation naming is very important for the early diagnosis of conditions such as Alzheimer's disease (AD). Research has demonstrated that anomia is one of the earliest symptoms of Alzheimer's disease (Huff, Corkin, & Growdon, 1986; Zec, 1993). Patients with AD have difficulties in retrieving the correct words when speaking, which results in circumlocution, i.e., uttering a large number of words in order to characterize an object whose specific name cannot be retrieved from semantic memory. Assessment of confrontation naming is very important in these cases, since it helps confirm the diagnosis of AD (Zec, 1993). Anomia is also a very frequent symptom in patients suffering from other conditions such as aphasia (Helm-Estabrooks & Albert, 2004), frontotemporal dementia (Weder, Aziz, Wilkins, & Tampi, 2007), post-epilepsy surgery (Ives-Deliperi & Butler, 2012), and even fragile X syndrome (Spinelli, De Oliveira Rocha, Giacheti, & Richieri-Costa, 1995).

Most standardized measures of naming ability use line drawings of the objects to be named. Several tests using this methodology have been designed, with the Boston Naming Test (BNT) being the most widely used. Two versions of this test have been released (Kaplan, Goodglass, & Weintraub, 1983, 2001). The BNT is comprised of 60 line drawings of objects, and even though it was developed in an English-speaking environment, it is widely employed in several non English-speaking countries. The BNT has been adapted to a variety of languages, including Malay (van Dort, Vong, Razak, Kamal, & Meng, 2007), Korean (Kim & Na, 1999), French-Canadian (Roberts & Doucet, 2011), French-Swiss (Thuillard-Colombo & Assal, 1992), Chinese (Cheung, Cheung, & Chan, 2004), Dutch (Marien, Mampaey, Vervaet, Saerens, & De Deyn, 1998), Swedish (Tallberg, 2005), Italian (Riva, Nichelli, & Devoti, 2000), and Greek (Patricacou, 2007).

One other naming test is the Visual Naming Test which is included in the Multilingual Aphasia Examination battery. This instrument has been adapted to Spanish (Rey & Benton, 1991). Nonetheless, this test consists of only 10 line drawings requiring 30 naming responses. Although some criterion validity data have been reported, norms and reliability data have not been published (Rey, Feldman, Rivas-Vazquez, Levin, & Benton, 1999). The Graded Naming Test (McKenna & Warrington, 1983) is another naming test that was developed in the United Kingdom and comprises 30 hierarchically organized items. There is no Spanish language adaptation of this test.

All of these tests have been developed in English-speaking contexts, and some of them have been adapted to Spanish, especially the BNT (Allegri et al., 1997; Pontón et al., 1992a; Quiñones-Ubeda, Peña-Casanova, Böhm, Gramunt-Fombuena, & Comas, 2004). However, researchers have invariably faced the obstacle that some of these items are inappropriate for Spanish-speaking populations. For example, among its figures, the BNT includes a pretzel, an igloo, and a beaver. Pretzel is a word that has no equivalent in Spanish, the igloo is frequently interpreted as a mud oven in some rural areas of Latin America, and the beaver dwells almost exclusively in North America. Consequently, many individuals cannot name it because it is unknown to them. Because of these problems, the construct under evaluation is confounded when the test is applied cross-culturally. In many cases, what is being assessed is not the naming ability (ability to retrieve words stored in semantic memory), but the extension of semantic memory; that is, whether the evaluee even *knows* the name of the object. Therefore, the measurement of naming ability is confounded by both language differences as well as adaptive test versions that are not equivalent to the original.

Consequently, these features dilute the validity and generalizability of the psychometric studies performed with the original test (Van de Vijver & Tanzer, 1997). Data obtained about the psychometric properties of the original tests are not applicable to these adaptations, since they are not parallel or interchangeable instruments.

Moreover, the difficulties in the cross-cultural administration of the BNT are apparent not only when utilized in non English-speaking countries, but also in English-speaking countries, such as New Zealand and Australia. Barker-Collo (2001) found that a sample of New Zealander university students had a significantly lower mean score than a sample from the United States of America. She suggested that such a difference was caused by the considerably lower percentage of correct answers on some less universal items such as "pretzel", "beaver", and "asparagus". Cruice, Worrall, and Hickson (2000) obtained similar results in Australia. They recommended the substitution of "beaver" and "pretzel" by "platypus" and "pizza" when the BNT was to be applied in Australian samples.

Recently Gollan, Weissberger, Runnqvist, Montoya, and Cera (2012) described the Multilingual Naming Test (MINT); a new naming test that was designed for bilingual speakers. This test consists of a set of 68 black-and-white line drawings selected and presented in order of estimated increasing difficulty. Although the MINT has been carefully designed for the assessment of subjects in a variety of languages (English, Spanish, Hebrew, and Mandarin Chinese), no psychometric data were reported (validity or reliability), and normative data were scarce. Thus far, the MINT seems a useful test for research rather than for clinical settings.

The one test that has been specifically developed to assess Hispanic population to date is the Texas Spanish Naming Test (TSNT; Marquez de la Plata et al., 2008). This test consists of 30 items selected according to the frequency of use in Spanish. Nevertheless, the normative sample of the TSNT is small (n = 55) and was limited to subjects over 64 years of age. Furthermore, the mean level of education of the subjects was very low (M = 5 years ± 6). Although the test seems to have acceptable evidence of validity, its norms limit its use in older Spanish-speaking adults with low levels of education.

In light of this history, there is an obvious need for a confrontation naming test more specific to the Spanish-speaking population. In this paper, it is described the development, validity and reliability studies, and normative data of the Córdoba Naming Test (CNT). The CNT was developed in the city of Córdoba, Argentina, and its norms are based on an Argentinean sample.

SUBJECTS AND METHODS

Materials development

The CNT was initially developed in 2002. In a first stage, a group of items was selected with the aim of including words with varying frequency of use in everyday living. An example of high frequency words were "pencil" and "book", while low frequency words were "chisel" and "Minotaur". Busca-Palabras (B-Pal), a software program that contains a large database of the frequency of use of Spanish words, establishes that *lápiz* (pencil) has a frequency of 6.96 times per million words while *formón* (chisel) appears 0.18 times per million words (Davis & Perea, 2005). In addition, we attempted to avoid using polisemic items, i.e., items with more than one name. The initial drawing set contained 45 items. Since there were no tools such as B-Pal at the time that this project was started, the items were placed in a hierarchical order of frequency according to the author's estimation of their frequency of use. The drawings were carefully designed with the goal of producing clear, non-ambiguous pictures that could elude naming failures resulting from defective recognition.

