

# Performance on the Boston Naming Test in Bilinguals

Christine Sheppard,<sup>1</sup> Shanna Kousaie,<sup>1</sup> Laura Monetta,<sup>2,3</sup> AND Vanessa Taler<sup>1,4</sup>

<sup>1</sup>Bruyère Research Institute, Ottawa, Canada

<sup>2</sup>Département de Réadaptation, Université Laval, Québec City, QC, Canada

<sup>3</sup>Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec, Québec City, QC, Canada

<sup>4</sup>School of Psychology, University of Ottawa, Ottawa, ON, Canada

(RECEIVED November 6, 2014; FINAL REVISION November 3, 2015; ACCEPTED November 18, 2015; FIRST PUBLISHED ONLINE December 21, 2015)

## Abstract

**Objectives:** We examined performance on the Boston Naming Test (BNT) in older and younger adults who were monolingual English or French speakers, or bilingual speakers of English and French ( $n = 215$ ). **Methods:** Monolingual participants completed the task in their native language, and bilingual participants completed the task in English, French, and bilingual (either-language) administrations. **Results:** Overall, younger and older monolingual French speakers performed worse than other groups; bilingual participants performed worst in the French administration and approximately two-thirds of bilingual participants performed better when responses were accepted in either language. Surprisingly, however, a subset of bilinguals performed worse when responses were accepted in either language as compared to their maximum score achieved in either English or French. This either-language disadvantage does not appear to be associated with the degree of balanced bilingualism, but instead appears to be related to overall naming abilities. Differential item analysis comparing language groups and the different administrations identified several items that displayed uniform and/or non-uniform differential item functioning (DIF). **Conclusions:** The BNT does not elicit equivalent performance in English and French, even when assessing naming performance in monolingual French speakers using the French version of the test. Scores were lower in French overall, and several items exhibited DIF. We recommend caution in interpreting performance on these items in bilingual speakers. Finally, not all bilinguals benefit from an either-language administration of the BNT. (*JINS*, 2016, 22, 350–363)

**Keywords:** Picture naming, Aging, Bilingualism, Differential item functioning, Scoring, French

## INTRODUCTION

Bilingualism is extremely prevalent worldwide, with estimates putting the proportion of the world's population that is bilingual at around 50% (Grosjean, 1989, 2008). In Canada, almost 20% of the population speaks both English and French, with numbers rising above 40% in Quebec (Statistics Canada, 2011). Bilingualism is known to affect performance on neuropsychological tasks, particularly on tests of language performance including picture naming and verbal fluency (for a review, see Bialystok, 2009).

The Boston Naming Test (BNT) is commonly used in clinical practice (Kaplan, Goodglass, & Weintraub, 1983). The task is composed of 60 line drawings arranged in order of increasing difficulty; shortened versions with 15 or 30 items have also been developed (Lansing, Ivnik, Cullum, & Randolph, 1999; Mack, Freed, Williams, & Henderson, 1992).

The BNT has been translated into French (Colombo & Assal, 1992; Roberts & Doucet, 2011), but norms and scoring criteria have not been established (Roberts & Doucet, 2011).

Studies have indicated that relative to monolinguals, bilinguals perform worse on naming tasks such as the BNT, in both accuracy (Bialystok, Craik, & Luk, 2008; Kohnert, Hernandez, & Bates, 1998) and response time (Gollan, Fennema-Notestine, Montoya, & Jernigan, 2007; Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Roberts, Garcia, Desrochers, & Hernandez, 2002) even when tested in their dominant language (Gollan et al., 2005; Ivanova & Costa, 2008). On the BNT, English monolinguals have been found to outperform French–English bilinguals when completing the task in English, and item difficulty on the test was found to differ between groups (Roberts et al., 2002). Further research with Spanish–English bilinguals has found that using an either-language scoring method, where the total number of items correctly named in two separate single-language administrations, improves performance in a subset of participants (Kohnert et al., 1998). Gollan et al. (2007) found this improvement in balanced but not unbalanced Spanish–English bilinguals.

Correspondence and reprint requests to: Vanessa Taler, School of Psychology, University of Ottawa, 136 Jean Jacques Lussier, Vanier Hall, Ottawa, Ontario, Canada K1N 6N5. E-mail: vtaler@uottawa.ca

These findings highlight the difficulty for clinicians in assessing language abilities of bilingual individuals, where baseline data or appropriate norms to assist in interpreting performance are typically not available.

Previous research, where participants' language status was typically not controlled, has found that naming performance declines with age (Burke & Mackay, 1997), with an estimated decline of 2–3 items on the 60-item BNT between the fifth and eighth decade of life (Zec, Burkett, Markwell, & Larsen, 2007; Zec, Markwell, Burkett, & Larsen, 2005). Further research, however, has suggested that aging may be associated with improved ability to name pictures, especially for more difficult words (Gollan & Brown, 2006). Increased variability in older, relative to younger, adults may indicate better preservation of ability in some participants than others.

Overall, it is well established that a participant's neuropsychological performance can be affected by the language in which they are tested (e.g., Gollan et al., 2007; Kohnert et al., 1998) as well as their language background. However, to date there exists little research examining cross-language performance on picture naming tasks specifically in English–French bilinguals in Canada. The present study aims to fill this gap.

### Differential Item Functioning

Individual items on any psychometric test may show different functioning in different groups of participants and/or when the test is administered in different languages. This phenomenon is referred to as differential item functioning (DIF). Critically, DIF is identified *after matching for the underlying ability across participants* (Zumbo, 1999). For example, in the context of the current study, it would be expected that a randomly selected individual who speaks only English, and a randomly selected individual who speaks only French, would have a similar probability of correctly naming any BNT item, in the case where the two speakers had comparable naming abilities overall. For items where the probability of a correct response differs between groups, the item is said to display DIF. Uniform DIF occurs when the item favors one group over another across all levels of ability being measured. For example, more monolingual English speakers may correctly name an item on the BNT than monolingual French speakers, regardless of overall naming ability. Non-uniform DIF, in contrast, occurs when there is a significant group by ability interaction, suggesting that the probability of responding correctly on an item is not the same across ability levels for the two groups (Marshall, Mungas, Weldon, Reed, & Haan, 1997; Zumbo, 1999). For example, the probability of responding correctly to an item on the BNT may be higher in monolingual English speakers than in monolingual French speakers when overall naming ability is low, but the reverse may be true when naming abilities are high.

A variety of neuropsychological assessment tools, including the BNT, have been shown to display DIF with sex (Jones & Gallo, 2002), education (Jones & Gallo, 2002; Teresi, Kleinman, & Ocepek-Welikson, 2000), ethnicity

(Marshall et al., 1997; Pedraza et al., 2009), and language (Marshall et al., 1997). A recent examination of DIF in a 30-item version of the Boston Naming Test (Pedraza et al., 2009) found that when comparing samples of African American and Caucasian older adults, 12 items displayed DIF using item-response theory methods, and 14 items displayed DIF using logistic regression analyses. Critically, the presence of DIF in clinical tools poses a threat to the construct validity of the tool, because performance on the test is influenced by factors other than that being assessed. In a clinical setting, the presence of DIF may, therefore, result in erroneous decisions about cognitive functioning.

While DIF has been examined in the BNT for different ethnicities, to date no study has examined DIF on this task in monolingual and bilingual samples. The present study reports on the performance of younger and older English–French bilingual, English monolingual and French monolingual Canadians on the Boston Naming Test. We compare monolingual performance to bilingual performance in English, French, and an either-language administration and identify individual items on the BNT that exhibited DIF in the monolingual samples and in the English-only, French-only, and bilingual administrations in the bilingual participants.

The objectives of this study were to: (a) examine performance on the BNT in English monolingual, French monolingual and English–French bilingual younger and older adults; (b) examine bilingual performance on the BNT in English, French and an administration where they can respond in either language; and (c) identify items exhibiting differential functioning in the English and French monolingual participants, as well as in the monolingual and bilingual participants completing the test in English, French, and an either-language administration.

