

# Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals\*

TAMAR H. GOLLAN

*Department of Psychiatry, University of California San Diego*

GALI H. WEISSBERGER

*Department of Psychiatry, University of California San Diego & San Diego State University*

ELIN RUNNQVIST

*Universitat de Barcelona, Spain*

ROSA I. MONTOYA

*Department of Psychiatry, University of California San Diego*

CYNTHIA M. CERA

*Department of Psychiatry, University of California San Diego*

(Received: September 29, 2010; final revision received: May 10, 2011; accepted: May 13, 2011; first published online 1 August 2011)

*This study investigated correspondence between different measures of bilingual language proficiency contrasting self-report, proficiency interview, and picture naming skills. Fifty-two young (Experiment 1) and 20 aging (Experiment 2) Spanish–English bilinguals provided self-ratings of proficiency level, were interviewed for spoken proficiency, and named pictures in a Multilingual Naming Test (MINT); in Experiment 1, the Boston Naming Test (BNT) was also used. Self-ratings, proficiency interview, and the MINT did not differ significantly in classifying bilinguals into language-dominance groups, but naming tests (especially the BNT) classified bilinguals as more English-dominant than other measures. Strong correlations were observed between measures of proficiency in each language and language-dominance, but not degree of balanced bilingualism (index scores). Depending on the measure, up to 60% of bilinguals scored best in their self-reported non-dominant language. The BNT distorted bilingual assessment by underestimating ability in Spanish. These results illustrate what self-ratings can and cannot provide, illustrate the pitfalls of testing bilinguals with measures designed for monolinguals, and invite a multi-measure goal-driven approach to classifying bilinguals into dominance groups.*

Keywords: proficiency, aging, picture naming test, language dominance, self-ratings

Research on bilingualism has in recent years accelerated at “a dizzying pace” (Kroll & de Groot, 2005). Yet, despite the now thousands of studies, there is still no standard method for determining language proficiency, degree of bilingualism, and language dominance. Uniformity in how language dominance is assessed is tremendously important for advancing knowledge about the effects of bilingualism on language processing and cognition, and for interpretation of outcomes observed in experimental studies, and in clinical settings. Some effects obtained will apply only to some types of bilinguals (e.g., the cognitive advantages of bilingualism may be observed only in highly proficient bilinguals), but without a system for classifying

bilinguals into types it will be impossible to identify precisely which aspect of bilingualism is critical in each case. A standard method for determining proficiency and dominance across multiple types of bilinguals would go a long way towards clarifying the associated theoretical implications.

One of the most broadly used approaches to assessing bilingual language proficiency are self-ratings (Li, Sepanski & Zhao, 2006). Bilinguals are often asked to rate their abilities in each language, and multiple studies have shown that self-ratings are significantly correlated with objectively measured proficiency on a broad variety of measures (e.g., in one study, significant correlations were reported between self-ratings and reading fluency, reading comprehension, picture naming, auditory comprehension, sound awareness, receptive vocabulary, and grammaticality judgment speed and accuracy; see Marian, Blumenfeld & Kaushanskaya, 2007). These correlations are often highly robust (significant at the  $p < .01$  level), and can also be moderate or large in size (especially for ratings of a non-dominant

\* The authors thank Jeanine Treffers-Daller, Michael Daller, Liz Peña, Alexandra Dunn, two anonymous reviewers and the editorial staff at *BLC* for their feedback on earlier versions of this manuscript. This research was supported by R01 grants (NIDCD 011492, NICHD050287 and NICHD 051030), and by a P50 (AG05131) from NIH/NIA. The authors have no conflicts of interest to report.

Address for correspondence:

Tamar H. Gollan, 9500 Gilman Drive, 0948, Department of Psychiatry, University of California, San Diego, La Jolla, CA 92093-0948, USA  
tgollan@ucsd.edu

language which were as high as .74 in some cases in Marian et al., 2007).

However, correlations between self-reported proficiency and objective measures of proficiency are far from perfect, and they do not address a different question, which is: How accurately can bilinguals classify themselves into language dominance groups? Some have argued that bilinguals are “notoriously bad” (Dunn & Fox Tree, 2009, p. 275) at providing such ratings (Hakuta & D’Andrea, 1992), and the issue of measuring bilingual language proficiency and dominance is timely (e.g., Bedore et al., 2011; Daller, 2011; Treffers-Daller, 2011), but no studies considered how accurately bilinguals report which language is dominant on a case-by-case basis. In clinical settings examinees are often asked which language they prefer and then are tested exclusively in that language. Thus, it is important to assess the accuracy of such reports for predicting language dominance (Lim, Rickard Liow, Lincoln, Chan & Onslow, 2008). Testing in a non-dominant language will underestimate performance, and testing in the dominant language may be more likely to distinguish patients from healthy controls (Gollan, Salmon, Montoya & da Pena, 2010), which is often the goal in clinical settings.