In the next stage, a pilot study was run. All of the 45 items of the initial version of the CNT were administered to a sample of 49 subjects in which both genders and a wide range of age and education were represented. The sample was comprised of people recruited from several sources: individuals attending teaching programs for older adults as well as acquaintances or relatives of the test administrators. Participants gave their consent to be tested. Subjects were grouped in the following age categories: 15–19 (n = 4), 20–29 (n = 6), 30–39 (n = 9), 40–49 (n = 7), 50–59 (n = 6), 60–69 (n = 6), and over 69 (n = 11). Three education categories were established: primary (1–7 years, n = 18), secondary (8–12 years, n = 16), and tertiary (more than 12 years, n = 15). Forty-nine percent of the subjects were female.

Administration of the CNT was performed by several psychology students and graduate psychologists. To insure standardized administration and scoring, all administrators underwent four training sessions with the test developers. After reviewing the test administration instructions, raters practiced administering the instrument, first with other raters and then with individuals who matched the sample used in the final studies. Data gathered during the training sessions were not included in the data analyses shown in this paper. Test administration was performed in different settings based on the availability (classrooms, laboratories, etc.). Test administration was always carried out under standardized conditions defined as a quiet, well-lit room containing only the rater and study participant.

In the third stage, data were analyzed in order to design the definitive version of the test. Thirty items were included in its final form. These items were hierarchically ordered from the easiest to the most difficult item according to the percentage of correct answers obtained in the pilot sample. The percentages of correct answers for each item are shown in Table 1.

Drawings appearing in boldface in Table 1 were excluded from the final version of the test. Criteria for the exclusion were the following:

(1) pictures that did not elicit a univocal answer, namely objects that were easily confused with other objects. For instance, the neuron was often interpreted as a tree;
(2) pictures that had more than one acceptable answer though not all of them were correct. For instance, the headphones ("auriculares") were called earphones ("audífonos") by some people. Although this answer is incorrect, it is not unusual in certain neurologically intact people in our society who typically use the term headphones for "earphones"; and
(3) finally, some items showing a high percentage of correct answers were excluded in order to maximize the possibility of obtaining a normal scoring distribution.

After the items were selected, an instructional manual was designed. The manual contained the following sections: materials, demographic data registry instructions, test administration instructions, criteria for test interruption, glossary, and instructions for

Table 1. Percentage of correct answers for each item in the pilot study

Item	Percentage
SILLA (CHAIR)	**100**
PIPA (PIPE)	100
LÁPIZ (PENCIL)	**98**
RASTRILLO (RAKE)	98
CANILLA (FAUCET)	**96**
DESTORNILLADOR (SCREWDRIVER)	94
LIBRO (BOOK)	92
CARRETILLA (BARROW)	**92**
TOBOGÁN (SLIDE)	**90**
ESPADA (SWORD)	88
JERINGA (SYRINGE)	88
CALCULADORA (POCKET CALCULATOR)	86
VIOLÍN (VIOLIN)	84
TRACTOR (TRACTOR)	84
COHETE (ROCKET)	**82**
CASSETTE (CASSETTE)	82
BRÚJULA (COMPASS)	78
SUBMARINO (SUBMARINE)	76
DELFIN (DOLPHIN)	73
TREBOL (CLOVER)	73
CRESTA (COMB)	71
BROCHA (PAINTBRUSH)	**63**
BARRILETE (KITE)	**63**
GUADAÑA (SICKLE)	63
CANGREJO (CRAB)	**61**
AURICULARES (HEADPHONES)	**59**
MICROSCOPIO (MICROSCOPE)	56
TELESCOPIO (TELESCOPE)	**53**
CARABELA (CARAVEL)	51
COBRA (COBRA)	49
ANTIPARRA (GOGGLES)	**47**
ALICATE (PLIERS)	**43**
PORTAAVIONES (AIRCRAFT CARRIER)	43
ESPUMADERA (SKIMMING LADLE)	**39**
BALLESTA (CROSSBOW)	39
ALFIL (BISHOP)	37
MINOTAURO (MINOTAUR)	37
SATURNO (SATURN)	31
DIRIGIBLE (ZEPPELIN)	29
ATOMO (ATOM)	29
FORMÓN (CHISEL)	27
NEURONA (NEURON)	**27**
TANGENTE (TANGENT)	12
ALJABA (QUIVER)	4
CLAVE DE FA (BASS CLEF)	2

Items in bold letter were excluded from the final version of the test.

scoring. Scoring on the CNT is conducted as follows: (1) every spontaneous correct response or a correct response given after a semantic cue is given one point. (2) No point is given for correct answers prompted by a phonological cue. (3) The final score is tabulated by adding all the correct spontaneous answers plus all the correct answers given following a semantic cue. (4) Test administration is discontinued after four consecutive non-spontaneous answers (even if they are correct or incorrect after the cues). This rule was made on the basis of the low percentage of subjects that gave correct answers after four consecutive failures. Forty-five percent of the pilot study gave no spontaneous correct answers after four consecutive failures, and an additional 44% gave up to three correct spontaneous answers. Only 12% gave more than three spontaneous correct answers after four consecutive failures. Thus, since that asking the subjects to retrieve names that they are unlikely to recall can be very frustrating, it was decided to discontinue test administration after four consecutive non-spontaneous answers. After these materials and procedures were established, validity, reliability, and normative studies were started.

Normative study

Four hundred and fifty-six cases were included in the sample. Inclusion of each case required that health backgrounds of the subject were explored through a set of questions. Subjects with any of the following diagnoses were excluded from the sample: stroke, loss of consciousness, traumatic head injury, central nervous system diseases, diabetes, chronic renal insufficiency, hepatic encephalopathy, non-treated thyroid disease, chronic headache, epilepsy, non-treated high blood pressure, severe cardiac failure, severe sleep disorders, coma, diagnosed psychiatric disease or illegal drug consumption. The recruiting procedure and test settings were identical to the pilot study. Mean age was 49.99 ± 20.77. Age range was 14–94 years old. The mean of years of education was 13.58 ± 4.55, and the education range was 2–31 years. Sixty-two percent of the sample were female. Regarding handedness, 88% was right-handed, 9% ambidextrous, and 3% left-handed. Subjects came from a variety of Argentinean cities and states (Córdoba, Resistencia, Rio Cuarto, Trelew, La Playosa, El Bolsón, Cruz del Eje, Carlos Paz, etceteras). Test administrators were psychology students and psychologists properly trained as described earlier. Tables 2–4 show the mean scores corresponding to CNT raw scores for each age, gender, and education group.

