BRAIN AND LANGUAGE **65,** 422–440 (1998)
ARTICLE NO. BL982001

Bilingual Performance on the Boston Naming Test: Preliminary Norms in Spanish and English

Kathryn J. Kohnert,*,† Arturo E. Hernandez,*,‡ and Elizabeth Bates*

**University of California, San Diego; †San Diego State University; and ‡University of California, Santa Barbara*

A total of 100 young educated bilingual adults were administered the Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 1983) in both Spanish and English. Three group performance scores were obtained: English only, Spanish only, and a composite score indicating the total number of items correctly named independent of language. The scores for the entire group were significantly greater in English than in Spanish. An additional set of analyses explored individual differences in picture naming performance across the two languages as measured by the BNT. For a subset of the larger group ($n = 25$) there were significant differences in composite over single language scoring, but no significant differences between Spanish and English. Item analyses of correct responses were conducted in both languages to explore the construct validity of the standardized administration of the BNT with this population. There was much greater variability in responses over the Spanish items for this bilingual group. The results of a correlation analysis of information obtained from the initial questionnaire with the BNT scores in each language is also reported. The practical implications of this preliminary bilingual BNT normative data are discussed.

INTRODUCTION

The U.S. currently ranks as the fifth largest Spanish-speaking country in the world (U.S. Bureau of the Census, 1996) with more than 20 million speakers. Of these identified Spanish speakers, the overwhelming majority are bilingual (Schick & Schick, 1991). There is a high rate of maintenance of the Spanish language for first generation immigrants, with a shift toward bilingualism in subsequent generations and greater English proficiency by the third or fourth generation (Bahrick, Hall, Goggin, Bahrick, & Berger,

This research was supported by the grants entitled ''Aging and Bilingualism'' (NIA Grant P50-AG05131-11) and ''Cross-linguistic Studies of Aphasia'' (NIH/NIDCD Grant 2-R01-DC00216-10). We thank Lorena Huerta, Eloisa Falcón, and Ronald Figueroa for their assistance on this project. Please address correspondence and reprint requests to Kathryn Kohnert at the Center for Research in Language, University of California, San Diego, 9500 Gilman Drive, Department 0526, La Jolla, CA 92092-0526. E-mail: kkohnert@crl.ucsd.edu.

0093-934X/98 $25.00

1994; Keefe & Padilla, 1987; Peñalosa, 1980; Sanchez, 1983). In addition, the elderly Hispanic population has been identified as the nation's most rapidly growing ethnic minority (ASHA, 1989, 1991). The bilingual adult clinical population has shown similar increases. For example, Arámbula (1992) has estimated that annually more than 66,000 Hispanic individuals in the U.S. will suffer an acquired neurological disability secondary to cerebrovascular accident (CVA), head trauma, infection, or chemical toxicity. These demographic trends, together with their cultural and linguistic implications for clinical service providers, are predicted to continue well into the next millennium.

The adequate assessment and differential diagnosis of bilingual adults with neurogenic communicative disorders presents clinicians with a substantial challenge. The three primary sources of data used by the clinical professional in this challenging process are (1) patient medical history and neurological examination results, (2) biographical information, and (3) individual performance on a battery of speech, language, and cognitive appraisal measures (Rosenbek, LaPointe, & Wertz 1989). Information from the neurological examination may direct, support, or confirm the diagnostic impressions gained from the behavioral assessment (e.g., a history of progressive memory loss and MRI documentation of cortical atrophy directs the clinician differently than reports of right hemiplegia and CT evidence of damage in the left perisylvian area). Biographical information is used to determine premorbid levels of cognitive and communicative functioning and to ascertain relevant individual, demographic, and cultural-linguistic information (e.g., languages spoken and in what contexts, the proficiency level of each language across modalities, handedness, education, etc.) which should, in turn, assist the professional in the proper administration and interpretation of individual performance on specific speech-language and cognitive appraisal measures.

In order to achieve this behavioral assessment and differential diagnosis of bilingual adults with suspected neurogenic communicative disorders, the differences between divergent cultural-linguistic factors and the potential effects of brain damage must be adequately determined (Paradis, 1997; Rosselli, Ardila, Florez, & Castro, 1990). A fundamental challenge in making this determination between neurolinguistic deficits and cultural-linguistic differences in a given individual is the lack of relevant normative data on widely employed language assessment measures (Arámbula, 1992; Lowenstein, Argüelles, Barker, & Duara, 1993; Reyes, 1995; Rosselli et al., 1990). Of potentially even greater theoretical and clinical significance is the little explored validity of such measures with culturally and linguistically diverse adult populations. Clearly there is a pressing need to obtain normative data on general language measures reflecting the demographic diversity of the U.S. population, while simultaneously exploring the validity of standardized diagnostic instruments across diverse clinical groups.

Confrontation naming has long been recognized as one of the most sensi-

tive tasks for identifying and quantifying neurogenic language deficits (e.g., Kremin, 1988; Mack, Freed, White-Williams, & Henderson, 1992; Neils et al., 1995; Welch, Doineau, Johnson, & King, 1996). The Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 1983) is the single most frequently used test of visual confrontation naming in the U.S. The BNT has been used in the evaluation of patients with focal left and right CVAs, with diffuse brain damage resulting from head injury, anoxia, or progressive dementia such as Alzheimer's disease (see Neils et al., 1995, for review). The BNT published norms include 84 normal adults (ages 18–59) with scores grouped according to those with greater or less than 12 years of education, and 82 aphasic adults with performance divided into 6 severity levels. A number of recent studies have extended the utility of the BNT by providing additional normative data related to gender, age, geographic region, educational level, and living environment (e.g., Guilford & Nawojczyk, 1988; LaBarge, Edwards, & Knesevish, 1986; Neils et al., 1995; Nicholas, Brookshire, MacLennan, Schumacher, & Porrazzo, 1989; Welch et al., 1996). Lowenstein and colleagues (1993) have also used the English-language BNT and a Spanish translation as part of their neuropsychological battery in a cross-linguistic comparison of Spanish and English speakers with mild-moderate levels of dementia. In addition, shortened equivalent versions of the BNT have been developed based on the graded nature of the original test items (Mack et al., 1992). These shortened versions have proven useful when repeated assessments requiring independent forms of a naming task are required (such as in research in Alzheimer's disease) or when administration of the complete 60-item BNT is not practical.