The question “Which language is your dominant language?” can also be viewed as inherently flawed given that for many bilinguals one language is dominant in one domain whereas a different language is dominant in another domain (e.g., at home versus at work; this issue is discussed at length by Grosjean, 2008). Evidence for this phenomenon can be found in the assessment of picture naming skills which improve for bilinguals when they are credited for producing a name in either language (for similar approaches see Bedore, Peña, García & Cortez, 2005; Kohnert, Hernandez & Bates, 1998). This improvement in naming scores with alternative scoring procedures is found in bilingual children (Bedore, et al., 2005; Pearson, Fernández & Oller, 1993; Umbel, Pearson, Fernández & Oller, 1992) in college-aged and middle-aged adult bilinguals (Gollan & Silverberg, 2001; Kohnert, et al., 1998), in aging bilinguals (Gollan, Fennema-Notestine, Montoya & Jernigan, 2007), and for bilinguals with Alzheimer’s disease (Gollan et al., 2010). Scores improve when names in either language are credited because bilinguals know some names in their non-dominant language that they do not know in their otherwise usually more dominant language. Thus, the usually non-dominant language may be dominant in some situations, and even if bilinguals could be accurate in saying which language is dominant overall, testing in just one language would still provide an incomplete assessment of language proficiency in some important ways.

A different approach to establishing which language is dominant is to test bilinguals in both languages on an objective measure. However, objective measures can be

biased if they are more difficult in one language than the other. Further complicating matters, it is not always clear how to design difficulty-matched measures across different languages. This can be particularly challenging with language pairs that are structurally distinct (e.g., English and Chinese differ greatly in orthography, phonology, and morphology; Lim et al., 2008), but will be present to at least some degree with any language pair (Grosjean, 1998). For example, the Boston Naming Test (BNT; Kaplan, Goodglass & Weintraub, 1983) was designed for monolingual English speakers, and is graded for difficulty in English such that relatively easy items appear at the beginning of the test and the most difficult items towards the end of the test. The final item is *abacus* an item that is quite difficult in English, but because abacuses are more common in China than they are in the USA, it is relatively easy to name in Mandarin. Thus, an item that is difficult in one language may be relatively easy in the other and vice versa (see also Kohnert et al., 1998).

One way around this problem is to create parallel versions of a test with different items for each language. However, this introduces a different problem which is how to establish the criterion of reference for difficulty. For example, it might be stipulated that a test is difficulty-matched for English and Spanish if monolingual speakers of similar age and education levels obtain equivalent scores on the test (Peña, 2007). This approach is becoming common practice in the field; for example, the Bilingual Aphasia Test (Paradis & Libben, 1987) has parallel versions with some overlapping and some different items for each language, and the Woodcock–Muñoz test (see Woodcock & Muñoz-Sandoval, 1996) has different items for testing in Spanish than the Woodcock–Johnson set has for testing in English (see Mather & Woodcock, 2001). Similarly the TVIP (Test de Vocabulario en Imágenes Peabody; Dunn, Padilla, Lugo & Dunn, 1986) was created by selecting subsets of Spanish-appropriate items from two versions of the PPVT (Peabody Picture Vocabulary Test; Dunn & Dunn, 1981, 1987). The use of different items in each language will work well for assessing proficiency in an individual target language, but not necessarily for comparing across languages given possible difficulties with matching monolingual speakers across cultures (e.g., a high school education in the USA may not be equivalent to a high school education in a different country; Byrd, Sanchez & Manly, 2005). In some respects this approach also seems to adopt the questionable assumption (Grosjean, 1989) that bilinguals should ideally be able to function like a monolingual in each language.