## METHODS

### Participants

Six groups of participants took part in this study: younger ( $n = 41$ ) and older ( $n = 31$ ) monolingual English speakers, younger ( $n = 30$ ) and older ( $n = 30$ ) monolingual French speakers; and younger ( $n = 47$ ) and older ( $n = 36$ ) bilingual speakers of both English and French. All participants had good self-reported health and no neurological or psychiatric history. Monolingual English and bilingual participants were recruited and tested in the Ottawa-Gatineau region, while monolingual French participants were recruited and tested in Quebec City. Across language groups, younger adults were recruited from local undergraduate populations, and older adults were recruited through word of mouth and advertisements in local community centers. The majority of participants across groups were right-handed (>90%) and language groups were matched for age and education (see Table 1).

Monolingual participants spoke only English or French and bilinguals spoke only French and English. All participants had minimal exposure to other languages. Twenty of 47 younger

**Table 1.** Demographic and neuropsychological performance by participant group (mean ± standard deviation)

	Younger adults						Older adults						Group differences <sup>b</sup>
	Monolingual English (n = 41; 15 males)		Monolingual French (n = 30; 10 males)		Bilingual (n = 47; 15 males)		Monolingual English (n = 31; 16 males)		Monolingual French (n = 30; 7 males)		Bilingual (n = 36; 19 males)		
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
Age (years)	21.66	1.62	21.80	2.47	21.57	2.33	72.26	6.43	72.60	6.59	70.69	5.86	YA < OA
Education (years)	15.56	1.14	15.13	1.38	15.68	1.49	15.62	2.87	16.20	2.57	16.14	2.85	N.S.
Coefficient of variability <sup>a</sup>													
English	0.22	0.10	—	—	0.26	0.13	0.25	0.10	—	—	0.26	0.13	N.S.
French	—	—	0.30	0.14	0.29	0.12	—	—	0.31	0.15	0.27	0.11	N.S.
MoCA (/30)	28.44	1.45	28.30	1.73	28.19	1.44	27.94	1.65	27.50	1.36	27.53	1.59	YA > OA
Digit Span													
Forwards (/16)	11.27	2.04	12.20	2.19	11.40	2.45	11.19	1.72	10.47	2.40	10.28	1.91	YA > OA
Reverse (/14)	6.88	2.05	8.40	2.66	7.83	2.56	7.87	2.05	7.07	2.03	7.06	2.19	YMF > YME; YMF > OMF; OBil
WCST (categories; /6)	4.27	1.03	4.70	0.47	4.58	0.92	3.50	1.22	4.07	0.69	3.50	1.21	YA > OA; MF > ME, Bil
Stroop													
Word	107.69	16.94	115.23	15.86	107.49	13.42	97.39	12.96	104.07	16.05	96.17	15.34	YA > OA; MF > ME, Bil
Color	76.97	14.30	86.77	14.37	76.29	10.56	64.87	11.73	73.10	14.07	59.94	12.57	YA > OA; MF > ME, Bil
Interference	52.00	11.95	48.67	8.26	53.29	9.52	34.77	8.16	32.40	9.55	36.72	8.57	YA > OA; Bil > MF
Verbal Fluency													
FAS	40.53	12.28	38.27	8.26	37.79	10.10	44.10	10.00	36.37	10.13	41.25	14.06	ME > MF
Animals	24.34	5.35	23.00	5.07	23.54	6.62	20.68	4.58	18.27	4.62	19.78	5.58	YA > OA
BNT (Lenient Scores)													
English	53.59	3.57	—	—	49.47	6.27	55.81	3.03	—	—	51.58	5.03	YA > OA; ME > Bil
French	—	—	50.27	4.25	38.98	8.84	—	—	47.87	6.04	46.11	5.45	YA > OA; MF > Bil; YMF > YBil; OBil > YBil

YA = younger adult; OA = older adult; ME = monolingual English; MF = monolingual French; Bil = bilingual; YMF = younger monolingual French; YME = younger monolingual English; YBil = younger bilingual; OME = older monolingual English; OMF = older monolingual French; OBil = older bilingual; MoCA = Montreal Cognitive Assessment; WCST = Wisconsin Card Sorting Test; BNT = Boston Naming Test

<sup>a</sup>Due to technical difficulty, data are missing for three younger and one older bilingual participants, one younger and one older monolingual French participants and two younger monolingual English participants.

<sup>b</sup>All group comparisons significant at  $p < 0.05$ .

**Table 2.** Language profile of younger and older bilingual participants

		Younger Adults	Older Adults
Language at home	Primarily English	48.95%	41.67%
	Primarily French	29.79%	30.56%
	Both equally	19.15%	27.78%
Language at work <sup>a</sup>	Primarily English	46.81%	13.89%
	Primarily French	6.38%	5.56%
	Both equally	44.68%	41.67%
Language of education	Primarily English	27.66%	16.67%
	Primarily French	31.91%	44.44%
	Both equally	40.43%	38.89%

<sup>a</sup>38.89% of older bilinguals were retired and did not note the primary language used at work.

(42.5%) and 25 of 36 older bilinguals (69.4%) reported French as their first language. All bilinguals had acquired a high degree of proficiency in both languages before age 13, and on average, younger adults acquired their second language at age 4, and older adults at age 6. All bilingual participants self-reported using both languages on a daily basis. The language profile by age group is provided in Table 2. All bilinguals provided a self-reported rating, on a 5-point Likert scale, of their English and French proficiency in the areas of auditory comprehension, reading, speaking and writing (1 = no ability and 5 = native-like ability; see Table 3). For the majority of bilinguals, the self-rated proficiency in English and French was equal and rated at “native-like ability” (i.e., 5) for auditory comprehension (77% of younger and 92% of older adults), reading (66% of younger and 81% of older adults), speaking (45% of younger and 83% of older adults) and writing (45% of younger and 70% of older adults).

### Measure of Bilingualism

In addition to self-report, English and French proficiency was assessed using an animacy judgment task that has been used to assess automaticity in participants' first and second languages (Segalowitz & Frenkiel-Fishman, 2005). In this task, participants must decide if stimuli are living (animate) or non-living (inanimate). Stimuli include 32 animate and 32 inanimate nouns in each language; no translation equivalents are included, and stimuli are presented in separate language blocks. Monolingual participants completed the task in their native language and bilingual participants completed first the English block and then the French block. For monolingual English and bilingual participants, this task was run using E-Prime software (Version 2.0) on a Dell laptop computer running Windows XP. For monolingual French participants, this task was run using E-Prime software (Version 2.0) on a Toshiba Portégé A600 laptop computer running Windows 7. We then calculated the coefficient of variability (CV) for each language administration by dividing the standard deviation of

response time for correct trials by the mean response time for correct trials in each language. This measure reflects cognitive efficiency due to reduced variability when processing is relatively more automatic, even when average response latencies are the same. The more similar the CVs in English and French, the more equally proficient the individual is assumed to be (Segalowitz & Segalowitz, 1993). Average CVs by group are provided in Table 1.

### Neuropsychological Battery

To assess general cognitive function, executive function and language abilities, all participants completed a neuropsychological battery that included the Montreal Cognitive Assessment (Nasreddine et al., 2005); the forward and backward digit span subtests of the Wechsler Adult Intelligence Scale-Third Edition (Wechsler, 1997); the 64-item Wisconsin Card Sorting Test (Grant & Berg, 1948); a version of the Stroop color-word interference test (Stroop, 1935) in which the number of items produced in 45s was recorded in each condition (word reading, color naming, and incongruent color naming); and category (animal) and letter (FAS) verbal fluencies (Benton & Hamsher, 1976). For monolingual participants, the battery was completed in their native language. For bilingual participants, the battery was completed in English, with the exception of the verbal fluency tasks, which were completed in English, French and in an administration where they could respond in either language. Scores from the English administration are presented in Table 1.

### Boston Naming Test

Monolingual participants completed the BNT (Kaplan et al., 1983) in their native language, while bilingual participants completed the test in three different administrations: English-only, French-only, and either language. In the either-language administration, bilingual participants were able to freely switch between French and English when naming each image (Gollan & Ferreira, 2009). All participants were asked to name all 60 items. The number of images spontaneously named correctly was calculated. We then compared strict and lenient scoring on the BNT, where lenient scoring included additional synonyms or alternate responses for select items. Accepted responses in English and French were based on the BNT scoring manual, and work by Roberts and Doucet (2011). Lenient and strict scoring criteria in each language for each item are provided in Appendix A.