For the normative data analysis, the regression-based method was followed. Initially promoted by Zachary and Gorsuch (1985) this approach is being used more regularly by norm developers (Heaton, Avitable, Grant, & Matthews, 1999; Parmenter,

Table 2. Mean and standard deviation corresponding to Cordoba Naming Test raw scores for each age group

Age group	*N*	Mean	*SD*
14–19	11	19.27	4.96
20–29	126	22.18	4.19
30–39	19	22.58	4.27
40–49	20	22.65	3.12
50–59	95	23.00	3.97
60–69	90	20.88	4.43
70–79	72	17.94	5.02
80–89	22	15.23	5.66
90–99	1	11.00	

Table 3. Mean and standard deviation corresponding to Cordoba Naming Test raw scores for each gender group

Gender	*N*	Mean	*SD*
Male	174	23.61	4.56
Female	282	19.44	4.37

Table 4. Mean and standard deviation corresponding to Cordoba Naming Test raw scores for each education group

Education level	*N*	Mean	*SD*
Primary (1–7)	67	16.22	4.17
Secondary (8–12)	113	20.31	4.81
Tertiary (>12)	276	22.50	4.22

Testa, Schretlen, Weinstock-Guttman, & Benedict, 2010; Testa, Winicki, Pearlson, Gordon, & Schretlen, 2009; Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006a, 2006b). One of the most important advantages of this method is that, unlike the traditional approach, age and education grouping are not arbitrarily defined. In the traditional methodology age groups are usually defined in terms of decades (20–29, 30–39, etc.). Educational groups are also rather arbitrarily defined (1–7, 8–12, and so forth). It is valid to question how different is the performance of a 29-year-old subject from another who is 30 years old. Likewise, it is legitimate to question if there is any difference in the performance of a subject with 12 years of education as compared to another having 13 years of education. Moreover, the traditional grouping method requires larger samples in order to get an acceptable reliability (Testa, S. M. et al., 2009).

In accordance with the method described by Testa, S. M. et al. (2009), raw scores were transformed to scaled scores (mean = 10, *SD* = 3) based on the cumulative frequency of the normative sample. The resulting scores are exhibited in Table 5.

Next, using the scaled scores as the dependent variable and age, years of education, sex, handedness, and age squared as predictive variables, a multiple regression analysis was performed. Gender was dummy coded with male = 1 and female = 2. Handedness was dummy coded with right-handed = 1, left-handed = 2, and ambidextrous = 3. Age-squared was included in order to smooth the polynomic relationship of age (Allison, 1999). The inclusion of this squared variable in the regression analysis has also been employed by other authors in developing normative data (Parmenter et al., 2010; Van der Elst et al., 2006a, 2006b).

Validity

In order to assess the validity of the CNT, a construct validity study was performed using the contrasted group method. Two groups, control and AD, were compared. The AD group was used as a model of anomia, since this is a very frequent

Table 5. Equivalence table for the transformation of raw scores into scaled scores

Scaled score	Raw score
19	30
16	29
15	28
14	27
13	26
12	25
11	23–24
10	21–22
9	19–20
8	17–18
7	16
6	14–15
5	12–13
4	9–11
3	7–8
1	1–6

feature of these patients. A one-way ANOVA was performed using the raw score of the CNT as the dependent variable. The total sample was comprised of 49 subjects, divided in two groups: (a) “Alzheimer’s disease group” (n = 23), which included individuals diagnosed with AD according to the NINCS-ARDRA criteria (McKhann et al., 1984). The diagnosis of AD was based on (1) a comprehensive neuropsychological test battery assessing memory, attention, constructional praxis, verbal fluency (phonological and semantic), reading, writing, calculation, and executive functioning (the CNT was not considered to determine the diagnosis) and (2) a detailed review of subjects’ medical history, including current symptoms, medical records of the patient, and present life circumstances. In many cases, additional medical procedures such as neurological assessment, psychiatric assessment, electrophysiological studies (quantified EEG, evoked potential P300), imaging (MRI, CT scan or SPECT), and laboratory examination were available for review; (b) “control group” (n = 26) comprised of a subsample of neurologically normal individuals from the normative study. AD subjects were recruited from a neurological clinic in Córdoba and the author’s neuropsychological private practice. All of them had been referred for a neuropsychological assessment due to complaints of cognitive decline. The diagnosis was made either by consensus of a group of professionals (composed of a neurologist, neuropsychologist and psychiatrist) or by the author, who is a neuropsychologist trained and experienced in the diagnosis of dementia.

The groups were matched in age, gender, and education as demonstrated by an ANOVA: age, $F(1, 47) = 1.18$, $p = .28$; education, $F(1, 47) = 0.36$, $p = .55$. Table 6 shows the demographic data for both groups. Patients with AD had a mean score of 105.26 ± 17.05 on the Mattis Dementia Rating Scale (Mattis, 1988).

Using the formula resulting from the regression analysis, the z-score for each subject (control and AD) was calculated. A logistic regression analysis was performed in order to determine an appropriate z-score cut-off for differentiating normal control

Table 6. Demographic data of Control and Alzheimer's disease groups

	Group	
	Control (n = 26)	Alzheimer's disease (n = 23)
Age	73.92 ± 6.67	76.3 ± 8.67
Education	11.73 ± 5.67	12.57 ± 3.82
Gender	Male 46%	Male 48%
	Female 54%	Female 52%

cases from individuals with AD. The use of logistic regression to evaluate the concurrent validity of a test is a usual procedure in neuropsychological literature (Jacobs & Donders, 2007; King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Strong & Donders, 2008). Next, the diagnostic accuracy of the derived z-score cutoff was evaluated. Using the DAG-Stat spreadsheet (Mackinnon, 2000), the following indexes were computed: sensitivity, specificity, efficiency (proportion of positives and negatives classified correctly by the test), predictive value of a positive test (probability that an observation with a positive test will be positive on the criterion), and predictive value of a negative test (probability that an observation with a negative test will be negative on the criterion).