In the current study we examined the utility of self-reported proficiency ratings for establishing spoken language dominance. As objective measures of spoken proficiency, participants were interviewed in each language by a bilingual experimenter using a structured oral proficiency interview (OPI). In addition, participants

named pictures in each language using the Multilingual Naming Test (MINT; a new naming test that was designed for bilingual speakers), and in Experiment 1 also the Boston Naming Test. Although self-report of language dominance has been criticized, we hypothesized that dominance ratings on the group level would be at least as reliable as correlations between self-report and measures of ability in each language because individuals may vary in their standards of excellence, and dominance ratings control for such differences but ratings of absolute level of ability do not. For example, some people might never rate themselves as Superior on any domain even though their abilities may in fact be superior in objective terms relative to others. Conversely, other individuals might overestimate their abilities relative to others. Ratings of language dominance would not be as affected by such differences given their focus on ability in one versus the other language within the same person, rather than on ability in each language relative to other people.

## Experiment 1: Young bilinguals

### Methods

#### Participants

A total of 112 young adults (56 bilinguals and 56 monolinguals) participated. Most were undergraduates at the University of California, San Diego (UCSD), and participated in exchange for course credit. A smaller number received payment (\$20) for their participation. Four bilinguals were excluded from further analyses because they had to leave before they could complete all of the tasks. In addition, 19 monolinguals were excluded for being partially bilingual. The criteria used to classify monolinguals were as follows: (a) must rate their ability to speak a language other than English as less than 5 (which corresponds to “intermediate middle” on the 10-point scale in the appendix), and (b) must report using English at least 95% of the time during childhood. These criteria were developed based on the bilingual data; all but two bilinguals rated their Spanish-speaking abilities as greater than 6 (the remaining two rated their Spanish-speaking ability as 5). In addition, all bilinguals rated their percentage of English use when growing up as between 10% and 93%. Participant characteristics are shown in Table 1, with bilinguals separated into three groups: Spanish-dominant bilinguals, who rated their Spanish as more proficient than their English ( $n = 10$ ); balanced bilinguals, who selected the same rating for each language ( $n = 7$ ); and English-dominant bilinguals, who rated their English as more proficient than their Spanish ( $n = 35$ ).

#### Materials and procedure

Participants signed consent forms and completed a Language History Questionnaire at the start of the testing

session, followed by an English vocabulary test (the Shipley Vocabulary Test; Shipley, 1946; which consists of 40 multiple-choice synonym identification questions), and a test of non-verbal reasoning skills (the Matrices Subtest of the Kaufman Brief Intelligence Test, Second Edition, KBIT-2; Kaufman & Kaufman, 2004; which consists of 46 designs with a missing element that participants complete by selecting an element from multiple-choice options). Participants began with the first item (rather than beginning at an age-specific start point). Raw Shipley and Matrices scores are shown in Table 1.

After completing these tests, participants were interviewed to assess spoken language proficiency, and then were asked to name pictures from the Boston Naming Test (BNT; Kaplan et al., 1983) and the Multilingual Naming Test (MINT) with test order (BNT, MINT), and language-of-testing (English, Spanish) in counterbalanced order between subjects. Monolinguals were tested in English only. Bilinguals were interviewed in both languages, and named pictures in both languages. To minimize language switching, the proficiency interview and naming tests were administered in succession in one language, followed by interview and then naming tests in the other language. Phonemic cues were not administered for either naming test, and participants were asked to name all pictures in both tests (i.e., testing did not begin in the middle of the test). Tasks were presented on a Macintosh computer with a 17-inch color monitor using PsyScope 1.2.5 (Cohen, MacWhinney, Flatt & Provost, 1993) and a bilingual experimenter recorded naming accuracy during testing, and testing sessions were also audio-recorded for later verification of scoring. The testing protocol took about an hour and a half to complete for most participants, and no more than two hours.

#### Self-ratings of language proficiency

As part of the questionnaire, participants were asked to rate their proficiency level using a 10-point scale modified and shortened from guidelines published by the American Council on the Teaching of Foreign Languages (ACTFL). ACTFL introduces ten categories used to classify a speaker's language abilities: Superior (10), Advanced High (9), Advanced Middle (8), Advanced Low (7), Intermediate High (6), Intermediate Mid (5), Intermediate Low (4), Novice High (3), Novice Mid (2), and Novice Low (1). The modified guidelines for spoken proficiency that were used here are shown in the appendix. The full-length guidelines as published by ACTFL can be obtained on the “publications” tab at <http://www.actfl.org/i4a/pages/index.cfm?pageid=1>.