Despite the diagnostic usefulness of confrontation naming tasks in investigating the visual perception and lexical-semantic abilities across a range of neurogenic communication deficits, it is difficult to interpret the results obtained from patients whose demographic profile is not reflected in the normative data. To our knowledge, normative data on the BNT has been limited to monolingual speakers. Hence a first step toward accurate interpretation of naming performance of bilingual clinical populations is to obtain normative data on healthy adults who are proficient speakers of both Spanish and English.

Aside from its practical value for clinical diagnosis, validation and norming of the BNT for Spanish–English bilinguals would also prove useful for basic research on the brain bases of language and language disorders in bilingual populations. Many researchers currently use the BNT as an index of naming ability that can be correlated with one or more experimental measures (Davis, 1993). However, one cannot simply assume that the BNT is a valid measure for any population other than the one for which that instrument was first developed and normed. For example, Gomez-Tortosa and her colleagues (1995) report a case study of a 25-year-old bilingual woman with a high school education who required surgical resection of an arteriovenous malformation (AVM) in the left perisylvian area. Their subject had immi-

grated to the U.S. from a Spanish-speaking country at age 10, receiving the bulk of her formal education in the U.S. She was considered to be a proficient bilingual, although this assumption is not backed by premorbid testing in both languages. The authors also claim that Spanish is this subject's ''native language,'' based on the widely held assumption that the first language (L1) remains the native language in bilinguals who acquire their second language (L2) by or after 10 years of age. Hence performance by this patient in L1 should be equal to or better than performance in L2. The primary post-operative test of language functioning was the BNT, administered in both English and Spanish. Unfortunately the BNT in English was not administered preoperatively, and only the first 30 items were administered in Spanish prior to surgical intervention. Post-operative testing revealed that performance on the BNT was better in English (44 of 60 pictures identified correctly) than in Spanish (32 of 60 pictures identified correctly). An additional 4 responses were considered by the authors to reflect ''phonemic paraphasias.'' This post-operative performance of reduced naming on the BNT in Spanish relative to English was interpreted to be a direct result of the surgical intervention. That is, resection of the AVM was believed to have affected Spanish naming performance while leaving English naming skills intact. ''Pre-operatively this patient had a mild naming deficit in both languages that was only evident during formal language assessment. Her additional post-operative language impairment selectively affected her Spanish'' (p. 324). Based on this result, the authors conclude that there may be different anatomical substrates for the representation of a bilingual's two languages within the left perisylvian area.

A number of concerns have been raised regarding the interpretation of results in the Gomez-Tortosa et al. (1995) study. For present purposes, only the aspects relevant to the administration and interpretation of the BNT are discussed. (The reader is referred to Hines, 1996, and Paradis 1996, 1997, for additional comments.)

First, results for both the Spanish and English administrations of the BNT were interpreted within a monolingual framework. That is, the authors assume that a proficient bilingual is in fact ''balanced'' across languages and linguistic domains; hence naming skills should be equivalent in Spanish and English, and comparable to monolingual speakers of either language. However, this kind of balance in performance is rarely observed in bilingual speakers. The age and context of second language acquisition as well as current patterns of language use can all influence the lexical abilities of a bilingual speaker, in both languages.

Second, the authors' interpretation of the post-operative BNT results rests on the further assumption that Spanish was their subject's dominant language prior to surgery. This is not a safe assumption either, because skills acquired in L2 often surpass the same skills in L1, particularly among immigrant populations immersed in a second language culture (e.g., Altarriba, 1992;

Haugen, 1977; Hernandez, Bates, & Avila, 1994, 1996; Heredia, 1995, 1996, 1997; Mägiste, 1992; Smolicz, 1983). Hence we should not be surprised to find that an individual who has received most of her formal education in English since age 10 is able to name more pictures in that language.

Third, the items depicted on the BNT are assumed to have equal frequency in both Spanish and English. However, lexical items do not always share the same lexical-semantic frequency cross-linguistically or cross-culturally (Sanfeliu & Fernandez, 1996; von Studnitz & Green, 1997). It cannot be assumed, for example, that ''embudo'' has the same frequency of occurrence for Spanish monolinguals as ''funnel'' does for their English-speaking counterparts. This fact has implications for the scaling properties of the BNT. Items on the English version are graded in difficulty, from more readily named (i.e., higher frequency items) to those that are more difficult to name (i.e., lower frequency items), based on the original monolingual English-speaking normative data used in test construction. A cross-linguistic item analysis of the BNT stimuli is not presented by Gomez-Tortosa and colleagues; instead, they assume that the same progression from easy to difficult items is maintained in the Spanish translation. In the absence of norming information for the Spanish version, this assumption is unwarranted.

Putting these factors together, we have reason to suspect that naming difficulty will be correlated to pre-morbid variations in language proficiencies, reflecting the age at which L2 was acquired and the contexts in which each language is currently used (Goggin, Estrada, & Villarreal, 1994). These factors within individuals will interact in turn with cultural-linguistic variables such as word frequency and familiarity. Given these concerns, the notion that different anatomical substrates for representation of a bilingual's two languages cannot be supported on the basis of the reported BNT results in the Gomez-Tortosa et al. (1995) study. Clearly normative data for diverse populations and studies which explore the validity of frequently employed neurolinguistic assessment measures with these culturally and linguistically diverse groups are needed to support both research and clinical practice.

The primary purpose of the current study was to establish preliminary normative data on the BNT for young Spanish–English bilingual adults educated in the United States. Accurate interpretation of a bilingual's performance on traditional language measures, in either of their languages, cannot be done on the basis of a norming sample which does not reflect their demographic characteristics. This preliminary data should assist clinicians and researchers in accurately interpreting naming performance on the BNT by Spanish–English bilinguals who share similar demographic characteristics as the current study participants.