### Procedure

Each monolingual participant completed testing in one session, while each bilingual participant completed testing in two sessions. For the verbal fluency and BNT, English, French and either-language administrations were completed in a different randomized order for each bilingual participant, wherein two administrations occurred in the first testing session (at the beginning and end of the session) and the final

**Table 3.** Mean rating and standard deviation for proficiency by modality for both English and French for bilingual younger ( $n = 47$ ) and bilingual older ( $n = 36$ ) adults

		Younger adults							Older adults								
		Full sample ( $n = 47$ )		English dominant ( $n = 21$ )		French dominant ( $n = 8$ )		Balanced ( $n = 18$ )		Full sample ( $n = 36$ )		English dominant ( $n = 10$ )		French dominant ( $n = 4$ )		Balanced ( $n = 22$ )	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
English	Listening	4.91	0.29	5.00	0.00	4.50	0.53	5.00	0.00	4.91	0.28	5.00	0.00	4.50	0.58	4.95	0.21
	Reading	4.89	0.39	4.95	0.22	4.50	0.76	5.00	0.00	4.94	0.24	5.00	0.00	4.75	0.50	4.95	0.21
	Speaking	4.89	0.32	5.00	0.00	4.25	0.46	5.00	0.00	4.86	0.36	5.00	0.00	4.00	0.00	4.95	0.21
	Writing	4.89	0.56	4.81	0.40	3.75	0.46	4.94	0.24	4.86	0.36	4.90	0.32	4.50	0.58	4.91	0.29
	Summed (/20) <sup>a</sup>	19.36	1.56	19.76	0.54	17.00	1.51	18.83	4.07	19.57	1.07	19.90	0.32	17.75	1.50	19.77	0.87
French	Listening	4.82	0.45	4.62	0.59	5.00	0.00	5.00	0.00	4.91	0.37	4.80	0.63	5.00	0.00	4.95	0.21
	Reading	4.64	0.57	4.24	0.62	5.00	0.00	5.00	0.00	4.80	0.41	4.40	0.52	5.00	0.00	4.95	0.21
	Speaking	4.39	0.69	3.81	0.51	4.75	0.46	5.00	0.00	4.91	0.28	4.80	0.42	5.00	0.00	4.95	0.21
	Writing	4.20	0.88	3.45	0.67	4.81	0.37	4.95	0.24	4.63	0.55	3.90	0.32	5.00	0.00	4.91	0.29
	Summed (/20) <sup>a</sup>	18.05	2.59	16.12	1.70	19.56	0.82	18.83	0.24	19.26	1.24	18.00	0.94	20.00	0.00	19.77	0.87

<sup>a</sup>Calculated by adding the self-rating between 1 and 5 for proficiency in each of the four modalities.

third administration occurred at the beginning of the second session. The study procedures adhered to federal guidelines for protection of human research participants and received ethical approval from the Research Ethics Board at the Bruyère Research Institute, Laval University, and the University of Ottawa. Participants signed an informed consent form before participation and were remunerated \$10/hr at the end of the study.

## Analyses

To identify items exhibiting differential functioning in the monolingual English and monolingual French participants, as well as in the monolingual and bilingual participants completing the test in English, French, and an either-language administration, we conducted a logistic regression analysis for each item in each comparison (monolingual-English vs. bilingual-English; monolingual-English vs. bilingual-either language; monolingual-French vs. bilingual-French; monolingual-French vs. bilingual-either language; and monolingual-English vs. monolingual-French). Compared to other methods commonly used to assess DIF (such as the Mantel-Haenzel or Simultaneous Item Bias Test procedures), logistic regression analyses have been shown to be as powerful at detecting uniform DIF and more powerful at detecting non-uniform DIF (Hidalgo & Lopez-Pina, 2004; Swaminathan & Rogers, 1990). The goal of the analysis is to identify items on the BNT that are not comparable in English and French in monolingual samples, as well as items that are not equivalent across monolingual and bilingual samples in English, French, and either-language administrations of the test.

Item response (pass/fail) was entered as the dependent variable, and condition (monolingual-English / monolingual-French/bilingual-English/bilingual-French/bilingual-either language) and total score were entered as independent variables.

Therefore, the logistic regression equation is:

$$Y = \beta_0 + \beta_1(\text{total}) + \beta_2(\text{condition}) + \beta_3(\text{total} * \text{condition})$$

To assess DIF, terms are successively added into the model to compare overall model fit. A common approach to identify items with DIF involves a simultaneous test of both uniform and non-uniform DIF using a two degrees of freedom chi-squared test; if a significant result is found, non-uniform DIF is identified when  $\beta_3 \neq 0$ , and uniform DIF is identified when  $\beta_2 \neq 0$  and  $\beta_3 = 0$  (Swaminathan & Rogers, 1990; Zumbo, 1999). However, simulation studies have shown that using separate one degree of freedom chi-squared tests for uniform and non-uniform DIF has higher power and reduces the rate of Type I errors, especially in smaller sample sizes (Jodoin & Gierl, 2001; Shimizu & Zumbo, 2005). Therefore, we used this approach, with BNT total score as the matching criterion. Before conducting the DIF analysis, a purification of the matching criterion was undertaken, where an initial logistic regression was conducted to identify items of potential DIF. These items were then removed and a new BNT total score was recalculated, and used as the matching criterion in subsequent DIF analyses (Zumbo, 1999).

To assess clinical relevance, effect sizes were calculated for each item to quantify the magnitude of uniform and non-uniform DIF. Following the recommendations outlined by Zumbo (1999),  $\Delta R^2$  values greater than 0.3 were classified as clinically relevant DIF.

## RESULTS

### Language Proficiency

To examine differences in self-rated proficiency in English and French, we summed the self-rating between 1 and 5 for proficiency in each modality for a total score out of 20

(see Table 3). Using a 2 (language) by 2 (age) analysis of variance (ANOVA), we found a main effect of language ( $F(1,80) = 9.208; p < .005; \eta_p^2 = 0.10$ ), such that proficiency in English was rated higher than proficiency in French ( $p < .01$ ) and a main effect of age ( $F(1,80) = 8.321; p < .01; \eta_p^2 = 0.094$ ), whereby older adults had higher self-rated proficiency than younger adults ( $p < .01$ ). A trend for a significant age\*language interaction ( $F(1,80) = 3.301; p = .071; \eta_p^2 = 0.040$ ) revealed that higher self-rated proficiency in English compared to French was seen in younger adults ( $p < .001$ ), but not older adults ( $p = .40$ ), and that older adults had a higher self-rated proficiency in French than did younger adults ( $p = .007$ ).

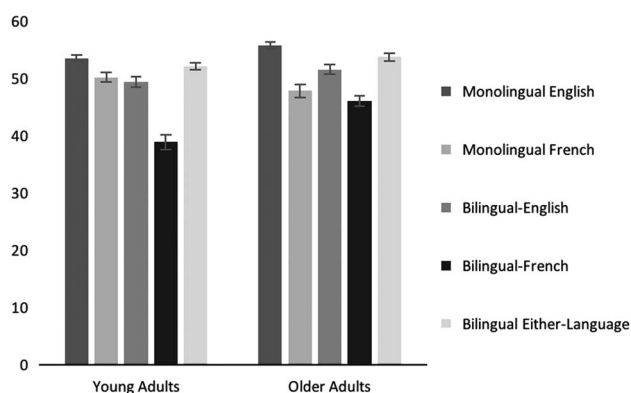
Table 1 presents the average coefficients of variability (CV) by group. A 2 (language) by 2 (age) ANOVA revealed that there was no significant effect of age ( $p = .752$ ), or language ( $p = .146$ ), indicating that both younger and older adults had similar CVs and that the CVs in French and English were not significantly different from one another.