#### Oral proficiency interview (OPI)

The proficiency interviews were based on the format used by ACTFL for assessing spoken language proficiency. Questions appropriate for Novice levels (levels 1–3) were excluded because of the focus on relatively proficient

Table 1. Means (range in parentheses), and SD of participant characteristics in Experiment 1. Bilingual means reported separately based on self-rated language dominance. All monolinguals are English-dominant. Statistical comparisons reported in this table are to be interpreted with caution given the small number of participants in some cells.

	Spanish-dominant			Balanced bilinguals			English-dominant			Monolinguals		
	bilinguals (n = 10)			(n = 7)			bilinguals (n = 35)			(n = 36)		
	M	SD	Range	M	SD	Range	M	SD	Range	M	SD	Range
Age	19.3	1.0	18–21	23.1 <sup>e</sup>	5.8	19–36	20.7*	2.2	18–29	20.1	1.7	18–25
% Female	80.0	n/a	n/a	71.0	n/a	n/a	80.0	n/a	n/a	70.0	n/a	n/a
Education	13.1	1.0	12–15	14.9 <sup>ee</sup>	1.9	13–19	13.9*	1.2	12–17	13.6	1.3	12–16
Age 1st Exposure to English	6.4	2.5	1–10	2.9 <sup>eee</sup>	2.3	0–6	3.0 <sup>***</sup>	2.7	0–10	0.6 <sup>†††</sup>	1.9	0–3.5
Age 1st Exposure to Spanish	0.9	2.5	0–8	0.1	0.4	0–1	0.6	1.5	0–6	10.2	4.4	0–14
% Currently Using Spanish	39.5	22.2	0–70	17.1 <sup>ee</sup>	7.6	5–30	13.6 <sup>***</sup>	12.7	.02–50	n/a	n/a	n/a
% Used Spanish growing up	62.0	15.3	40–90	39.3 <sup>eee</sup>	9.3	25–50	35.2 <sup>***</sup>	12.6	7–61	n/a	n/a	n/a
How often speak bilinguals <sup>1</sup>	5.3	1.8	2–7	4.2	1.7	2–6	4.5	1.8	1–7	n/a	n/a	n/a
How often speak to bilinguals growing up <sup>1</sup>	6.1	1.9	1–7	6.0	1.0	5–7	6.1	1.1	3–7	n/a	n/a	n/a
How often switch languages <sup>2</sup>	3.6	1.5	1–5	3.4	1.7	1–5	3.1	1.2	1–5	n/a	n/a	n/a
How often switched growing up <sup>2</sup>	3.5	1.6	1–5	3.1	1.4	1–4.5	2.9	1.3	1–5	n/a	n/a	n/a
Primary parent education level	10.2	4.5	4–16	12.6	4.9	4–18	10.8	3.9	1–18	15.1 <sup>†††</sup>	2.8	10–20
Secondary parent education level	9.9	4.5	3–16	10.4	5.5	4–18	10.3	4.8	2–20	15.1 <sup>†††</sup>	2.6	12–20
Shipley Vocabulary Test	28.4	3.2	24–33	29.7	3.6	26–36	29.7	3.0	21–37	31.6 <sup>††</sup>	2.9	26–38
Matrices Reasoning Subtest	33.8	3.6	29–40	39.3 <sup>eee</sup>	2.1	36–41	35.3	4.3	27–45	39.3 <sup>††</sup>	4.2	28–46
Self Ratings of Proficiency <sup>3</sup>												
English speaking	7.9	1.1	6–9.5	9.6 <sup>eee</sup>	0.8	8–10	9.5 <sup>***</sup>	0.7	7–10	9.6	0.8	8–10
English listening	8.4	0.8	7–10	9.7 <sup>eee</sup>	0.5	9–10	9.7 <sup>***</sup>	0.7	8–10	9.7	0.7	8–10
English writing	7.6	0.7	7–9	9.4 <sup>eee</sup>	1.0	8–10	9.5 <sup>***</sup>	0.8	7.5–10	9.4	0.8	8–10
English reading	8.4	1.0	7–10	9.6 <sup>ee</sup>	0.8	8–10	9.6 <sup>***</sup>	0.8	7–10	9.6	0.6	8–10
Spanish speaking	9.5	0.8	8–10	9.6	0.8	8–10	7.8 <sup>***</sup>	1.0	5–9	2.8 <sup>†††</sup>	1.0	1–4
Spanish listening	9.4	0.8	8–10	9.4	0.9	8–10	8.8*	0.9	6.5–10	3.4 <sup>†††</sup>	1.4	2–7
Spanish writing	8.3	1.3	6–10	8.1	1.4	7–10	7.2 <sup>**</sup>	1.2	4–9	3.1 <sup>†††</sup>	1.5	1–7
Spanish reading	9.0	1.2	7–10	8.9	1.2	7–10	7.9 <sup>**</sup>	1.4	4–10	3.5 <sup>†††</sup>	1.4	1–7