Reliability

Stability of the CNT was assessed using the test–retest method. The sample included 80 subjects and was comprised of students and acquaintances of the administrators. These individuals were not part of the pilot or normative data samples. Mean age was 28.2 ± 11.4, and the mean of years of education was 15.3 ± 2.7. Sixty-two percent of the subjects were female. Handedness was distributed as follows: 91% right handed, 5% left handed, and 4% ambidextrous. Test–retest interval was 45.22 ± 15.55 days, with a minimum of 18 and a maximum of 77 days. Pearson correlation coefficient and Student's t-test were applied. The tests were administered by psychology students who had been properly trained in the administration of this test for the purpose of collecting the data. The administrators were allowed to collect data only after their competency in administering this test had been established. The test setting was the same as in the pilot study.

RESULTS

Normative study

All the predictors with the exception of "handedness" had a significant influence on the CNT scaled score. Results showed a direct relationship between education and performance. Males obtained higher scores on average than females. Age influenced CNT score, as subjects over 70 performed progressively worse on the test. The R^2 explained 43% of the variance in the dependent variable.

Table 7. Multiple regression analysis results

	β	*SE*	*B*	*SE*	*t*(451)	*p*
Intercept			7.2375	0.8620	8.3960	.0000
Education	0.3573	0.0393	0.2359	0.0259	9.0968	.0000
Sex	–0.3679	0.0359	–2.2740	0.2221	–10.2397	.0000
Age^2	–1.4178	0.2130	–0.0021	0.0003	–6.6551	.0000
Age	1.2949	0.2113	0.1874	0.0306	6.1274	.0000

$R = .65$, $R^2 = .43$, $F(4, 451) = 83.870$, $p < 000$; Standard Error of Estimation: 2.2865.

Results of the multiple regression analysis are displayed in Table 7. In accordance with these results the following prediction equation was formulated:

Predicted Scaled Score= constant + Bage (age) + $Bage^2$ (age^2) + Bsex(sex) + Beducation (education)

Handedness was excluded from the equation, since it did not contribute significantly to the prediction of the dependent variable.

In the clinical setting, the performance of an individual will be assessed by calculating his/her demographically adjusted *z*-score. To calculate the *z*-score, the predicted scaled score will be subtracted from the actual scaled score, and the resulting difference will be divided by the standard error of estimate.

The following formula will be used to calculate the adjusted *z*-score:

z-score = (Actual Scaled Score-Predicted Scaled Score)/Standard Error of Estimation

Figure 1 shows the frequency distribution of the raw scores of the CNT. The distribution has a slight negative skewness. Using the formulae discussed above to convert the individual raw scores to adjusted *z*-scores, the results were plotted in Figure 2. As

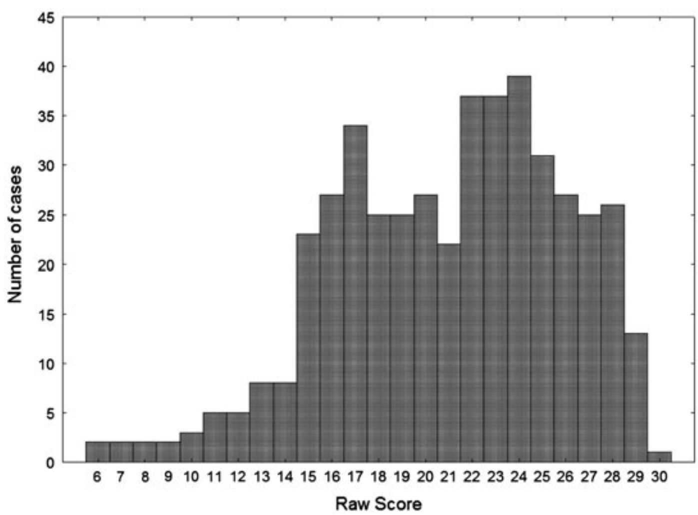

Figure 1. Histogram of the frequency distribution of raw scores of the Cordoba Naming Test.

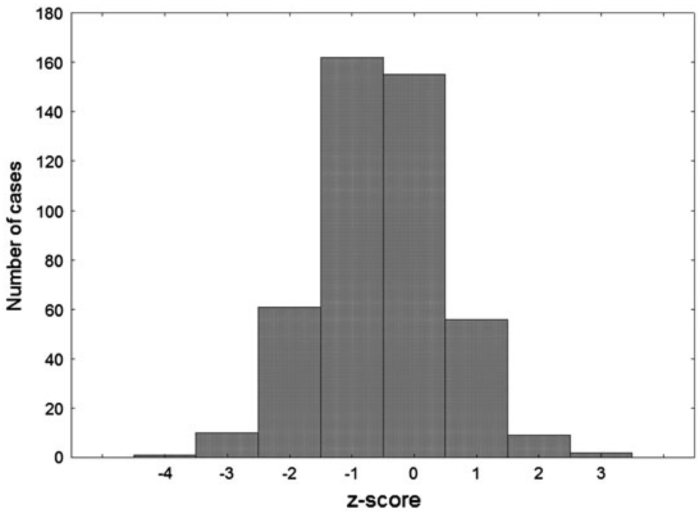

Figure 2. Histogram of the frequency distribution of adjusted *z*-scores of the Cordoba Naming Test.

can be observed, the resulting distribution is normalized. The Shapiro–Wilk *W*-test for normality was non-significant (p = .32), indicating that the distribution can be considered normal (Shapiro & Wilk, 1965).

Using the previous formulae, it is estimated that, for instance, a raw score of 14 in a 60-year-old woman with 12 years of education would yield a *z*-score of –1.4. For a man of the same age, years of education, and same raw score would yield a *z*-score of –2.4.

Validity

Mean score on the CNT for the Control group was 17.89 ± 5.02; while for the AD group it was 11.52 ± 6.1. Results showed that there is a significant difference between the control and AD groups, $F(1, 47)$= 16.03, p = .0002. The effect size was calculated yielding a coefficient η^2 = .25 which can be considered as “large” (Cohen, 1988). With this effect size the percentage of non-overlapping between samples is 62%.

The logistic regression analysis determined that a *z*-score of –1.2 is maximally sensitive and specific in differentiating the AD and control groups. Table 8 illustrates the results of the cases classification using this cut-off score. The resulting indices are

Table 8. Classification accuracy of the Cordoba Naming Test

Criterion	Test		Total
	Alzheimer’s disease	Control	
Alzheimer’s disease	17	6	23
Control	6	20	26
Total	23	26	49

as follows: sensitivity 74%; specificity 77%, efficiency 76%, predictive value of a positive test 74%, and predictive value of a negative test 77%.