## Boston Naming Test

Results by participant group for the BNT following lenient scoring are presented in Figure 1. Overall, for both younger and older adults following lenient scoring criteria, monolingual-English participants outperformed all other groups, and bilingual performance was lowest in the French administration. BNT performance following strict scoring criteria demonstrated similar patterns of results (data not shown). Given the similarities between lenient and strict scoring, all subsequent analyses are presented with the lenient scoring data. Results with the strict scoring data can be found in Appendix B.

### Analysis 1: Group effects

We used a 2 (age)  $\times$  3 (language group) ANOVA to examine the effects of age and language background on BNT performance, with lenient scores as the dependent variable. Each bilingual participant's highest score (from the three administrations) was entered into the analysis. Overall, no



**Fig. 1.** Average number ( $\pm$  standard error) of images named under lenient scoring criteria by age and language group.

main effect of age group was observed ( $p = .36$ ), but an effect of language group was observed ( $F(2,209) = 31.52; p < .001; \eta_p^2 = 0.23$ ), revealing that monolingual French participants had fewer correct responses than monolingual English and bilingual participants, both for younger ( $p < .01$ ) and older ( $p < .05$ ) adults. An age by language group interaction ( $F(2,209) = 5.87; p < .01; \eta_p^2 = 0.05$ ) demonstrated that older adults outperformed younger adults in the monolingual English ( $p < .05$ ) and bilingual groups ( $p = .057$ ), but the opposite effect was observed in the monolingual French participants, with younger adults outperforming older adults ( $p < .05$ ).

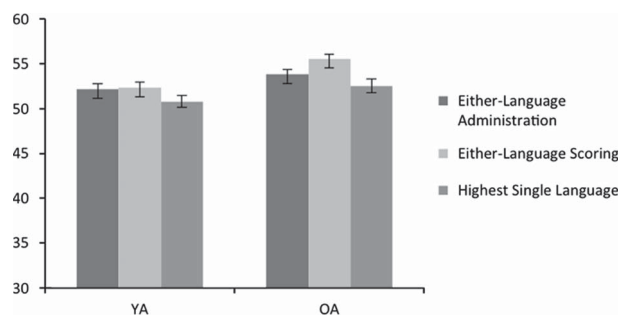
### Analysis 2: Effects of language of administration

Bilingual participants completed two single language administrations of the BNT (in English and in French) and an either-language administration, where they could respond in both English and French. We compared the effects of language of administration on the performance of bilingual participants using a repeated measures ANOVAs, with language of administration as the within-subjects factor and age as the between-subjects factor. Because assumptions of sphericity were not met, we applied the Greenhouse-Geisser adjustment; adjusted values are reported here. Overall, older adults outperformed younger adults ( $F(1,81) = 17.57; p < .001; \eta_p^2 = 0.18$ ), and bilingual participants performed better in the either-language administration than in the English-only administration, with lowest performance in the French-only administration ( $F(1.25, 101.08) = 78.10; p < .001; \eta_p^2 = 0.49$ ). An interaction between age and language of administration ( $F(1.25, 101.08) = 6.06; p < .02; \eta_p^2 = 0.07$ ) revealed that older adults outperformed younger adults in the French-only administration ( $p < .001$ ) but not in the English-only ( $p = .10$ ) or either-language administrations ( $p = .08$ ).

### Analysis 3: Effects of an either-language administration

Under the either-language administration, participants could respond to each item in the language of their choice. In younger adults, 66.6% of BNT items were named in English, while 26.3% of BNT items were named in French ( $p < .001$ ; participants were unable to name the remaining 7.1% of items). In older adults, 47.6% of items were named in English, while 45.2% of items were named in French ( $p = .814$ ; participants were unable to name the remaining 7.2% of items).

We wished to examine whether the either-language administration used in the present study produced similar findings to the either language scoring protocol used in previous research (Gollan et al., 2007). The either-language scoring protocol used in previous research calculated the bilingual advantage by adding the total number of items correctly named when the task was administered first in one language and then in the other (Gollan et al., 2007). In the present study, in contrast, we conducted an either-language



<sup>1</sup>Either-language administration: participants can respond to each item in whichever language they prefer

<sup>2</sup>Either-language scoring: either-language advantage calculated by adding the total number of items correctly named when the task was administered first in one language and then the other

<sup>3</sup>Highest single language score: highest score achieved in either English or French

**Fig. 2.** A comparison of performance in an either-language administration<sup>1</sup>, either-language scoring<sup>2</sup> and highest single language score<sup>3</sup> for both younger and older adults following lenient scoring (average number of images named  $\pm$  standard error).

administration in which participants were encouraged to provide their response to each item in whichever language they preferred, similar to what was done by Gollan and Ferreira (2009). Using separate paired samples *t* tests for younger and older adults, we found that an either-language administration yielded lower scores than an adding of total score in either language across two administrations in older adults ( $t(35) = -4.20$ ;  $p < .001$ ), but not younger adults ( $t(46) = -0.49$ ;  $p = .63$ ). Averages by condition are presented in Figure 2.

### Either-language administration advantage

To determine how many participants benefited from an either-language administration of the BNT, we compared performance in the either-language administration to the maximum score achieved in either English or French. We found that a subset of younger and older participants performed worse on the BNT when they could respond in either language, relative to their maximum score (see Table 4). As shown in Figure 3, in both younger and older bilinguals with an either-language advantage, there were significant differences between the score achieved in the either-language administration and the maximum single-language score ( $t(8.472) = 26$ ;  $p < .001$  for younger and  $t(6.625) = 22$ ;  $p = .001$  for older adults). The difference between these two scores was also significant in younger and older bilinguals with an either-language disadvantage ( $t(-4.583) = 9$ ;  $p = .001$  and  $t(-5.186) = 10$ ;  $p < .001$ , respectively). As shown in Table 4, across both age groups, a higher proportion of participants in each group had higher maximum scores in the English administration of the BNT as compared to the French.

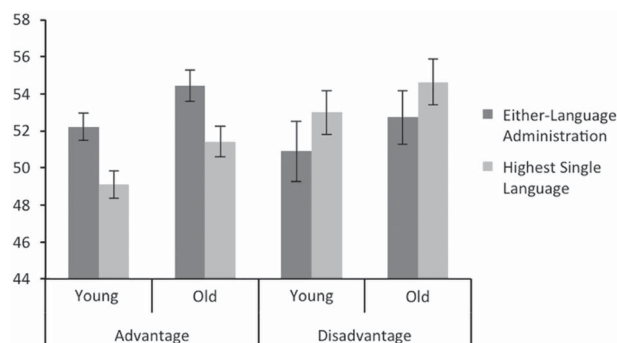
To examine differences in overall naming ability between participants with an either-language advantage or disadvantage, we used separate independent samples *t* tests for younger and older adults to compare the maximum score

**Table 4.** Cross-tabulations for the highest single-language BNT administration by bilingual advantage group in younger and older bilingual adults under lenient scoring

		Younger adults ( $n = 47$ )	Older adults ( $n = 36$ )
Advantage	English	$n = 21$	$n = 18$
	French	$n = 6$	$n = 4$
	Total advantage group:	$N = 27$	$N = 22$
Disadvantage	English	$n = 10$	$n = 8$
	French	$n = 0$	$n = 3$
	Total disadvantage group:	$N = 10$	$N = 11$
Equal	English	$n = 7$	$n = 2$
	French	$n = 3$	$n = 1$
	Total equal group:	$N = 10$	$N = 3$

achieved in either single language administration. We found that younger and older adults with an either-language advantage had lower maximum scores in either English or French ( $t(-2.335) = 35$ ;  $p = .025$  and  $t(-2.199) = 31$ ;  $p = .036$ , respectively).