<sup>e</sup>Marginally significant t-test comparing Spanish-dominant to balanced bilinguals (p < .10)

<sup>ee</sup>Significant t-test comparing Spanish-dominant to balanced bilinguals (p < .05)

<sup>eee</sup>Significant t-test comparing Spanish-dominant to balanced bilinguals (p < .01)

\*Marginally significant t-test comparing Spanish-dominant to English-dominant (p < .10)

\*\*Significant t-test comparing Spanish-dominant to English-dominant (p < .05)

\*\*\*Significant t-test comparing Spanish-dominant to English-dominant (p < .01)

†Significant t-test comparing English-dominant to monolinguals (p < .05)

††Significant t-test comparing English-dominant to monolinguals (p < .01)

<sup>1</sup>The following seven-point scale was used: 1 = rarely or never, 2 = less than one hour per day, 3 = about one hour per day, 4 = about 2 hours per day, 5 = about 3–4 hours per day, 6 = about 5 hours per day, 7 = 6 or more hours per day.

<sup>2</sup>The following five-point scale was used: 1 = just once to switch out of English, 2 = occasionally, 3 = two or three times in each conversation, 4 = several times in each conversation, 5 = lot or sometimes even constantly.

<sup>3</sup>Self-ratings were based on the following 10-point scale: 1 = novice low, 2 = novice middle, 3 = novice high, 4 = intermediate low, 5 = intermediate middle, 6 = intermediate high, 7 = advanced low, 8 = advanced middle, 9 = advanced high, 10 = superior.

early bilinguals. Two sets of six interview questions were created. The first question in each set was relatively easy and could be answered mostly in the present tense (e.g., “Where did you grow up? How is it similar to or different

from San Diego?”). The second question in each interview set asked speakers to describe a picture (either the *Cookie Theft* picture from the Boston Diagnostic Aphasia Exam, or a picture of a scene depicting a broken window, the child

who broke the window hiding behind a bush, and an adult accusing a different child of breaking the window). The third and fourth questions were designed to elicit past and future tense constructions (e.g., “Tell me about your first day at UCSD. What was it like? What do you remember most about it?” and “Tell me about what you will do next week. Where will you be and what will you be doing each day?”). The last two questions in each set were designed to provide speakers with an opportunity to produce more difficult constructions typical of educated native speakers (e.g., “Some parents think that bilingual children will not do as well in school as monolingual children. Others say bilingualism is an advantage. What do you think? How would you try to convince someone that your view is the right one?”). Monolinguals completed only one set of interview questions in English (with question set counterbalanced between subjects). Bilinguals completed both sets (one in each language with counterbalanced assignment of question set to language between subjects).

Participants were interviewed by one of two proficient Spanish–English bilingual experimenters who assigned each participant a rating using the same guidelines as those shown in the appendix. After data collection, a third multilingual experimenter listened to all of the proficiency interview recordings and assigned each participant a rating for each language (using the same scale). Perhaps because of the truncated range of bilingual proficiency levels (no low-proficiency bilinguals were tested), and because two different raters provided the initial ratings, the correlation between the final ratings (provided by the single third rater) and initial ratings (some of which were provided by one experimenter and some by a second experimenter) were not very high; for English was  $r = .55$ ,  $p < .01$ , and for Spanish it was  $r = .60$ ,  $p < .01$ . However, the average difference between the third rater and the initial two raters was quite small; just over half a point of difference on average for both languages ( $M = 0.72$ ;  $SD = 0.58$  for English, and  $M = 0.87$ ;  $SD = 0.73$  for Spanish). Thus, on average, the ratings matched each other within a difference of less than one point on the 10-point scale in both languages. For internal consistency, the ratings provided by the third rater were used in all statistical analyses reported below (with the exception of one initial rating for one person in one language because the recording was corrupted and thus the third experimenter could not rate this interview).