METHOD

Subjects. The participants in this study were 100 right-handed adult bilinguals of Mexican–American descent recruited from UCSD, UCSB, and the San Diego community having >12 years of formal education. Spanish was the primary home language (L1, learned from birth)

TABLE 1
Responses from Language History Questionnaire

	Speaking	Listening	Reading	Writing	Est. daily use
Spanish	6.11 (*SD* .9)	6.64 (*SD* .7)	5.92 (*SD* 1)	5.46 (*SD* 1)	30%
English	6.54 (*SD* .7)	6.77 (*SD* .6)	6.64 (*SD* .8)	6.48 (*SD* .9)	70%

Note. The mean group ($n = 100$) self-ratings on 7-point scale for skills are shown in each language. The last column indicates the mean percentage of current daily use for each language.

with English acquisition (L2) prior to the age of 8. Individuals were given course credit or paid $5.00 for their participation. All subjects completed health and language history questionnaires as well as self-ratings of speaking, listening, reading, and writing skills in each of their languages. General exclusion criteria included left-handedness, a history of speech, language, hearing, or uncorrected visual deficits, proficiency or prolonged exposure to languages other than those tested (i.e., Spanish or English) or a medical history potentially resulting in compromised neurological status (e.g., seizures, head injury, anoxia, etc.).

This young-adult bilingual group consisted of 41 males and 59 females. The mean age was 20.82 years ($SD = 2.6$) with an average educational level of 14.4 years ($SD = 1.7$). The mean age of English acquisition was 4.6 ($SD = 3.0$) years. English was the primary language of education for all participants. The mean self-rated proficiency levels in each language for speaking, listening, reading and writing on a seven-point scale (from a low of ''only a few words'' to a ceiling of ''native speaker'') are reported in Table 1.

Stimuli. The BNT consists of 60 black and white pictures graded in naming difficulty in English. That is, the test is constructed so that those objects consistently named correctly by their normative sample (e.g., bed, tree, pencil) are presented prior to pictures of less familiar objects which are less likely to be named correctly (e.g., palette, protractor, abacus). Each pictured object is presented individually in a spiral bound book. In the standardized administration respondents are asked to name each object and, when unable to do so, are provided with semantic and/or phonemic cues. Correct identification of a picture without cueing or with only a semantic cue is awarded one point. The recommended starting point for testing is item No. 30 with credit given for the previous non-tested items if this proves to be a valid basal score. Test administration is discontinued when a ceiling of 6 consecutive errors is reached. Accuracy and gross measures of latency are recorded by the examiner. (As described in the following section, these recommended basal and ceiling scoring procedures were not used in the current study but are provided here as a reference for subsequent discussion.)

Procedure. The 60-item BNT was administered two times to each participant, once in English and once in Spanish, with order of presentation counterbalanced across study participants. Instructions were given in the language of test administration. As this was an attempt to explore the validity of this measure with a specified bilingual population, the entire set of pictures was shown to each participant, in each language (that is, no pre-established basal or ceiling values were used). A Spanish protocol and semantic cues were developed based on the original English BNT. The test was administered by a bilingual speech-language Pathologist or a trained bilingual research assistant. Total correct scores were defined as answers given spontaneously or with a semantic cue only.

Scoring and data analysis. The total number of pictures named in each language was recorded. Alternative responses reflecting dialectal or acceptable lexical variations in Spanish were credited (e.g., ''sierra'' or ''serucho'' were both accepted for the pictured saw). In addition to the single language totals, a composite score indicating non-duplicated responses from both English and Spanish was derived for each bilingual participant. That is, the composite score reflected the total number of pictures named by an individual from the possible 60, independent of language. As in the single language testing condition of the BNT, the greatest

earned score possible is 60. That is, if ''bed'' and its Spanish equivalent ''cama'' are both used to name the first picture on the BNT only one point is awarded because the labels are cross-language duplicates. However, if the picture ''snail'' is identified in *either* Spanish or English, one point is awarded in the composite scoring method. Group means, standard deviations, and 95% confidence intervals are reported for Spanish, English, and Composite BNT scores.

We also anticipated that there would be significant variability within this relatively large sample of Spanish–English college students, both in terms of naming proficiency in each language and in the potential benefit derived from the single-language vs composite-language scoring methods. To explore this variability, a cross-language difference score was obtained for each individual by subtracting his/her Spanish BNT score from his/her English BNT total. We classified as ''English dominant'' or ''Spanish dominant'' any individual whose difference score was greater than one standard deviation from the mean difference score for the group as a whole (English dominant if the English score was larger; Spanish dominant if the Spanish score was larger). Any individual whose difference score fell within one standard deviation of the mean was treated as Balanced (relative to the sample as a whole; this does not mean that competence in the two languages is identical). The three raw BNT scores, the difference score itself, and these three classifications were all entered into correlational analyses with the various scores on the language history questionnaire, permitting a validation of BNT performance against independent self-reports of language proficiencies with this group of bilingual participants.

Finally, item analyses were conducted to determine whether the graded design of the BNT (from easily named items to more difficult items) was maintained in both languages. The total number of subjects who named each picture correctly was determined for each language, and ordered from ''easy'' (highest proportion named) to ''hard'' (lowest proportion named) in English. The corresponding values for Spanish were inspected visually, and through correlation over items.

RESULTS

Overall BNT results. Summary data from the 100 young bilingual adults in the current study are presented in Table 2. The sample mean score in English was 46.66 (*SD* 6.64) [95% confidence interval of 34.06–59.26], in Spanish the mean was 32 (*SD* 8.83) [CI of 15.23–48.78], and the mean Composite score was 48.59 (*SD* 5.57) [CI 38.01–59.17]. These mean values are

TABLE 2
BNT Results

	Mean	*SD*	95% confidence interval
Spanish***	32.00	8.83	15.23 to 48.78
English	46.66	6.64	34.06 to 59.26
Composite	48.59	5.57	38.01 to 59.17

Note. The group means, standard deviations, and 95% confidence intervals for the Spanish, English, and Composite BNT scores are shown ($n = 100$). The Spanish scores were significantly different from both the English and Composite scores ($p < .0001$).

close to the respective means for English (44 out of 60) and Spanish (32 out of 60) reported by Gomez-Tortosa for the post-operative performance of their Spanish–English bilingual patient.