Reliability

The test–retest correlation coefficient was r = .90 (p < .00). Mean scores were 22.04 ± 4.43 for the test, and 24.08 ± 4.26 for the re-test. In order to detect a possible learning effect, a t-test for dependent samples was performed. Test–retest differences were significant $t(79) = –9.19$, $p < .01$, revealing a learning effect across administrations of the test. The Reliable Change Index was calculated following the formula derived by Chelune, Naugle, Lüders, Sedlak, and Awad (1993). The following formula was employed:

$$\mathrm{RCI} = \frac{((\mathrm{X}_2 - \mathrm{X}_1) - (\mathrm{M}_2 - \mathrm{M}_1))}{\mathrm{SED}}$$

where:

X_1 is the observed pre-test score
X_2 is the observed post-test score
M_1 is the group mean pre-test score
M_2 is the group mean post-test score

SED is the Standard Error of the Difference which is calculated with the following formula: $\mathrm{SED} = \sqrt{2.(SEM)^2}$.

SEM is the Standard Error of Measurement which can be obtained with this formula: $SEM = SD\sqrt{1 - r_{xx}}$

Where:

SD is the standard deviation of the test and r_{xx} is the reliability coefficient of the test.

Using this formula, an increase of more than five points or a decrease of more than one point indicates a reliable change in the subject's performance.

DISCUSSION

The results above confirm that the CNT fulfills the basic psychometric requirements for a psychological test. The validity and reliability indexes are appropriate and the normative sample is large and heterogeneous. In comparison with other Spanish confrontation naming tests (Marquez de la Plata et al., 2008; Pontón et al., 1992a, 1992b) the CNT has a larger and more heterogeneous sample of Spanish-speaking individuals.

The regression based-norms approach used here remedies difficulties encountered with traditional procedures. For instance, with this procedure, it is estimated that a 39-year-old female with 12 years of education who scores 20 on the test will have a z-score of –0.3. This is the same z-score that a 40-year-old female with the same characteristics (except age) would obtain. Likewise, a 49-year-old male with 12 years of education who scores 20 on the CNT will obtain a z-score of –1.3; exactly the same *z*-score that a 50-year-old subject with the same characteristics would obtain. These examples prove that although both subjects belong to different age decades the *z*-scores are the same. These results would hardly be the same if the traditional approach was used to calculate the transformed score. Indeed, this can be noticed when employing the BNT norms developed by Tombaugh and Hubley (1997). Using these norms for an individual 69 years old whose education is between 13 and 21 years of school, and whose score is 60 his/her percentile is 75; meanwhile for an individual with the same characteristics but who is 70 years old his/her percentile is above 90. It is very unlikely that 1 year of variation in age can make such a difference in the confrontation naming performance, thus it is obvious that such a difference is produced by the method used to develop the norms. Furthermore, there is evidence that regression-based norms are more sensitive in the detection of cognitive impairment than traditional norms (Parmenter et al., 2010).

The impact of demographic variables is probably the most important difference between the CNT and other naming tests, since gender is not usually reported as influencing naming ability. In this study, males had significantly higher scores than females. This could be due to the inclusion of tools and weapons among the item pool (sickle, rake, screwdriver, chisel, sword, crossbow, aircraft carrier, etc.), which are generally associated with male interests. Although research with the BNT has not typically reported performance differences between genders, a study performed by Randolph, Lansing, Ivnik, Cullum, and Hermann (1999) found that males scored significantly higher than females on the BNT. This difference was also interpreted by the authors as the consequence of bias due to the presence of several items related to male interests. Zec, Burkett, Markwell, and Larsen (2007a, 2007b) also found a non-significant trend for men to score on average slightly higher than women. Although this trend was not significant the authors also provided normative data stratified by gender because of the consistent effect of gender across age groups and educational levels on the scores of the BNT.

Education exerts a considerable influence on the CNT score. However, this is not unexpected, since the test was designed to reflect this effect. The goal was to prevent the ceiling effect observed in other naming tests in which most of the subjects correctly answer the majority of the items, independent of their educational level. This ceiling effect makes impossible to detect a decline in naming ability in highly educated subjects. For example, an individual with a university degree might have a decline in his/her naming ability but this will go unnoticed, since his/her performance will be within normal limits. Consequently, such a test will generate false negatives in cases like this. The CNT, by including low frequency items, might be able to detect performance declines in highly educated subjects. Furthermore, the inclusion of low frequency items increases the sensitivity of the test since low usage frequency words are more likely to reveal difficulties in naming skills as compared to high usage frequency words (McCarthy & Warrington, 1990). This situation has been observed even in patients with semantic dementia (Jefferies, Patterson, Jones, & Lambon Ralph, 2009). At the other end of the continuum, the CNT allows for adequate assessment of the performance of subjects with low education. With a mean of 16.22 and standard deviation of 4.17, both high and low scorers of the low education group can be identified (see Table 4). This is an important feature of the test in light of the strong influence of education on neuropsychological test performance as well as the particular performance characteristics of individuals with low education on these tests (Ardila et al., 2010; Ostrosky-Solís, Ardila, Rosselli, López-Arango, & Uriel-Mendoza, 1998).

Data inspection demonstrated that age influence starts after 70 years old, where performance decays noticeably in comparison with younger groups. This finding with the CNT is similar to results produced by other research with the BNT in which age influenced performance after 70 or 80 years old (Albert, Heller, & Milberg, 1988; Allegri et al., 1997; MacKay, Connor, & Storandt, 2005; Ross, Lichtenberg, & Christensen, 1995; Zec, Markwell, Burkett, & Larsen, 2005; Zec, Burkett, Markwell, & Larsen, 2007a, 2007b).

This investigation also indicates that patients with AD performed significantly lower than controls on the CNT. This finding provides evidence of the validity of this measure for identifying word finding difficulties. Supporting these findings are the data of studies with the BNT using contrasted groups methodology which yielded very similar outcomes. Results show that patients with AD scored significantly lower than control subjects (Henry, Crawford, & Phillips, 2004; Testa, J. A. et al., 2004; Williams, Mack, & Henderson, 1989). Moreover, evidence about the difference in naming ability between AD patients and control subjects has been strengthen by meta-analysis research (Henry et al., 2004).