#### *Multilingual Naming Test (MINT)*

A set of 68 black-and-white line drawings were selected and presented in order of estimated increasing difficulty. To cater the test to multilingual speakers, target pictures were selected from a variety of sources with the following constraints. First, pictures with cognate names (i.e., translation equivalents that are similar in form across languages were excluded; e.g., *pyramid* is *pirámide* in

Spanish; see Gollan, et al., 2007 for an analysis of cognate effects on the BNT). Cognates were excluded in an attempt to maximize the extent to which the test measures language-specific knowledge without influence from the other language. Second, an attempt was made to include a range of item difficulty but with a greater proportion of medium difficulty items than typically included in naming tests designed for monolinguals (e.g., the BNT; Kaplan et al., 1983). The rationale here was that sensitivity to bilingual naming skills might be better with a slightly easier test given that bilinguals often obtain lower naming scores than monolinguals, and bilinguals might be completely unfamiliar with some of the very low frequency items towards the end of the test (e.g., Gollan & Brown, 2006; Gollan, Montoya, Cera & Sandoval, 2008; Roberts, Garcia, Desrochers & Hernandez, 2002). Inclusion of a greater range of medium difficulty items might be especially important for assessing naming ability in a non-dominant language (given that items that are too difficult would simply elicit “don’t know” responses).

Finally, these criteria were applied with consideration of four languages including Spanish, English, Mandarin Chinese, and Hebrew to allow for eventual cross-study comparison of bilinguals of different language combinations (though here we present only the Spanish–English data). To this end, several bilingual experimenters were consulted during initial item selection including two Spanish–English bilinguals, two Hebrew–English–Spanish trilinguals, and three Mandarin–English bilinguals. The initial item set was piloted with a larger set of words in English, Spanish, Hebrew, and Mandarin ( $n \approx 5$  per language). Items were eliminated if they were cognates with English words, seemed to be more difficult to name in one language than in the others, or had multiple names in any of the four languages. Thus, the resulting item set might be relatively culture-neutral when compared with an item set designed for use with just one (or even just two) languages. However, we caution the test would be unlikely to work for other languages (i.e., be biased against or for languages that were not included in piloting and item development; e.g., cognate status is something that would vary across language pairs and could have powerful effects on naming scores; e.g., Costa, Caramazza & Sebastián-Gallés, 2000; Costa, Santesteban & Caño, 2005; Gollan & Acenas, 2004; Gollan et al., 2007; Roberts & Deslauriers, 1999).

Table 2 illustrates the material characteristics with means for BNT items as a point of comparison (a full list of items is also shown in Supplementary Materials.<sup>1</sup>

<sup>1</sup> Naming rates for each item by age group and proficiency level are available online, on the Journal’s website, as Supplementary Materials accompanying the present article (see the online version of the present article via [journals.cambridge.org/bil](https://journals.cambridge.org/bil)).

Table 2. Multilingual Naming Test item characteristics.

		Multilingual Naming Test		Boston Naming Test	
		M	SD	M	SD
ENGLISH	Length in syllables	1.37	0.54	2.00**	0.92
	Length in phonemes	3.91	1.29	5.32**	1.94
	Corpus of Contemporary American English	62.21	165.80	20.14	70.54
	Celex frequency	69.94	183.80	24.51	91.49
	Log frequency	1.22	0.71	0.64**	0.59
	English Noun Lemma frequency	57.82	116.45	23.64	87.95
	Kučera & Francis frequency	59.74	143.31	22.97	90.22
SPANISH	Length in syllables	2.68	0.89	3.05	1.28
	Length in phonemes	6.00	1.92	6.80*	2.62
	Corpus del Español	14.89	33.29	7.78	38.92
	Lexesp frequency	30.96	66.61	16.97	82.56
	Log frequency	0.97	0.66	0.56*	0.55

\*Significant difference between the MINT and the BNT items at  $p < .05$  level

\*\*Significant difference between the MINT and the BNT items at  $p < .01$  level