A one-way analysis of variance comparing the Spanish, English, and composite scores revealed a main effect of language: $F(2, 97) = 161.3350$, $p < .0001$. Pairwise comparisons using the Tukey HSD test (preserving alpha at .05) showed that the BNT scores were significantly greater in English than in Spanish, suggesting that the majority of our subjects are English dominant (see below). Furthermore, the composite scores were significantly greater than the Spanish scores, but they were not significantly greater than the English scores alone. This means that, for the group as a whole, the pictures that were named in Spanish comprise a subset of the words that were named in English.

Individual differences and sub-group comparisons. A cross-language mean difference score (MDS) was obtained for each individual by subtracting his/her BNT Spanish score from his/her BNT English score. Thus, those individuals who had similar scores in both languages would have a smaller absolute MDS (e.g., subject No. 28 had a BNT Spanish Score of 48 and a BNT English Score of 50, yielding an absolute MDS of 2) than those participants who were clearly stronger in one of the languages (e.g., subject No. 24 had a BNT Spanish score of 22 and a BNT English score of 50, yielding an absolute MDS of 28). Although this scoring method would permit division into three groups (English-dominant, Spanish-dominant, Balanced), here there are only two groups as dominance is defined as a difference score greater than one standard deviation (7 raw points) from the mean difference score for the group as a whole. A total of 25 study participants were classified as Balanced, and 75 were classified as English-dominant. A 2 (group) $\times$ 3 (language/method of scoring: Spanish, English, Composite) analysis of variance was conducted on total naming scores, with group as a between-subjects variable and scoring system as a within-subjects variable. The analysis revealed a main effect of score type ($F[2, 94] = 87.58$, $p < .0001$) and a significant interaction of group and score type ($F[2, 94] = 43.32$, $p < .0001$). Pairwise comparisons ($p < .05$) showed that Spanish BNT scores for the Balanced group were greater than Spanish scores for the English-dominant group, while English BNT scores for the English-dominant group were greater than those of the Balanced group. There were no reliable between-group differences in the Composite scores.

Within the Balanced language group, the mean BNT score in English was 42.04 (*SD* 8.02), the mean in Spanish was 40.88 (*SD* 7.01), and the mean Composite score was 46.48 (*SD* 6.77). A one-way analysis of variance within this group alone revealed a significant effect of language/scoring method ($F(2, 22) = 4.11$, $p < .02$). Post hoc comparisons using the Tukey HSD test with the preserved family-wise alpha of .05 revealed a significant difference between the mean BNT Spanish score and the Composite score as well as

between the mean BNT English score and the Composite score. For this group, the difference between the English and Spanish mean BNT scores was not significant (which is, of course, what we would expect given the criterion by which this group was derived).

Within the English-dominant group, the mean English BNT score was 48.2 (*SD* 5.33), the mean in Spanish was 29.04 (*SD* 7.27), and the mean Composite score of 49.29 (4.96). A one-way analysis of variance within this group revealed a significant effect of language/scoring method for this group ($F(2, 72) = 275.68$, $p < .0001$). Post hoc comparisons using the Tukey HSD test with a family-wise alpha of .05 indicated a significant difference between the mean English and Spanish BNT scores, and between the Spanish and Composite BNT scores. In contrast to the first group, the difference between the English BNT and the Composite scores did not reach significance.

Individual differences potentially affecting the results of the BNT scores were further explored with correlational analyses. As Table 1 shows, these participants tend (as a group) to rate their English skills higher than their Spanish skills in all modalities except for passive listening. However, there is some variability in ratings within the group. BNT scores for each language were correlated with these self-ratings of proficiency in each language, across modalities (i.e., speaking, listening, reading, writing), together with language history (e.g., age of English acquisition), and current patterns of use (e.g., percent of time spent each day speaking Spanish). Results are summarized in Table 3. Although these correlations are modest (ranging from $-.42$ to $+.37$), most of them are significant and easily interpretable, providing cross-validation for both the English and the Spanish naming scores. English BNT scores are positively correlated with age, education, and with all self-ratings of English proficiency; they are negatively correlated with age of English (L2) acquisition (a typical finding in this literature), and with self-ratings of speaking, reading, and writing proficiency in Spanish. Interestingly, there is no relationship between English BNT scores and self-ratings of listening skills in Spanish, suggesting that passive understanding of their first language has not been eroded by increased proficiency in English in this population. Spanish BNT scores are positively correlated with age, education, the amount of Spanish spoken every day, and with all self-ratings of Spanish proficiency; they are negatively correlated with self-ratings of speaking and listening in English, but not with self-ratings of reading and writing in English. There is no relationship between Spanish naming scores and age of L2 acquisition, which means that Spanish naming skills are not necessarily damaged by age of English (L2) exposure.

Item analyses. The item analysis in English revealed an overall pattern of increasing word naming difficulty, generally consistent with what is considered to be construct validity of this language assessment measure. That is, during the English BNT administration items presented earlier in the test were more likely to be named correctly than those items presented during

TABLE 3
BNT Correlations

	Spanish-BNT	English-BNT
Gender	−.081	−.117
Age	.257**	.3416***
Education	.2106*	.2853**
Age of L2 acquisition	.1183	−.416**
Daily use of Spanish	.2856**	−.1450
Spanish speaking	.3041***	−.2218*
Spanish–listening	.2828**	−.0289
Spanish–reading	.2431**	−.1848*
Spanish–writing	.3168***	−.2187*
English–speaking	−.2675**	.3707***
English–listening	−.1738*	.3165***
English–reading	−.0419	.4403***
English–writing	−.0731	.3957***

Note. Correlations for the obtained BNT scores in Spanish and English with information obtained from the language history questionnaire and participants' self-rated skills in each language (on a 7-point scale) are shown.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

the last half of testing (Fig. 1). In contrast, the item analysis in Spanish revealed a highly variable response pattern to the BNT stimulus items (Fig. 2). The pattern of correct responses was much more varied throughout the stimulus presentation so that the actual ordering of the test items was not a good predictor of naming accuracy. A numerical listing of items together