We then conducted subsequent analyses in younger and older adults to compare demographic characteristics (age, education), language background (including first language, CV in English and French, and self-rated proficiency in English and French) and performance on the neuropsychological battery by either-language advantage status (i.e., benefited, did not benefit, and disadvantaged). In younger adults, we found that, compared to the advantaged group, the disadvantaged participants had higher CVs in French ( $p < .05$ ), and higher scores on the forward digit span ( $p < .01$ ). No other significant differences in demographic characteristics, language background or neuropsychological performance emerged. Similarly, for the older adults, no significant group differences were observed for language



<sup>1</sup>Either-language administration: participants can respond to each item in whatever language they prefer

<sup>2</sup>Highest single language: Highest score achieved in either English or French

**Fig. 3.** A comparison of the either-language administration<sup>1</sup> and the maximum score<sup>2</sup> achieved in either English or French by either-language advantage status for younger and older adults under lenient scoring (average number of images named  $\pm$  standard error).

background or neuropsychological performance, although the advantaged group was older than the disadvantaged group ( $p < .05$ ). Thus, we could not identify any meaningful demographic, language background, or neuropsychological differences between those participants who experienced an advantage *versus* those who did not.

*Analysis 4: Degree of bilingualism*

We were interested in examining whether the either-language advantage was related to degree of bilingualism (i.e., how balanced the bilingual is). Following Gollan, Weissberger, Runqvist, Montoya, and Cera (2012), degree of bilingualism was calculated using both subjective (i.e., the sum of the self-rating between 1 and 5 for proficiency in each modality for a total score out of 20 for English and French) and

objective (i.e., co-efficient of variability in English and French) measures of language proficiency, as well as with BNT performance in English and French. For all measures of bilingualism, we divided the lowest score by the highest score, meaning that values closer to one indicate a higher degree of balance. We then examined the correlations between the size of the either-language advantage and degree of bilingualism. For all three measures of balance, the results suggested that degree of bilingualism was not related to the size of the either-language advantage for younger or older adults under strict or lenient scoring ( $p > .5$  in all cases).

*Analysis 5: Item analysis*

To determine whether each item exhibited differential item functioning in the different conditions, we conducted a series of

**Table 5.** Items displaying uniform<sup>a</sup> and non-uniform<sup>b</sup> DIF ( $\Delta R^2$ ) for the lenient scoring of the BNT for younger and older participants

		Item				
		ME <sup>c</sup> -MF <sup>d</sup>	ME <sup>c</sup> -BE <sup>d</sup>	ME <sup>c</sup> -BEL <sup>d</sup>	MF <sup>c</sup> -BF <sup>d</sup>	MF <sup>c</sup> -BEL <sup>d</sup>
Younger adults	Uniform	Seahorse* (0.35)	Cactus* (0.93)	Cactus <sup>±</sup> (0.49)	Comb <sup>±</sup> (0.06)	Seahorse <sup>±</sup> (0.41)
		Cactus* (0.42)	Tripod* (0.10)	Escalator <sup>±</sup> (0.17)	Octopus <sup>±</sup> (0.04)	Globe* (0.24)
		Hammock* (0.15)	Protractor <sup>±</sup> (0.05)	Noose <sup>±</sup> (0.09)	Seahorse <sup>±</sup> (0.10)	Knocker* (0.40)
		Knocker <sup>±</sup> (0.40)			Globe* (0.17)	Compass <sup>±</sup> (0.11)
		Accordion* (0.10)			Stilts <sup>±</sup> (0.12)	Tongs <sup>±</sup> (0.35)
		Asparagus* (0.23)			Harp* (0.06)	Yoke* (0.15)
		Compass* (0.21)			Stethoscope* (0.07)	Trellis* (0.10)
		Tripod* (0.09)			Noose <sup>±</sup> (0.10)	
		Tongs* (0.34)			Latch <sup>±</sup> (0.10)	
		Sphinx* (0.11)			Tripod* (0.05)	
	Non-uniform	Protractor <sup>±</sup> (0.03)			Trellis* (0.08)	
		Wreath <sup>‡</sup> (0.20)	Stilts <sup>‡</sup> (0.08)	Stilts <sup>‡</sup> (0.13)	Dart <sup>‡</sup> (0.05)	Dominos <sup>°</sup> (0.10)
		Harmonica <sup>‡</sup> (0.18)	Hammock <sup>°</sup> (0.09)	Protractor <sup>°</sup> (0.10)	Harp <sup>‡</sup> (0.14)	
		Hammock <sup>‡</sup> (0.21)	Stethoscope <sup>‡</sup> (0.09)		Sphinx <sup>‡</sup> (0.12)	
		Accordion <sup>‡</sup> (0.08)	Accordion <sup>‡</sup> (0.13)		Trellis <sup>‡</sup> (0.14)	
		Protractor <sup>°</sup> (0.11)	Asparagus <sup>‡</sup> (0.14)		Palette <sup>‡</sup> (0.10)	
			Protractor <sup>‡</sup> (0.14)		Protractor <sup>‡</sup> (0.06)	
Older adults	Uniform	Volcano* (0.08)	Pretzel* (0.11)	Dart <sup>±</sup> (0.22)	Octopus* (0.23)	Snail <sup>±</sup> (0.16)
		Stilts <sup>±</sup> (0.08)	Wreath* (0.23)	Escalator <sup>±</sup> (0.24)	Mushroom <sup>‡</sup> (0.22)	Harmonica <sup>±</sup> (0.19)
		Knocker <sup>±</sup> (0.11)	Harmonica <sup>±</sup> (0.27)	Funnel <sup>±</sup> (0.28)	Snail <sup>±</sup> (0.16)	Knocker* (0.16)
		Pelican <sup>±</sup> (0.12)	Funnel <sup>±</sup> (0.14)	Asparagus* (0.15)	Rhinoceros* (0.13)	
		Tripod* (0.02)	Compass* (0.12)	Compass* (0.09)	Dominos* (0.22)	
		Scroll <sup>±</sup> (0.10)	Scroll <sup>±</sup> (0.11)	Tripod* (0.21)	Cactus* (0.19)	
		Tongs* (0.25)		Scroll <sup>±</sup> (0.13)	Unicorn* (0.33)	
		Yoke <sup>±</sup> (0.12)			Funnel <sup>‡</sup> (0.11)	
		Trellis <sup>±</sup> (0.03)			Sphinx* (0.20)	
					Trellis* (0.21)	
	Non-uniform	Volcano <sup>°</sup> (0.23)	Pretzel <sup>°</sup> (0.14)	Noose <sup>‡</sup> (0.12)	Bench <sup>°</sup> (0.28)	N/A
		Latch <sup>‡</sup> (0.09)		Compass <sup>‡</sup> (0.04)	Compass <sup>‡</sup> (0.14)	
		Tripod <sup>‡</sup> (0.09)		Scroll <sup>°</sup> (0.03)	Sphinx <sup>‡</sup> (0.08)	

ME = monolingual-English; MF = monolingual-French; BE = bilingual-English administration; BF = bilingual-French administration; BEL = bilingual-either language administration.

<sup>a</sup>Uniform DIF occurs when the item favours one group over another across all levels of naming ability.

<sup>b</sup>Non-uniform DIF occurs when the probability of a correct response varies by naming ability.

<sup>c</sup>Focal Group.

<sup>d</sup>Reference Group.

<sup>±</sup>Item favours focal group over reference group.

<sup>\*</sup>Item favours reference group over focal group.

<sup>‡</sup>For individuals with low naming ability, item favors reference group; for individuals with high naming abilities, item favors focal group.

<sup>°</sup>For individuals with low naming ability, item favors focal group; for individuals with high naming abilities, item favors reference group.