Item characteristics were obtained using a program called N-WATCH (Davis (2005) for English, and using Buscapalabras (Davis & Perea, 2005) for Spanish, and from the Corpus del Español (Davies, 2002). Frequency counts for English are from the Count of Contemporary American English (Davies, 2008), CELEX (Baayen, Piepenbrock & Gulikers, 1995) and Kučera & Francis (1967), and for Spanish from the LEXESP database (Sebastián-Gallés, Martí, Cuetos & Carreiras, 2000). Consistent with the goal of making the MINT a little easier than the BNT, the MINT names are shorter (in syllables and number of phonemes) and higher frequency in both languages than the BNT names. Given other selection restrictions we did not attempt to match across languages for length; thus, English words tended to be shorter on average than Spanish words. The means also suggest that the English names are higher frequency than the Spanish names, but note that the validity of this comparison is compromised by the fact that the frequency counts were not matched across languages, and that the frequency databases for Spanish were based on texts from many countries, whereas nearly all of the bilinguals in the current study originated from Mexico. It should also be noted that monolingual frequency counts may not be as accurate for bilingual speakers. A complete list of names used most often to name MINT pictures, any alternative names that were counted as correct (e.g., *teeter totter* was accepted as a correct response for *seesaw*), and naming rates for each item by age group and proficiency level can be found in Supplementary Materials.

## Results

Table 3 reports the means (Ms) and standard deviations (SDs) for bilinguals' self-rated spoken language proficiency, the oral proficiency interview (OPI) ratings, and proportion correct on the MINT and the BNT in English and in Spanish broken down by self-rated dominance groups. For ease of exposition we group together the OPI, MINT and BNT scores under the term OBJECTIVE measures because they do not rely on bilinguals' self-ratings (note however, that the OPI is technically not objective in the sense that the interview scores exist in the minds of the interviewers). Briefly summarized, results reveal significant correlations between measures, but these are far from perfect. Self-report, proficiency interview, and the MINT (but not the BNT) agreed with each other in classifying bilinguals into groups, but when considering degree of language dominance (rather than simple classification into groups) the naming tests (especially the BNT) classified bilinguals as more English-dominant than the self-ratings and proficiency interviews.

### Correlations between measures

We began by considering correlations between measures in each language, a language dominance score, and an index score designed to measure the degree of balanced bilingualism. We calculated a dominance score for each of the four measures (self-ratings, OPI, MINT, and BNT) by subtracting the Spanish scores from the English scores (thus negative difference scores reflect

Table 3. Mean (M) and standard deviation (SD) of self-ratings, oral proficiency interview ratings, and proportion correct on naming tests in each language for bilinguals in each self-rated language-dominance subgroup in Experiments 1 and 2.

	Self-rated spoken proficiency				Oral proficiency interview				Multilingual Naming Test (MINT) <sup>a</sup>				Boston Naming Test (BNT)			
	English		Spanish		English		Spanish		English		Spanish		English		Spanish	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Young bilinguals	Self-rated dominance				Self-rated dominance				Self-rated dominance				Self-rated dominance			
	7.9	1.1	9.5	0.8	8.3	0.9	8.5	0.8	.826	.056	.843	.069	.688	.103	.680	.158
	9.6	0.8	9.6	0.8	9.3	0.8	8.1	0.9	.908	.048	.752	.092	.829	.103	.531	.131
	9.5	0.7	7.8	1.0	8.8	1.0	7.5	1.2	.905	.043	.694	.136	.814	.099	.459	.156
	9.6	0.8							.954	.031			.899	.053		
Older bilinguals	Self-rated dominance				Self-rated dominance				Self-rated dominance				Self-rated dominance			
	6.9 <sup>b</sup>	1.3	9.4	1.0	6.9	1.6	9.1	1.3	.810	.158	.869	.080				
	7.6	0.8	7.6	0.8	8.7	0.3	7.0	1.8	.892	.085	.804	.100				
	9.6	0.7	6.8	1.3	8.9	1.2	6.7	0.9	.935	.035	.679	.067				

<sup>a</sup>MINT and BNT scores are reported as proportion correct for ease of comparison across tests given different number of items (68 in the MINT and only 60 in the BNT).

<sup>b</sup>Older adults self-ratings were adjusted from a seven-point scale to a 10-point scale to for ease comparison to other scores ((rating/7) × 10). Actual ratings for English were 5.57 (SD = 1.13) and for Spanish 5.75 (SD = 1.11; see also Table 7).