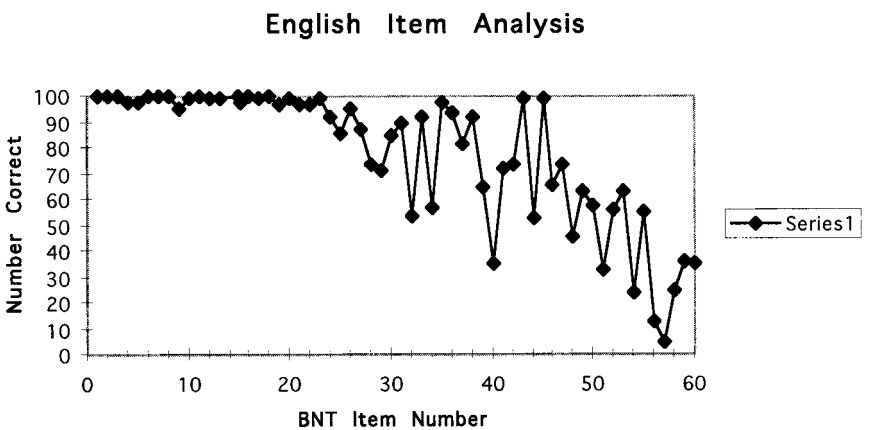

FIGURE 1

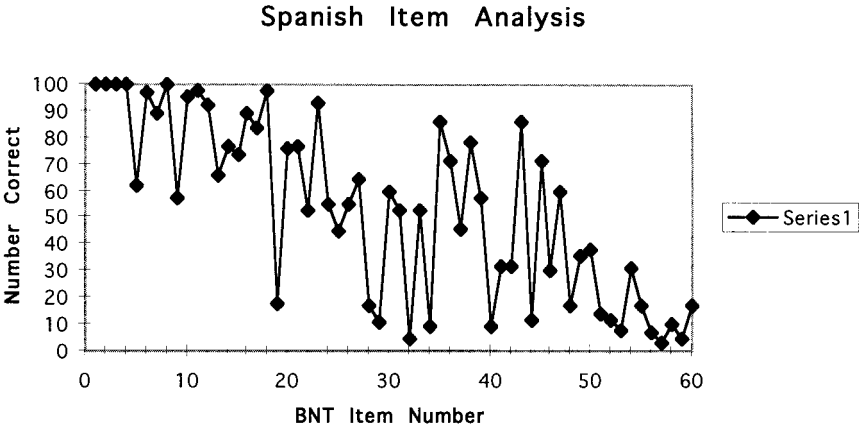

FIGURE 2

with the percentage of individuals passing each numbered item in Spanish and in English is presented in Table 4. This table also shows the direction and percentage of change between each item for both Spanish and English as a further indication of variability across consecutive stimulus items (the ''between item difference score''). The absolute value of the difference score in English (a measure of the total between-item variability regardless of direction) was 783 and for Spanish it was 1436, with means of 13.05 and 23.93, respectively. The split half mean between item difference in English was 3.03 for the first 30 items and 23.07 for the latter half of the items. In Spanish, however the split half means were more equivalent, with means of 20.93 and 26.93 for the first and second halves, respectively. This indicated that the sequential organization of the BNT items was not graded in difficulty (from easiest to hardest) in Spanish, as the test is constructed to be in English. More generally, however, there was a positive correlation between the degree of difficulty in English and the degree of difficulty in Spanish (as indicated by an over-items correlation of .61 between the two languages). The implications of these combined results are discussed in the next section.

DISCUSSION AND IMPLICATIONS

A total of 100 Spanish–English bilingual adults with ≥12 years of formal education participated in this study. All participants learned Spanish as a first language in the home environment, and began to learn English before the age of 8. The majority of their formal education was in English. Participants considered themselves to be proficient (albeit generally not ''balanced'') bilinguals. Individual administration of the Boston Naming Test, a measure of visual confrontation naming, was carried out in both languages to obtain preliminary normative data for healthy adult bilinguals. We found that pic-

TABLE 4

BNT item	Spanish %	English %	Spanish between-item difference score	English between-item difference score
1. bed	100	100	—	—
2. tree	100	100	0	0
3. pencil	100	100	0	0
4. house	100	100	0	0
5. whistle	62	98	−38	−2
6. scissors	97	100	35	2
7. comb	89	100	−8	0
8. flower	100	100	11	0
9. saw	57	95	−43	−5
10. toothbrush	95	99	38	4
11. helicopter	98	100	3	1
12. broom	92	99	−6	−1
13. octopus	66	99	−26	0
14. mushroom	77	100	11	1
15. hanger	74	98	−3	−2
16. wheelchair	89	100	15	2
17. camel	84	99	−5	−1
18. mask	98	100	14	1
19. pretzel*	18	97	−80	−3
20. bench	76	99	58	2
21. racquet	77	97	1	−2
22. snail	53	97	−24	0
23. volcano	93	53	40	2
24. seahorse	55	99	−38	−7
25. dart	45	92	−10	−6
26. canoe	55	86	10	−9
27. globe	64	95	−9	−8
28. wreath	17	87	−47	−13
29. beaver	11	74	−6	−3
30. harmonica	60	71	−49	14
31. rhinoceros	53	85	−7	5
32. acorn	5	90	−48	−36
33. igloo	53	54	48	38
34. stilts	9	92	−34	−35
35. dominoes	86	57	75	41
36. cactus	71	98	−15	−4
37. escalator	46	94	−25	−12
38. harp	78	82	32	10
39. hammock	57	65	−21	−27
40. knocker	9	35	−48	−30
41. pelican	32	72	23	37
42. stethoscope	32	74	0	2
43. pyramid	86	99	54	25
44. muzzle	12	53	−74	−46
45. unicorn	71	99	59	46
46. funnel	30	66	−41	−33