In this research, only patients with a mild-to-moderate degree of impairment were included. This is reflected in the mean score obtained by the patients with AD on the MDRS (Fernández & Scheffel, 2003). Thus, the difference between the mean scores on the CNT of both groups is not due to a global cognitive impairment effect but to rather specific changes in naming ability.

The diagnostic accuracy indices of the test could be considered as “Fair” in all the cases (sensitivity, specificity, accuracy, predictive value of a positive test and predictive value of a negative test), according to the criteria established by Cicchetti (2001). Although these indices are rather weak, it is necessary to consider that most of the neuropsychological tests are developed to evaluate a given ability. Even though the impairment of that ability might be frequent in a particular disease group, impairment in a single ability is typically not pathognomonic of that disease, i.e., it is not a pervasive sign of a certain disease. For example, while naming impairment is often found in patients with AD, not every AD patient shows this symptom. In fact, even when anomia is a sign that may appear in any stage of the AD, it is pervasive only in moderate and severe stages (Testa, J. A. et al., 2004).

In this study, we followed the criteria suggested by Strauss, Sherman, and Spreen (2006) for the assessment of reliability of neuropsychological tests. Hence, the reliability index of the CNT is “Very high” (.90), meaning that the test has adequate stability, which is especially important when it may need to be administered serially across time to the same patient. Research with the BNT has found similar results. Flanagan and Jackson (1997) found a .91 coefficient after a 1–2-week test–retest interval. Dikmen, Heaton, Grant, and Temkin (1999) found a .92 coefficient after an 11-month test–retest interval. Therefore, the stability study of the CNT yielded findings that were quite comparable to the results of stability assessment obtained with similar naming tests.

As mentioned earlier, a large increase in the CNT scores in successive administrations is necessary to demonstrate a considerable improvement in performance (six points), whereas a very small decrease (more than one point) means a significant decline. We believe that the significant learning effect on the CNT is responsible for this substantial difference in the magnitude of change based on the direction (i.e., increase or decrease in score). In that regard, the sample composition of the reliability study might be related to this effect. Previous data has shown that a more pronounced learning effect in neuropsychological tests is observed in younger samples as compared to older ones (Dikmen et al., 1999). Furthermore, a longitudinal study with the BNT with repeated administrations of the test failed to show large increases in the scores, despite several administrations along 10 years (Zec et al., 2005). Thus, the use of a young sample in this research for the reliability study limits the generalization of the results to older samples. Considering that this test is intended to be applied mainly in elderly samples, data on repeated administrations should be interpreted cautiously, since learning effect in this population could be less important.

An additional observation could also help explain the difference between the reliability results presented here and the data reported by other researchers. Although test administrators of the CNT did not provide the sample subjects with the correct name after a failed item, they reported that after the first administration, several subjects admitted to having deliberately searched for the names of some of the objects they could not name. This was probably motivated by their feeling embarrassed over failing to name some items and further motivated by their anticipation of a re-test session at a later time. Considering that the re-test was relatively soon after the test (45.22 ± 15.55 days) it is likely that this unpredicted behavior from the test takers significantly increased the re-test scores. Therefore, the considerable learning effect observed in this study was probably not the result of any intrinsic instability of the test, but instead, the consequence of the methodological design. Had the subjects been unaware of the re-test session or had the interval between administrations been longer, the learning effect could have been lower or irrelevant.

Due to the paucity of confrontation naming tests designed for Spanish-speakers, the CNT might be considered as a test suitable for Spanish-speaking countries. Several issues support this suggestion:

(1) the item selection was based on the use of those words in Spanish, instead of being a translation or adaptation of a test designed in a different language (this feature averts confounds related to using culturally/linguistically inappropriate items);
(2) it has been shown that it has satisfactory psychometric properties (validity and reliability); and
(3) it has a large normative sample.

However, this *a priori* assumption should be justified with empirical data. Despite polisemic items being excluded, some items names may vary across countries, especially due to idiomatic regionalisms. This, in turn, might affect the frequency usage of the item, which would call for a replacement of it within the item hierarchical order. Although one might assume that it would be necessary to develop local norms for each country, Ardila (2007) has suggested that norms could be used across countries if there is evidence that no performance difficulties are found across various same-language populations in different countries. Indeed, there is some evidence demonstrating that there are no significant differences in the performance of some tests across Spanish-speaking countries (Ramírez, Ostrosky-Solís, Fernández, & Ardila-Ardila, 2005). Nonetheless, data will be necessary to establish the stability of the CNT norms across Spanish-speaking countries.

In summary, the CNT is a valid and reliable test for the assessment of confrontation naming skills. A large heterogeneous normative sample was developed using Argentinean subjects representing a wide range of age and education as well as both genders. The CNT is potentially useful for the assessment of confrontation naming across the multiplicity of Spanish-speaking countries.

REFERENCES

Albert, M. S., Heller, H. S., & Milberg, W. (1988). Changes in naming ability with age. *Psychology and Aging, 3*, 173–178. doi:10.1037/0882-7974.3.2.173

Allegri, R. F., Mangone, C. A., Fernández Villavicencio, A., Rymberg, S., Taragano, F. E., & Baumann, D. (1997). Spanish Boston naming test norms. *The Clinical Neuropsychologist, 11*, 416–420. doi:10.1080/13854049708400471

Allison, P. D. (1999). *Multiple Regression. A primer*. Newbury Park, CA: Pine Forge Publications.

Ardila, A. (2007). The impact of culture on neuropsychological test performance. In B. P. Uzzell, M. Pontón, & A. Ardila (Eds.), *International Handbook of Cross-Cultural Neuropsychology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ardila, A., Bertolucci, P. H., Braga, L. W., Castro-Caldas, A., Judd, T., Kosmidis, M. H., … Rosselli, M. (2010). Illiteracy: The neuropsychology of cognition without reading. *Archives of Clinical Neuropsychology, 25*, 689–712. doi:10.1093/arclin/acq079

Barker-Collo, S. L. (2001). The 60-Item Boston Naming Test: Cultural bias and possible adaptations for New Zealand. *Aphasiology, 15*, 85–92. doi:10.1080/02687040042000124

Chelune, G. J., Naugle, R. I., Lüders, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology, 7*, 41–52. doi:10.1037/0894-4105.7.1.41

Cheung, R. W., Cheung, M. C., & Chan, A. S. (2004). Confrontation naming in Chinese patients with left, right, or bilateral brain damage. *Journal of the International Neuropsychological Society, 10*, 46–53. doi:10.10170S1355617704101069