**Table 6.** Items displaying uniform<sup>a</sup> and non-uniform<sup>b</sup> DIF for the strict scoring of the BNT for younger and older participants

		Item				
		ME <sup>c</sup> -MF <sup>d</sup>	ME <sup>c</sup> -BE <sup>d</sup>	ME <sup>c</sup> -BEL <sup>d</sup>	MF <sup>c</sup> -BF <sup>d</sup>	MF <sup>c</sup> -BEL <sup>d</sup>
Younger adults	Uniform	Seahorse* (0.32)	Cactus* (0.63)	Cactus* (0.53)	Seahorse <sup>±</sup> (0.12)	Mask <sup>±</sup> (0.60)
		Dart <sup>±</sup> (0.14)	Tripod* (0.13)	Escalator <sup>±</sup> (0.21)	Globe* (0.14)	Seahorse <sup>±</sup> (0.47)
		Escalator <sup>±</sup> (0.29)	Protractor <sup>±</sup> (0.06)	Noose <sup>±</sup> (0.10)	Stilts <sup>±</sup> (0.13)	Dart* (0.19)
		Knocker <sup>±</sup> (0.50)			Muzzle <sup>±</sup> (0.03)	Globe* (0.19)
		Asparagus* (0.26)			Unicorn <sup>±</sup> (0.14)	Knocker* (0.36)
		Compass* (0.22)			Noose <sup>±</sup> (0.11)	Compass <sup>±</sup> (0.12)
		Tripod* (0.07)			Tripod* (0.05)	Tongs <sup>±</sup> (0.34)
	Tongs* (0.33)					
	Sphinx* (0.10)					
	Non-uniform	Hammock <sup>±</sup> (0.21)	Stilts <sup>‡</sup> (0.10)	Octopus <sup>‡</sup> (0.17)	Harp <sup>°</sup> (0.10)	Hammock <sup>°</sup> (0.31)
		Pelican <sup>‡</sup> (0.09)	Hammock <sup>‡</sup> (0.09)	Stilts <sup>‡</sup> (0.11)	Sphinx <sup>‡</sup> (0.09)	Compass <sup>‡</sup> (0.08)
		Accordion <sup>‡</sup> (0.09)	Pelican <sup>‡</sup> (0.11)	Hammock <sup>‡</sup> (0.11)	Trellis <sup>°</sup> (0.15)	
		Asparagus <sup>‡</sup> (0.21)	Accordion <sup>‡</sup> (0.11)	Knocker <sup>°</sup> (0.06)	Palette <sup>‡</sup> (0.09)	
		Protractor <sup>°</sup> (0.12)	Asparagus <sup>‡</sup> (0.18)	Protractor <sup>°</sup> (0.12)	Protractor <sup>‡</sup> (0.07)	
Abacus <sup>‡</sup> (0.12)		Protractor <sup>°</sup> (0.13)				
Older adults	Uniform	Dart <sup>±</sup> (0.23)	Pretzel* (0.12)	Mushroom* (0.62)	Hanger <sup>±</sup> (0.27)	Camel <sup>±</sup> (0.36)
		Acorn* (0.09)	Wreath* (0.19)	Snail <sup>±</sup> (0.13)	Wheelchair <sup>±</sup> (0.20)	Snail <sup>±</sup> (0.15)
		Escalator <sup>±</sup> (0.35)	Funnel <sup>‡</sup> (0.11)	Dart <sup>±</sup> (0.20)	Camel <sup>±</sup> (0.17)	Dart* (0.17)
		Knocker <sup>±</sup> (0.06)	Tripod* (0.08)	Escalator <sup>±</sup> (0.25)	Snail <sup>±</sup> (0.32)	Harmonica <sup>±</sup> (0.32)
		Pelican <sup>±</sup> (0.13)	Sphinx* (0.06)	Funnel <sup>‡</sup> (0.29)	Harmonica <sup>±</sup> (0.39)	Knocker* (0.23)
		Accordion* (0.15)		Asparagus* (0.14)	Igloo* (0.27)	Noose* (0.12)
		Asparagus* (0.17)		Compass* (0.09)	Dominos* (0.21)	
	Compass* (0.08)		Tripod* (0.22)	Cactus* (0.19)		
	Tongs* (0.16)			Funnel <sup>±</sup> (0.15)		
				Sphinx* (0.20)		
				Trellis <sup>±</sup> (0.24)		
	Non-uniform	Muzzle <sup>°</sup> (0.13)	Stethoscope <sup>‡</sup> (0.15)	Noose <sup>‡</sup> (0.11)	Volcano <sup>‡</sup> (0.14)	N/A
			Unicorn <sup>‡</sup> (0.13)		Unicorn <sup>‡</sup> (0.13)	
					Scroll <sup>‡</sup> (0.07)	

ME = monolingual-English; MF = monolingual-French; BE = bilingual-English administration; BF = bilingual-French administration; BEL = bilingual-either language administration.

<sup>a</sup>Uniform DIF occurs when the item favours one group over another across all levels of naming ability.

<sup>b</sup>Non-uniform DIF occurs when the probability of a correct response varies by naming ability.

<sup>c</sup>Focal Group.

<sup>d</sup>Reference Group.

<sup>±</sup>Item favours focal group over reference group.

<sup>\*</sup>Item favours reference group over focal group.

<sup>‡</sup>For individuals with low naming ability, item favors reference group; for individuals with high naming abilities, item favors focal group.

<sup>°</sup>For individuals with low naming ability, item favors focal group; for individuals with high naming abilities, item favors reference group.

logistic regression analyses for each of the 60 items (see methods section for a description of the analyses). Because item difficulty may vary by age, we analyzed older and younger adults separately. The items displaying DIF varied between scoring criteria; therefore, results for lenient and strict scoring criteria are presented in Tables 5 and 6, respectively. Across comparisons for both lenient and strict scoring, a large number of items displayed DIF, with effect sizes ranging from negligible to large (i.e., over 0.3). For both younger and older adults, it appears that the largest number of DIF items appear when comparing monolingual French participants to monolingual English participants and to the bilingual participants completing the test in French.

## DISCUSSION

The purpose of the present study was to: (a) examine the performance of younger and older bilingual speakers of English

and French on the Boston Naming Test to that of age-matched monolingual English and French speakers; (b) compare bilingual participants' performance in each language and in an either-language administration, in which they could provide responses in the language of their choice; and (c) identify items displaying DIF by language group and/or language of administration.

Overall, monolingual English speakers outperformed bilinguals, who outperformed monolingual French speakers. Higher performance was observed in older than younger adults for the monolingual English and bilingual participants, while the opposite trend was observed in the monolingual French participants. In bilinguals, the French administration yielded the lowest scores; this finding, coupled with the finding that monolingual French participants had the poorest performance on the BNT relative to other groups, provides support for the notion that the BNT is not equivalent in

English and French. The either-language administration improved performance in both younger and older bilingual adults, suggesting that for many bilinguals, vocabulary is not evenly divided across languages. For example, participants may know certain words in the language of education (e.g., “protractor”), and other items in the language of the home (e.g., “wreath”).

Importantly, not all bilingual participants benefitted from an either-language scoring protocol, and in fact a subset of participants achieved lower scores in the bilingual administration relative to their highest single-language score. The benefits of an either-language protocol appear to be related to the participants’ naming abilities; participants who experienced an advantage with an either-language protocol had lower maximum BNT scores (in either English or French) than participants who experienced a disadvantage.

Additionally, correlation analyses suggested that the either-language advantage was not related to the degree of balanced bilingualism. This finding is inconsistent with previous research indicating an advantage only in balanced bilinguals (Gollan et al., 2007). However, the study by Gollan et al. used an alternate method to determine either-language advantage, whereby credit was given for a correct response in either single-language administration. Thus the findings are not directly comparable.

Preliminary logistic regression analysis indicated items with DIF by language group and/or language of administration. Findings suggested that after matching for underlying naming ability, several items are not equivalent in French or for our bilingual sample. The large number of DIF items specifically identified when comparing the monolingual English and French participants suggested that the test is not equivalent in English and French. Clinicians using the BNT in neuropsychological testing with English–French bilinguals may wish to exercise caution in interpreting performance on the items displaying large amounts of DIF (that is, cactus, seahorse, knocker, tongs, mask, hammock, escalator, mushroom, snail, camel, and harmonica, all of which had an effect size of greater than 0.3), or when terminating testing when the eight items unsuccessfully named include these items. There are several items that did not show DIF between groups or language of administration. Future research should investigate the validity of these items in participants from diverse geographical regions and educational levels.

Given that bilingual participants completed the BNT twice in one testing session and a third time in an additional testing session, it is possible that testing order may have impacted our results. However, when we compared BNT performance in administrations one, two, and three, we found no significant differences between scores, suggesting that order of administration is not driving the findings presented in this research.