TABLE 4—*Continued*

BNT item	Spanish %	English %	Spanish between-item difference score	English between-item difference score
47. accordion	60	74	30	8
48. noose	17	46	−43	−28
49. asparagus	36	63	19	17
50. compass	38	58	−2	−5
51. latch	14	33	−24	−25
52. tripod	12	56	−2	23
53. scroll	8	63	−4	7
54. tongs	31	24	28	−39
55. sphynx	17	55	−14	31
56. yoke	7	13	−10	−42
57. trellis	3	5	−4	−8
58. palette	10	25	7	20
59. protractor	5	36	−5	11
60. abacus	17	35	12	−1

Note. BNT items are shown in the order of presentation along with the percentages of subjects naming the picture correctly in each language. The ''between-item difference scores'' show the direction and percentage of increase or decrease in naming accuracy between consecutive responses in each language, based on the inter-stimulus response variation from the first two columns.

ture-naming performance as measured by the BNT was significantly better in English than in Spanish for this study sample. Seventy-five percent of the sample was English dominant and 25% was relatively (albeit imperfectly) balanced. These findings were consistent with the participants' responses on the language history questionnaires, including self-ratings of relative language skills.

Although BNT performance for these bilinguals was significantly better in English than in Spanish, a direct comparison of the obtained English scores with those of the available monolingual norms is not warranted. With respect to the linguistic variability inherent in a bilingual–monolingual comparison, the expectation that ''dominant'' bilinguals will (in their stronger language) behave like monolinguals is highly suspect on both practical and theoretical grounds (e.g., Grosjean, 1992, 1997; Gutierrez-Clellen, 1996; Hernandez, Bates, & Avila, 1994, 1996; Paradis, 1997; Reyes, 1995). It is also, at the very least, uninformative. That is, in comparing the overall group results from the current study to the available monolingual English normative data, little information could be gained which would help differentiate normal from impaired naming performance. For example, the normative data on monolingual adults with >12 years of education indicates a mean score of 55.71 (Kaplan, Goodglass, & Weintraub, 1983) whereas the mean BNT total

in English for our young-adult bilingual sample was 46.66, with a 95% confidence interval of 34–59. Differences in scores between the monolingual norms and the current study results (for groups equated on years of education) are likely explained by a combination of cultural, linguistic, and experiential variables. Age differences within the adult category across studies may also play a role (i.e., the Goodglass and Kaplan norms are based on 31 adults ranging from 18 to 59 years of age, and were collected in the 1960s when items like ''abacus'' may have been recognizable to a broader population). The positive correlation of age and education with *both* Spanish and English BNT performance found in the current study supports this latter notion of increased naming skill in bilinguals, in both their languages, even across the period of young adulthood covered by this study.

At the same time, however, it is worth underscoring that the respective means for English (46.7) and Spanish (32) in our study are strikingly similar to the post-operative means in English (44) and Spanish (32) in the bilingual case study reported by Gomez-Tortosa et al. (1995). Assuming that Spanish performance should be equal to or better than English performance in this bilingual (who acquired English at 10 years of age and was educated almost exclusively in English after that point), Gomez-Tortosa et al. conclude that the operation itself has selectively impaired the neural representation of Spanish. Our data suggest an alternative interpretation: this woman might have displayed a similar pattern of English dominance if she had been tested pre-operatively, with both versions of the full BNT. In summary, it is imperative that representative group normative data be used to adequately interpret performance on the BNT. Given the normal variability in language performance within any bilingual group, it is also important to consider normal individual differences potentially affecting test performance.

In the current study individual differences were looked at in two ways. First, we derived a cross-language mean difference score (MDS) of BNT performance for each participant. Based on the distribution of the MDS (i.e., Spanish BNT–English BNT) two different sub-groups were identified: a relatively balanced group (MDS less than one standard deviation) and an English-dominant group. Despite the fact that Spanish was the first language for all the participants in our study, with exposure to English varying from birth to age 8, none of the participants was Spanish-dominant by our criterion (see Table 3).

The first group ($n = 25$) was those individuals who were relatively balanced in their performance on the BNT across the two languages (MDS $\leq$ 1 standard deviation). In this ''balanced language'' group there were no reliable differences between Spanish and English BNT scores. In addition, this subgroup also benefited significantly from the composite-over single-language scoring method in which each non-duplicate picture named, independent of language, is credited. That is, the alternative scoring method was a much better indicator of this group's confrontation naming skills than the

single language only testing. Clear implications from these findings are that a dual-language approach to both assessment and intervention are indicated for a substantial portion (25%) of our overall sample.

In contrast, the second group ($n = 75$) was relatively ''unbalanced'' in their cross-language performance on the BNT. These individuals scored significantly better in English than in Spanish and showed no reliable benefit from the composite- over single-language scoring procedure. Here the implications for single- over dual-language assessment and training are less clear. These combined results highlight the magnitude of inter-individual variability even within a relatively homogeneous group of bilinguals, and reinforce the need for additional considerations in interpreting bilingual performance on static language measures. It is also worth noting that there were no between-group differences in the composite scores obtained on the BNT (i.e., the English dominant group did not outperform the balanced bilingual group in terms of the total number of pictures adequately identified).

The second way in which we looked at individual differences was through the use of self-reports of language history, current use, and skill ratings across modalities in Spanish and English. Measuring language proficiency in healthy adult bilinguals is not a trivial task. There is no consensus on what ''proficiency'' is and even less on how to quantify the set of language skills that might comprise cross-linguistic proficiency. Most empirical studies in the adult bilingual literature use self-report instruments as their primary or sole measure of language ability (e.g., see Grosjean 1997). Rarely, however, are results of these studies reported in terms of this preliminary information (see Goggin, Estrada, & Villarreall, 1994, for an exception). It is therefore unclear if these subject self-ratings are indeed effective indicators of linguistic abilities. In the current study the relationship between self-reports and obtained scores on the BNT was explored through a correlation analysis.