Cicchetti, D. V. (2001). The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology, 23*, 695–700. doi:10.1076/jcen.23.5.695.1249

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cruice, M. N., Worrall, L. E., & Hickson, L. M. H. (2000). Boston Naming Test results for healthy older Australians: A longitudinal and cross-sectional study. *Aphasiology, 14*, 143–155. doi:10.1080/026870300401522

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods, 37*, 665–671. doi:0.3758/BF03192738

Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test-retest reliability and practice effects of expanded Halstead-Reitan neuropsychological test battery. *Journal of the International Neuropsychological Society, 5*, 346–356. doi:10.1017/S1355617799544056

Fernández, A. L., & Scheffel, D. L. (2003). A study on the criterion validity of the Mattis Dementia Rating Scale. *International Journal of Testing, 3*, 49–58. doi:org/10.1207/S15327574IJT0301_3

Flanagan, J. L., & Jackson, S. T. (1997). Test-retest reliability of three aphasia tests: Performance of non-brain-damaged older adults. *Journal of Communication Disorders, 30*, 33–43. doi:10.1016/S0021-9924(96)00039-1

Gollan, T. H., Weissberger, G., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A multi-lingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition, 15*, 594–615. doi:10.1017/S1366728911000332

Goodglass, H., & Wingfield, A. (1997). Word-finding deficits in aphasia: Brain—behavior relations and clinical symptomatology. In H. Goodglass & A. Wingfield (Eds.), *Anomia Neuroanatomical and Cognitive Correlates* (pp. 3–30). San Diego, CA: Academic Press.

Heaton, R. K., Avitable, N., Grant, I., & Matthews, C. G. (1999). Further cross validation of regression-based neuropsychological norms with an update for the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology, 21*, 572–582. doi:10.1076/jcen.21.4.572.882

Helm-Estabrooks, N., & Albert, M. L. (2004). *Manual of Aphasia and Aphasia Therapy* (2nd ed.). Austin, TX: Pro-Ed.

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: a metaanalysis. *Neuropsychologia, 42*, 1212–1224. doi:10.1016/j.neuropsychologia.2004.02.001

Huff, F. J., Corkin, S., & Growdon, J. H. (1986). Semantic impairment and anomia in Alzheimer's disease. *Brain & Language, 28*, 235–249. doi:/10.1016/0093-934X%2886%2990103-3

Ives-Deliperi, V. L., & Butler, J. T. (2012). Naming outcomes of anterior temporal lobectomy in epilepsy patients: a systematic review of the literature. *Epilepsy & Behavior, 24*, 194–198. doi:10.1016/j.yebeh.2012.04.115

Jacobs, M. L., & Donders, J. (2007). Criterion validity of the California Verbal Learning Test-Second Edition (CVLT-II) after traumatic brain injury. *Archives of Clinical Neuropsychology, 22*, 143–149. doi:/10.1016/j.acn.2006.12.002

Jefferies, E., Patterson, K., Jones, R. W., & Lambon Ralph, M. A. (2009). Comprehension of concrete and abstract words in semantic dementia. *Neuropsychogy, 23*, 492–499. doi:10.1037/a0015452

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

Kaplan, E. F., Goodglass, H., & Weintraub, S. (2001). *The Boston Naming Test* (2nd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Kim, H. L., & Na, D. L. (1999). Normative data on the Korean version of the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology, 21*, 127–133. doi:10.1076/jcen.21.1.127.942

King, J. H., Sweet, J. J., Sherer, M., Curtiss, G., & Vanderploeg, R. D. (2002). Validity indicators within the Wisconsin Card Sorting Test: Application of new and previously researched multivariate procedures in multiple traumatic brain injury samples. *The Clinical Neuropsychologist, 16*, 506–523. doi:10.1076/clin.16.4.506.13912

MacKay, A., Connor, L. T., & Storandt, M. (2005). Dementia does not explain correlation between age and scores on Boston Naming Test. *Archives of Clinical Neuropsychology, 20*, 129–133. doi:10.1016/j.acn.2004.03.006

Mackinnon, A. (2000). A spreadsheet for the calculation of comprehensive statistics of the assessment of diagnostic tests and inter-rater agreement. *Computers in Biology and Medicine, 30*, 127–134. doi:10.1076/clin.16.4.506.13912

Marien, P., Mampaey, E., Vervaet, A., Saerens, J., & De Deyn, P. P. (1998). Normative data for the Boston Naming Test in native Dutch-speaking Belgian elderly. *Brain and Language, 65*, 447–467. doi:10.1006/brln.1998.2000

Marquez de la Plata, C., Vicioso, B., Hynan, L., Evans, H. M., Diaz-Arrastia, R., Lacritz, L., & Cullum, C. M. (2008). Development of the Texas Spanish Naming Test: A test for Spanish speakers. *The Clinical Neuropsychologist, 22*, 288–304. doi:10.1080/13854040701250470

Mattis, S. (1988). *Dementia Rating Scale. Professional Manual*. Odessa, FL: Psychological Assessment Resources.

McCarthy, R. A., & Warrington, E. K. (1990). *Cognitive Neuropsychology*. New York, NY: Academic Press.

McKenna, P., & Warrington, E. (1983). *Graded Naming Test*. Windsor: Nelson.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCS – ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology, 34*, 939–944.

Ostrosky-Solís, F., Ardila, A., Rosselli, M., López-Argango, G., & Uriel-Mendoza, V. (1998). Neuropsychological test performance in illiterate subjects. *Archives of Clinical Neuropsychology, 13*, 645–660. doi:10.1016/S0887-6177%2897%2900094-2

Parmenter, B. A., Testa, S. M., Schretlen, D. J., Weinstock-Guttman, B., & Benedict, R. H. B. (2010). The utility of regression-based norms in interpreting the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *Journal of the International Neuropsychological Society, 16*, 6–16. doi:10.1017/S1355617709990750

Patricacou, A., Psallida, E., Pring, T., & Dipper, L. (2007). The Boston Naming Test in Greek: Normative data and the effects of age and education on naming. *Aphasiology, 21*, 1157–1170. doi:10.1080/02687030600670643

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

Pontón, M. O., Satz, P., Herrera, L., Young, R., Ortiz, F., D'Elia, L., Furst, C., & Namerow, N. (1992b). The Neuropsychological Screening Battery for Hispanics (NeSBHIS): Preliminary report. *Journal of Experimental and Clinical Neuropsychology, 14*, 62.