The research reported here includes a preliminary sample of participants, and future research should attempt to replicate these results with a larger sample. Moreover, our sample included primarily balanced bilingual speakers of English and French, although there was some evidence to suggest

that our sample of older bilinguals had a stronger knowledge of French than the younger adults. Thus, these results may not be generalizable to unbalanced bilingual speakers or bilinguals whose languages are not French and English. However, we note that our monolingual and bilingual samples were closely matched for age, and were largely homogeneous in terms of cultural background (over 90% of participants in each age and language group were born in Canada, with the remainder born in the United States or the United Kingdom), as well as educational and professional background. Thus, the present findings are likely due to participants’ language background rather than other socio-economic or educational factors. However, the opposing age effects in the language groups leave open the possibility that performance differences may be related to other demographic or language proficiency differences across the language groups.

We also note that differences in performance on the French and English versions of the BNT may result from characteristics of the items themselves, such as frequency differences across languages, or orthographic differences. For example, several single-word items in English, such as “noose,” translate as multi-word items in French (“nœud coulant”). Although we did not find evidence for performance differences based on these factors in our data, we cannot definitively address these issues based on the current data. Moreover, age of acquisition effects may vary by age group, and this factor may play a role in the observed age differences. Future research should further explore these possibilities.

## CONCLUSIONS

The present study suggests that the BNT is not of equivalent difficulty in French and English, and indicates the importance of taking language background into account when interpreting performance on naming tasks in bilingual adults, as has been observed in previous studies (for a review and discussion, see Rivera Mindt et al., 2008). We identify items that display uniform and non-uniform DIF by language or condition (listed in Tables 5 and 6), and recommend caution when interpreting an English–French bilingual’s inability to name these items. Future research should be directed at developing and norming a new naming task that is appropriate for assessing language function in English–French bilingual speakers.

Not all bilingual participants benefitted from being allowed to provide responses in either language, and in fact a subset of participants performed worse in the either-language administration than in their highest scoring single language administration. Future research with a larger sample of bilinguals experiencing a disadvantage in an either-language administration protocol is needed to identify factors associated with this finding.

While the present study has focused on BNT performance in a specific and well-defined population, it is our hope that these findings may serve as a model for the general effect of bilingualism on neuropsychological test performance, particularly when the test(s) involve primarily verbal input and/or output.

## ACKNOWLEDGMENTS

This research was supported by a Catalyst grant from the Canadian Institutes of Health Research awarded to Vanessa Taler and Shanna Kousaie (Grant # 112241), an Alzheimer Society of Canada Research Grant awarded to Vanessa Taler, Laura Monetta, and Shanna Kousaie (Grant #1423), and a Natural Sciences and Engineering Research Council of Canada Discovery Grant awarded to Vanessa Taler (Grant #386467-2012). The authors declare no conflict of interest. We would like to thank Julien Blacklock, Chloe Corbeil, Dominique Fijal, and Maude Lemieux for their assistance with data collection.

## REFERENCES

- Benton, A.L., & Hamsher, K. (1976). *Multilingual Aphasia Examination Manual*. University of Iowa, Iowa City: AJA Associates.
- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, 12, 3–11.
- Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 34, 859–873.
- Burke, D.M., & Mackay, D.G. (1997). Memory, language and ageing. *Philosophical Transactions of the Royal Society: Biological Sciences*, 352, 1845–1856.
- Colombo, F.T., & Assal, G. (1992). Boston Naming Test - French language adaptation and short forms. *European Review of Applied Psychology*, 42, 67–73.
- Gollan, T.H., & Brown, A.S. (2006). From Tip-of-the-Tongue (TOT) data to theoretical implications of two steps: When more TOTs mean better retrieval. *Journal of Experimental Psychology: General*, 135(3), 462–483.
- Gollan, T.H., Fennema-Notestine, C., Montoya, R.I., & Jernigan, T.L. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society*, 13, 197–208.
- Gollan, T.H., & Ferreira, V.S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology. Learning Memory and Cognition*, 35(3), 640–665. doi:10.1037/a0014981
- Gollan, T.H., Montoya, R.I., Fennema-Notestine, C., & Morris, S.K. (2005). Bilingualism affects picture naming but not picture classification. *Memory and Cognition*, 33, 1220–1234.
- Gollan, T.H., Weissberger, G.H., 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, 592–615.
- Grant, D.A., & Berg, E.A. (1948). A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology*, 38, 404–411.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, 3–15.
- Grosjean, F. (2008). *Studying Bilinguals*. Oxford: Oxford University Press.
- Hidalgo, M.D., & Lopez-Pina, J.A. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement*, 64(6), 903–915. doi:10.1177/0013164403261769
- Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production. *Acta Psychologica*, 127, 277–288.
- Jodoin, M.G., & Gierl, M.J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement Education*, 14, 329–349.
- Jones, R.N., & Gallo, J.J. (2002). Education and sex differences in the mini-mental state examination: Effects of differential item functioning. *The Journal of Gerontology, Series B: Psychological Sciences and Social Sciences*, 57B(6), 548–558.
- Kaplan, E.F., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test*. Philadelphia, PA: Lea & Febiger.
- Kohnert, K.J., Hernandez, A.E., & Bates, E. (1998). Bilingual performance on the Boston Naming Test: Preliminary norms in Spanish and English. *Brain and Language*, 65, 422–440.
- Lansing, A.E., Ivnik, R.J., Cullum, C.M., & Randolph, C. (1999). An empirically derived short form of the Boston Naming Test. *Archives of Clinical Neuropsychology*, 14, 481–487.
- Mack, W.J., Freed, D.M., Williams, B.W., & Henderson, V.W. (1992). Boston Naming Test: Shortened versions for use in Alzheimer's disease. *Journal of Gerontology*, 47, P154–P158.
- Marshall, S.C., Mungas, D., Weldon, M., Reed, B., & Haan, M. (1997). Differential item functioning in the mini-mental state examination of English- and Spanish-speaking older adults. *Psychology and Aging*, 12, 718–725.
- Nasreddine, Z.S., Phillips, N.A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53, 695–699.
- Pedraza, O., Graff-Radford, N.R., Smith, G.E., Ivnik, R.J., Willis, F. B., Petersen, R.C., & Lucas, J.A. (2009). Differential item functioning of the Boston Naming Test in cognitively normal African American and Caucasian older adults. *Journal of International Neuropsychological Society*, 15(5), 758–768.
- Rivera Mindt, M., Arentoft, A., Kubo Germano, K., D'Aquila, E., Scheiner, D., Pizzirusso, M., ... Gollan, T.H. (2008). Neuropsychological, cognitive, and theoretical considerations for evaluation of bilingual individuals. *Neuropsychology Review*, 18, 255–268.
- 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(3), 254–267.
- Roberts, P.M., Garcia, L.J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston Naming Test. *Aphasiology*, 16, 635–645.
- Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & Cognition*, 33, 644–653.
- Segalowitz, N., & Segalowitz, S.J. (1993). Skilled performance, practice and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, 14, 369–385.
- Shimizu, Y., & Zumbo, B.D. (2005). A logistic regression for differential item functioning primer. *Japan Language Testing Association Journal*, 7, 110–124.
- Statistics Canada (2011). Population by knowledge of official language, by province and territory (2006 Census). Retrieved from <http://www40.statcan.ca/01/cst01/DEMO15-eng.htm>
- Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.