Information obtained from the language history questionnaire and the self-ratings of language skills was positively correlated with performance on the BNT in the following ways: Higher self-ratings in Spanish across modalities were associated with higher Spanish BNT scores, and higher self-ratings in English across the language skill modalities were associated with higher English BNT scores. This supports the use of well-designed questionnaires and self-rating scales with young-adult bilinguals as one independent measure of relative language proficiency in a research assessment battery. It also suggests that similar types of self-report instruments might be useful in clinical settings. There were several negative correlations in the current study, although their interpretation is somewhat less straightforward. For example, greater age at the time of second language (English) acquisition was related to lower English BNT scores, even though all study participants began learning English prior to 8 years of age. It is possible that this relationship would diminish across adulthood (as indicated by the previously noted positive relationship found between subject age at the time of testing and total English BNT scores). The negative result may also reflect demographic factors (e.g.,

complex patterns of language maintenance and use) that are not detected by this questionnaire. The second set of reliable negative correlations indicate some kind of trade-off in English and Spanish proficiency. For example, higher scores on the Spanish BNT were associated with lower self-ratings of speaking and listening in English, but were unrelated to self-rated proficiency in English reading and writing (the language in which they have received all or most of their formal education). Conversely, higher scores on the English BNT were negatively correlated with self-rated proficiency in Spanish, with one interesting exception: Spanish listening skills. The profile that emerges here is one that is familiar to many young Spanish–English bilinguals in southern California: the ability to understand Spanish-speaking family members remains despite immersion in an English-language educational setting, but the ability to read and write in Spanish may suffer, and to some extent there may also be some erosion in the ability to speak Spanish. In general, the data reported here suggest that patterns of continuing language use may be as important in the acquisition and maintenance of proficiency in two languages as the much-heralded issue of age of acquisition (i.e., the issue that is usually discussed in the framework of critical period theories of second language learning). Certainly our data suggest that the first language does not necessarily remain the stronger or more fluent language for this population of bilinguals.

In addition to the need for demographically representative norms for widely used language measures and the importance of individual variation within these groups, it is also necessary to explore the validity of these measures themselves. That is, can we appropriate tests constructed on one language group and apply them directly to another? This practice has come under serious scrutiny in research with bilingual children (e.g., Gutierrez-Clellen, 1996; Kayser, 1995; Umbel, Pearson, Fernandez, & Oller, 1992) but has rarely been questioned with research and/or clinical practice with proficient bilingual adults. The BNT and related measures of vocabulary assess knowledge of a specific set of lexical items. The pictures and words used to assess this lexical knowledge are, by design, graded in difficulty so that language deficits can be identified. Easily named high frequency lexical items are placed at the beginning of the test, followed by lower frequency items that are more difficult to name. In the standardized administration of the BNT the recommended starting point for testing is item No. 30. Credit is given for the previous non-tested items if this proves to be a valid basal score. Test administration is discontinued when a ceiling of 6 consecutive errors is reached. The item analysis for our English data revealed an overall response pattern that was generally consistent with the design of the BNT and its original norming base. That is, those items at the beginning of the test were more likely to be named correctly than those occurring at the end. In contrast, the item analysis in Spanish revealed a highly variable response pattern to the BNT stimulus items throughout the measure, so that the actual ordering of the test items was not a valid predictor of naming accuracy. This

variability calls into question the practice of using predetermined basal and ceiling scores in the administration and interpretation of tests that are employed with groups for whom they were not constructed. The use of such predetermined scoring criteria may lead to erroneous interpretations of the presenting skills, deficits, and recovery patterns in the linguistically diverse adult neurogenic population. Various shortened versions of the BNT in English have also been developed for repeated testing of clinical and research populations. The current study findings indicate that these shortened versions cannot be applied with equal success to bilingual populations.

CONCLUSIONS

Despite the significant challenges posed by attempts to characterize naming skills in bilingual adults, the prominent role of naming deficits in neurogenic language impairments clearly indicates that it is a challenge worth facing. The question is not whether we should use naming as a language measure with culturally and linguistically diverse adults, but rather, how this should be done. The current study takes a step in this direction by providing normative data on the BNT for a specific group of Spanish–English bilinguals. It also exposes some of the problems that are encountered in using standardized tests with populations for which they were not originally developed. Future investigations exploring assessment issues with other adult bilinguals are needed, most significantly with elderly and clinical populations.

It should also be emphasized that although the normative data reported in the current study can serve as a useful guideline for researchers and clinical service providers who are working with bilingual adults, the heterogeneity of the Spanish–English bilingual population in the United States is considerable. Educational level, dialect, country of origin, geographical location, socioeconomic status, gender, age, proficiency of each language, and context and age of acquisition of each of the languages must all be considered in the assessment of language skills. Even within a relatively homogeneous young bilingual group, such as that included in this study, significant individual differences exist. Part of the value of the present study is that it does, in fact, highlight the unique performance of this group of bilingual individuals on a commonly used test of confrontation naming.

REFERENCES

Altarriba, J. 1992. The representation of translation equivalents in bilingual memory. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (Vol. 83, pp. 157–174). Amsterdam: Elsevier.

American Speech-Language-Hearing Association. 1989. *Office of minority concerns newsletter,* 10 (1). Rockville, MD: American Speech-Language-Hearing Association.

American Speech-Language-Hearing Association. 1991. Cultural diversity in the elderly population. *Asha,* **33,** 66.

Arámbula, G. 1992. Acquired neurological disabilities in Hispanic adults. In H. W. Langdon & L. L. Cheng (Eds.), *Hispanic children and adults with communication disorders* (pp. 373–407). Gaithersburg, MD: Aspen.

Bahrick, H. P., Hall, L. K., Goggin, J. P., Bahrick, L. E., & Berger, S. A. 1994. Fifty years of language maintenance and language dominance in Bilingual Hispanic immigrants. *Journal of Experimental Psychology: General,* **123,** 264–283.

Davis, G. A. 1993. *A survey of adult aphasia and related language disorders* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Goggin, J. P., Estrada, P., & Villarreal, R. P. 1994. Picture-naming agreement in monolinguals and bilinguals. *Applied Psycholinguistics,* **15,** 177–193.

Gomez-Tortosa, E., Martin, E., Gaviria, M., Charbel, F., & Ausman, J. 1995. Selective deficit of one language in a bilingual patient following surgery in the left perisylvian area. *Brain and Language,* **48,** 320–325.