Quiñones-Ubeda, S., Peña-Casanova, J., Böhm, P., Gramunt-Fombuena, N., & Comas, L. (2004). Estudio normativo piloto de la segunda edición del Boston Naming Test en una muestra española de adultos jóvenes (20–49 años). *Neurología, 19*, 248–253.

Ramírez, M., Ostrosky-Solís, F., Fernández, A. L., & Ardila-Ardila, A. (2005). Fluidez verbal semántica en hispanohablantes: un análisis comparativo. *Revista Española de Neurología, 41*, 463–468.

Randolph, C., Lansing, A., Ivnik, R. J., Cullum, C. M., & Hermann, B. P. (1999). Determinants of confrontation naming performance. *Archives of Clinical Neuropsychology, 14*, 489–496. doi:10.1093/arclin/14.6.489

Rey, G. J., & Benton, A. L. (1991). *Multilingual Aphasia Examination-Spanish. Manual of Instructions*. Iowa City: AJA Associates.

Rey, G. J., Feldman, E., Rivas-Vazquez, R., Levin, B. E., & Benton, A. (1999). Neuropsychological test development and normative data on Hispanics. *Archives of Clinical Neuropsychology, 14*, 593–601. doi:10.1016/S0887-6177(99)00008-6

Riva, D., Nichelli, F., & Devoti, M. (2000). Developmental aspects of verbal fluency and confrontation naming in children. *Brain and Language, 71*, 267–284. doi:10.1006/brln.1999.2166

Roberts, P. M., & Doucet, N. (2011). Performance of French-speaking Quebec adults on the Boston Naming Test. *Canadian Journal of Speech-Language Pathology and Audiology, 35*, 254–267.

Ross, T. P., Lichtenberg, P. A., & Christensen, B. K. (1995). Normative data on the Boston Naming Test for elderly adults in a demographically diverse, medical sample. *The Clinical Neuropsychologist, 9*, 321–325. doi:10.1080/13854049508400496

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*, 591–611. doi:10.2307/2333709

Spinelli, M., De Oliveira Rocha, A. C., Giacheti, C. M., & Richieri-Costa, A. (1995). Word finding difficulties, verbal paraphasias, and verbal dyspraxia in ten individuals with fragile X syndrome. *American Journal of Medical Genetics, 60*, 39–43. doi:10.1002/ajmg.1320600108

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms and Commentary* (3rd ed.). New York, NY: Oxford University Press.

Strong, C. H., & Donders, J. (2008). Validity of the Continuous Visual Memory Test (CVMT) after traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 30*, 885–891. doi:10.1080/13803390701858224

Tallberg, I. M. (2005). The Boston Naming Test in Swedish: Normative data. *Brain and Language, 94*, 19–31. doi:10.1016/j.bandl.2004.11.004

Testa, J. A., Ivnik, R. J., Boeve, B., Pedersen, R. C., Pankratz, V. S., Knopman, D., Tangalos, E., & Smith, G. E. (2004). Confrontation naming does not add incremental diagnostic utility in MCI and Alzheimer's disease. *Journal of the International Neuropsychological Society, 10*, 504–512. doi:10.1017/S1355617704104177

Testa, S. M., Winicki, J. M., Pearlson, G. D., Gordon, B., & Schretlen, D. J. (2009). Accounting for estimated IQ in neuropsychological test performance with regression-based norms. *Journal of the International Neuropsychological Society, 15*, 1012–1022. doi:10.1017/S1355617709990713

Thuillard-Colombo, F., & Assal, G. (1992). Adaptation française du test de dénomination de Boston versions abrégées. *Revue Européenne de Psychology Appliquée, 42*, 67–71.

Tombaugh, T. N., & Hubley, A. M. (1997). The 60-item Boston Naming Test: Norms for cognitively intact adults aged 25 to 88 years. *Journal of Clinical and Experimental Neuropsychology, 19*, 922–932. doi:10.1080/01688639708403773

Van der Elst, W., Van Boxtel, M. P. J., Van Breukelen, G. J. P., & Jolles, J. (2006a). The Stroop Color-Word Test: Influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment, 13*, 62–79. doi:10.1177/1073191105283427

Van der Elst, W., Van Boxtel, M. P. J., Van Breukelen, G. J. P., & Jolles, J. (2006b). The Concept Shifting Test: Adult normative data. *Psychological Assessment, 18*, 424–432. doi:10.1037/1040-3590.18.4.424

Van de Vijver, F., & Tanzer, N. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 47*, 263–279. doi:10.1016/j.erap.2003.12.004

van Dort, S., Vong, E., Razak, R. A., Kamal, R. M., & Meng, H. P. (2007). Normative data on a Malay version of the Boston Naming Test. *Jurnal Sains Kesihatan Malaysia, 5*, 27–36.

Weder, N. D., Aziz, R., Wilkins, K., & Tampi, R. R. (2007). Frontotemporal dementias: A review. *Annals of General Psychiatry, 6*, 15. doi:10.1186/1744-859X-6-15

Williams, B. W., Mack, W., & Henderson, V. W. (1989). Boston Naming Test in Alzheimer's disease. *Neuropsychologia, 27*, 1073–1079. doi:10.1016/0028-3932%2889%2990186-3

Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology, 41*, 86–94. doi:10.1002/1097-4679(198501)41:1<86::AID-JCLP2270410115>3.0.CO;2-W

Zec, R. F. (1993). Neuropsychological functioning in Alzheimer's disease. In *Neuropsychology of Alzheimer's Disease and other Dementias*. New York, NY: Oxford University Press.

Zec, R. F., Burkett, N. R., Markwell, S. J., & Larsen, D. L. (2007). Normative data stratified for age, education, and gender on the Boston Naming Test. *The Clinical Neuropsychologist, 21*, 617–637. doi:10.1080/13854040701339356

Zec, R. F., Burkett, N. R., Markwell, S. J., & Larsen, D. L. (2007a). A Cross-sectional study of the effects of age, education, and gender on the Boston Naming Test. *The Clinical Neuropsychologist, 21*, 587–616. doi: 10.1080/13854040701220028

Zec, R. F., Markwell, S. J., Burkett, N. R., & Larsen, D. L. (2005). A longitudinal study of confrontation naming in the "normal" elderly. *Journal of the International Neuropsychological Society, 11*, 716–726. doi:10.1017/S13556177050897