- Swaminathan, H., & Rogers, H.J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361–370.
- Teresi, J., Kleinman, M., & Ocepek-Welikson, K. (2000). Modern psychometric methods for detection of differential item functioning: Application to cognitive assessment measures. *Statistics in Medicine*, 19, 1651–1683.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale - Third Edition*. San Antonio, TX: The Psychological Corporation.
- Zec, R.F., Burkett, N.R., Markwell, S.J., & Larsen, D.L. (2007). A cross-sectional study of the effects of age, education, and gender on the Boston Naming Test. *The Clinical Neuropsychologist*, 21, 587–616.
- 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.
- Zumbo, B.D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modelling as a unitary framework for binary and likert type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

## APPENDIX A

### Strict and Lenient Scoring Criterion for the BNT in English and French

BNT Item	English		French	
	Correct Response (Strict Scoring)	Alternate Response (Lenient Scoring)	Correct Response (Strict Scoring)	Alternate Response (Lenient Scoring)
1	Bed		Lit	
2	Tree		Arbre	
3	Pencil		Crayon	
4	House	Home	Maison	
5	Whistle		Sifflet	
6	Scissors	Shears	Ciseaux	
7	Comb		Peigne	
8	Flower		Fleur	
9	Saw	Hand Saw	Scie	Egoïne
10	Toothbrush		Brosse à dents	
11	Helicopter	Chopper	Hélicoptère	
12	Broom		Balais	
13	Octopus	Squid	Pieuvre	
14	Mushroom	Toadstool	Champignon	
15	Hanger	Clothes hanger	Cintre	Support
16	Wheelchair		Chaise roulante/d'invalide	
17	Camel	Dromedary	Chameau	Dromadaire
18	Mask	False-Face, Halloween Mask	Masque	
19	Pretzel		Bretzel	Pretzel
20	Bench	Park Bench	Banc	
21	Racquet		Raquette	
22	Snail		Escargot	Calimaçon
23	Volcano		Volcan	
24	Seahorse	Horsefish	Hippocampe	Cheval de mer
25	Dart		Dard	Fléchette
26	Canoe		Canot	Chaloupe
27	Globe		Globe terrestre	Mappe monde
28	Wreath		Couronne	
29	Beaver		Castor	
30	Harmonica	Mouth Organ	Harmonica	Musique à bouche
31	Rhinoceros	Rhino	Rhinocéros	Rhino
32	Acorn		Gland	
33	Igloo		Igloo	
34	Stilts		Echasse	
35	Dominos		Dominos	
36	Cactus		Cactus	

## APPENDIX A: (Continued)

BNT Item	English		French	
	Correct Response (Strict Scoring)	Alternate Response (Lenient Scoring)	Correct Response (Strict Scoring)	Alternate Response (Lenient Scoring)
37	Escalator		Escalateur	Escalier Roulant
38	Harp		Harpe	
39	Hammock		Hamac	
40	Knocker		Heutoir	Marteau de porte
41	Pelican		Pélican	
42	Stethoscope		Stéthoscope	
43	Pyramid		Pyramide	
44	Muzzle		Muselière	
45	Unicorn		Licorne	Unicorne
46	Funnel		Entonnoir	
47	Accordion	Squeeze Box, Concertina	Accordéon	
48	Noose		Nœud Coulant	Corde de pendaison
49	Asparagus		Asperge	
50	Compass		Compas	
51	Latch, Bolt		Clenche	Barrure
52	Tripod		Trépied	
53	Scroll		Parchemin	
54	Tongs		Pinces	
55	Sphinx		Sphinx	
56	Yoke	Oxbow	Joug	
57	Trellis		Trellis	
58	Palette		Palette	
59	Protractor		Rapporteur (d'angles)	
60	Abacus		Boulier compteur	Boulier, Abaque

## APPENDIX B

## Results Using the Strict Scoring Criteria of the BNT

*Analysis 1: Group effects*

Overall, no main effect of age group was observed ( $p = .49$ ), but an effect of language group was seen ( $F(2,209) = 35.49$ ;  $p < .001$ ;  $\eta_p^2 = 0.25$ ), whereby monolingual French participants performed more poorly than monolingual English and bilingual participants. An interaction between age and language group ( $F(2,209) = 5.17$ ;  $p < .01$ ;  $\eta_p^2 = 0.05$ ) revealed that younger adults outperformed older adults in the monolingual French group ( $p < .05$  for both) but the opposite trend was revealed for monolingual English ( $p = .05$ ) and bilingual groups ( $p = .09$ ).

*Analysis 2: Effects of language administration*

Results revealed an effect of age ( $F(1,81) = 13.16$ ;  $p < .001$ ;  $\eta_p^2 = 0.14$ ), where older adults outperformed younger adults ( $p < .001$ ), and an effect of language of administration ( $F(1.23,99.75) = 95.60$ ;  $p < .001$ ;  $\eta_p^2 = 0.54$ ), where bilingual participants performed highest in the either-language administration, followed by the English-only administration, with lowest performance seen in the French-only administration ( $p < .001$  in all cases). A language by age

interaction ( $F(1.23,99.76) = 4.97$ ;  $p < .05$ ;  $\eta_p^2 = .06$ ) demonstrated that older adults outperformed younger adults in the French administration ( $p < .001$ ) but not the English ( $p = .121$ ) or the either-language ( $p = .165$ ) administrations.

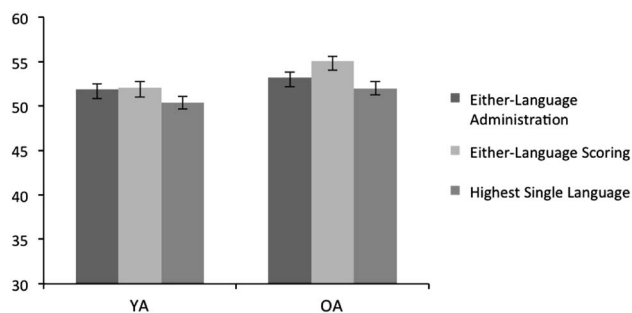
*Analysis 3: Effects of an either-language administration*

Using separate paired samples *t*-tests for younger and older adults, we found that an either-language administration (where participants could respond in both English and French) yielded lower scores than an adding of total score in either language across two administrations (see Gollan et al., 2007) in older ( $t(35) = -4.28$ ;  $p < .001$ ) but not younger ( $t(46) = -0.55$ ;  $p = .58$ ) adults. Averages by condition are presented in Figure 1.

*Either-language administration advantage*

Relative to their highest scoring single language administration, 27 younger and 23 older adults displayed an advantage in the either-language administration, while 8 younger and 11 older adults experienced a disadvantage. The remaining 12 younger and two older bilingual participants had equal performance in the either-language administration relative to their highest scoring single language administration. As shown in Figure 2, younger and older bilinguals with an

either-language advantage scored significantly higher in the either-language administration compared to their higher scoring single-language administration ( $t(8.989) = 25$ ;  $p < .001$  and  $t(6.625) = 22$ ;  $p < .001$ , respectively), while younger and older bilinguals with an either-language disadvantage scored significantly worse ( $t(-4.112) = 8$ ,  $p = .003$  and  $t(-6.197) = 10$ ,  $p < .001$ , respectively). Separate independent samples  $t$ -tests for younger and older adults revealed that in younger adults ( $t(-2.906) = 0.33$ ;  $p < .01$ ) but not older adults ( $t(-1.965) = 31$ ;  $p = .058$ ),

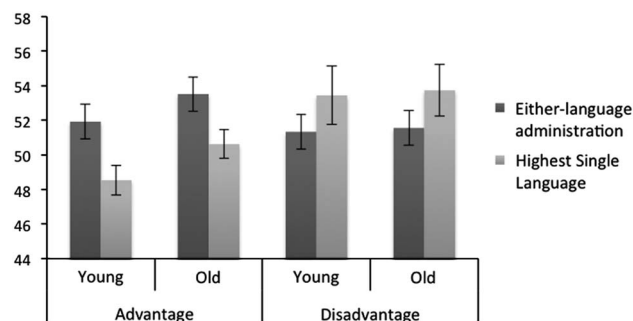


**Figure A1.** A comparison of performance in an either-language administration, either language scoring (an adding of total score in either language across two administrations) and highest single language score for both younger and older adults following strict scoring (average number of images named  $\pm$  standard error).

bilinguals with an either-language advantage had lower maximum scores in either English or French compared to those with an either-language disadvantage.

#### Analysis 4: Degree of bilingualism

Self-rated proficiency index, co-efficient of variability index, and BNT index were not correlated with the size of the either-language advantage in younger or older adults ( $p > .6$  in all cases).



**Figure A2.** A comparison of the either-language administration and the maximum score achieved in either English or French by either-language advantage status for younger and older adults under strict scoring (average number of images named  $\pm$  standard error).