Grosjean, F. 1992. *Life with two languages: An introduction to bilingualism.* Cambridge, MA: Harvard Univ. Press.

Grosjean, F. 1997. Processing mixed languages: Issues, findings, and models. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism* (pp. 255–278). Mahwah, NJ: Erlbaum.

Guilford, A. M., & Nawojczyk, D. C. 1988. Standardization of the Boston Naming Test at the kindergarten and elementary school levels. *Language, Speech, and Hearing Services in the Schools,* **19,** 395–400.

Gutierrez-Clellen, V. F. 1996. Language diversity: Implications for assessment. In K. N. Cole, P. S. Dale, & D. J. Thal (Eds.), *Assessment of communication and language: Vol. 6. Communication and language intervention series* (pp. 29–56). Baltimore: Brookes.

Haugen, E. 1977. Norm and deviation in bilingual communities. In P. A. Hornby (Ed.), *Bilingualism: Psychological, social, and educational implications* (pp. 91–102). New York: Academic Press.

Heredia, R. 1995. *Concreteness effects in high frequency words: A test of the revised hierarchical and the mixed models of bilingual memory representations.* Unpublished doctoral dissertation, University of California, Santa Cruz.

Heredia, R. 1996. Bilingual memory: A re-revised version of the hierarchical model of bilingual memory. *CRL Newsletter,* **10.**

Heredia, R. 1997. Bilingual memory and hierarchical models: A case for language dominance. *Current Directions in Psychological Science,* **6**(2), 34–39.

Hernandez, A. E., Bates, E., & Avila, L. X. 1994. Sentence interpretation in Spanish-English bilinguals: What does it mean to be in-between. *Applied Psycholinguistics,* **15,** 417–466.

Hernandez, A. E., Bates, E., & Avila, L. X. 1996. Processing across the language boundary: A cross modal priming study of Spanish-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **22,** 846–864.

Hines, T. M. 1996. Failure to demonstrate selective deficit in the native language following surgery to the left perisylvian area. *Brain and Language,* **54,** 168–169.

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

Kayser, H. (Ed.). 1995. *Bilingual Speech Pathology: An Hispanic Focus.* San Diego: Singular Publishing.

Keefe, S. E., & Padilla, A. M. 1987. *Chicano ethnicity.* Albuquerque: Univ. of New Mexico Press.

Kremin, H. 1988. Naming and its disorders. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (pp. 307–328). Amsterdam: Elsevier.

LaBarge, E., Edwards, D., & Knesevish, J. 1986. Performance of normal elderly on the Boston Naming Test. *Brain and Language,* **27,** 380–384.

Lowenstein, D. A., Argüelles, T., Bartker, W. W., & Duara, R. 1993. A comparative analysis of neuropsychological test performance of Spanish-Speaking and English-speaking patients with Alzheimer's Disease. *Journal of Gerontology,* **48**(3) 142–149.

Mack, W. J., Freed, D. M., White-Williams, B., Henderson, V. W. 1992. Boston Naming test: Shortened versions for use in Alzheimer's disease. *Journal of Gerontology,* **47**(3), 154–158.

Mägiste, E. 1992. Second language learning in elementary and high school students. *European Journal of Cognitive Psychology,* **4**(4), 355–365.

Neils, J., Baris, J. M., Carter, C., Dell'aira, A. L., Nordloh, S. J., Weiler, E., & Weisiger, B. 1995. Effects of age, education, and living environment on Boston Naming Test performance. *Journal of Speech and Hearing Research,* **38,** 1143–1149.

Nicholas, L. E., Brookshire, R. H., MacLennan, D. L., Schumacher, J. G., & Parrazzo, S. A. 1989. Revised administration and scoring procedures for the Boston Naming Test and norms for non-brain-damaged adults. *Aphasiology,* **3,** 569–580.

Paradis, M. 1996. Selective deficit in one language is not a demonstration of different anatomical representation: Comments on Gomez-Tortosa et al. (1995). *Brain and Language,* **54,** 170–173.

Paradis, M. 1997. The cognitive neuropsychology of bilingualism. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism* (pp. 331–354). Mahwah, NJ: Erlbaum.

Peñalosa, F. 1980. *Chicano sociolinguistics: A brief introduction.* Rowley, MA: Newbury House.

Reyes, B. A. 1995. Considerations in the assessment and treatment of neurogenic disorders in bilingual adults. In H. Kayser (Ed.), *Bilingual speech-language pathology: An Hispanic focus* (pp. 153–182). San Diego: Singular.

Rosenbek, J. C., LaPointe, L. L., & Wertz, R. T. 1989. *Aphasia: A clinical approach.* San Diego: Singular.

Rosselli, M., Ardila, A., Florez, A., & Castro, C. 1990. Normative data on the Boston Diagnostic Aphasia Examination in a Spanish-speaking population. *Journal of Clinical and Experimental Neuropsychology,* **12**(2), 313–322.

Sanchez, R. 1983. *Chicano discourse: Socio-historic perspectives.* Rowley, MA: Newbury House.

Sanfeliu, M. C., & Fernandez, A. 1996. A set of 254 Snodgrass-Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers,* **28**(4) 537–555.

Schick, F. L., & Schick, R. 1991. *Statistical handbook on U.S. Hispanics.* Phoenix, AZ: Oryx Press.

Smolicz, J. J. 1983. Modification and maintenance: Language among schoolchildren of Italian background in South Australia. *Journal of Multilingual and Multicultural Development,* **4,** 313–337.

Umbel, V. M., Pearson, B. Z., Fernandez, M. C., & Oller, D. K. 1992. Measuring bilingual children's receptive vocabularies. *Child Development,* **63,** 1012–1020.

U. S. Bureau of the Census. 1996. The Hispanic population in the United States. *In Current Population Reports.* Washington, DC: Government Printing Office.

von Studnitz, R. E., & Green, D. W. 1997. Lexical decision and language switching. *The International Journal of Bilingualism,* **1**(1), 3–24.

Welch, L. W., Doineau, D., Johnson, S., & King, D. 1996. Educational and gender normative data for the Boston Naming Test in a group of older adults. *Brain and Language,* **53,** 260–266.