Spanish dominance, and positive scores reflect English dominance; see Figure 1, on next page). Index scores were calculated for each of the four measures by dividing the score in whichever language produced the lower score by the score in the other language (which produced a higher score; see Figure 2). For example, a bilingual who named 60 pictures in English and only 30 in Spanish on the BNT would be classified as 50% bilingual according to the BNT (as would a bilingual who named 30 pictures in English and 60 in Spanish). Or using ratings as another example, a bilingual with a Superior rating for English (i.e., 10) and an Advanced Middle rating for Spanish (i.e., 8) would be classified as 80% bilingual. Bilingual index scores range from 0 to 1 and measure the extent to which knowledge of each language is similar (ignoring direction of dominance and ignoring absolute ability level; see also Gollan et al., 2010). The bilinguals tested here all scored at least 79% correct in their dominant language, and between 38% and 94% correct in their non-dominant language on the MINT (thus no bilinguals had extremely low scores in both languages, and all were at least moderately proficient bilinguals).

Table 4 shows the between measure correlations. As previously reported (e.g., Marian et al., 2007), there were significant correlations between self-reported proficiency in English (the dominant language for most participants) and objective measures (OPI, MINT, and BNT scores) ranging from  $r = .281$  to  $r = .503$ , and correlations tended to be higher between self-reported level of proficiency in Spanish (the non-dominant language for most participants) ranging  $r = .425$  to  $r = .520$ . Interestingly, and providing evidence against claims that bilinguals cannot accurately report which language is dominant (Dunn & Fox Tree, 2009), the correlations between self-reported ratings of language dominance and objective measures of language dominance tended to be higher, ranging from  $r = .585$  to  $r = .622$ . Thus, bilinguals were at least as accurate, or even more accurate, in estimating which of their own two languages is dominant as they were at estimating their absolute level of ability in each language.

In contrast, the correlations between self-rated index scores and objective index scores were substantially smaller and only marginally significant, ranging from  $r = .197$  to  $r = .268$ . Thus, whereas bilinguals are relatively accurate in indicating which language is dominant, they are relatively less able to estimate the extent of difference in proficiency between languages (ignoring language dominance and focusing instead on the extent to which knowledge of the two languages is similar or balanced). Finally, objective measure index scores were strongly correlated with each other, ranging from  $r = .669$  to  $r = .858$ . Taken together, these correlations suggest that self-report measures can predict language dominance (though their utility for this purpose is far from perfect), and

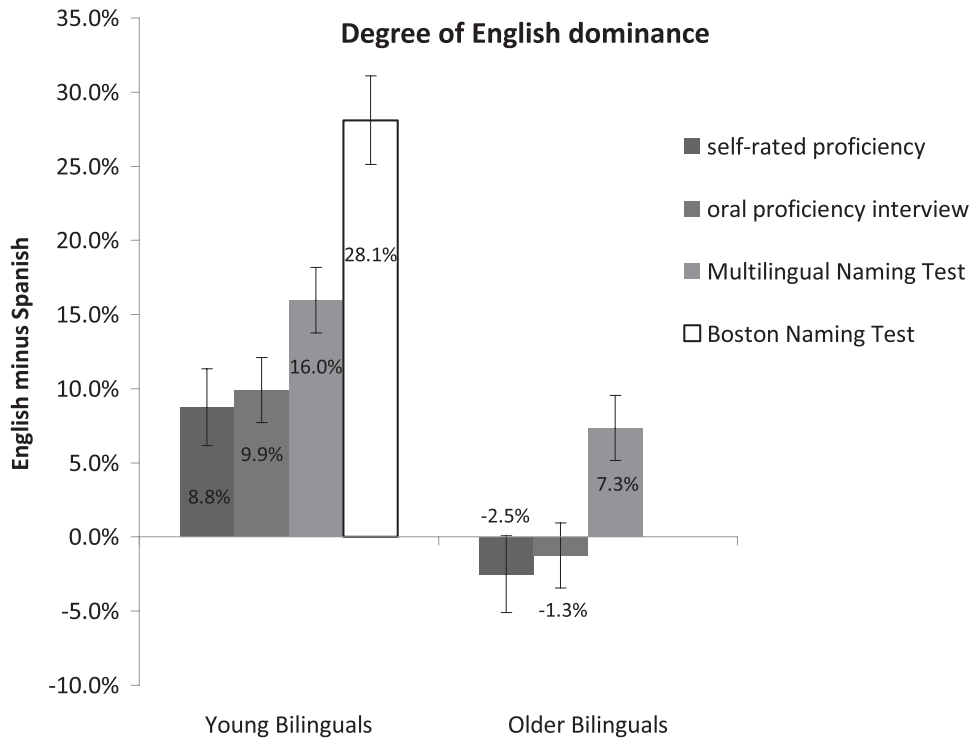


Figure 1. Average degree of English dominance for self-report and objective measures in Experiment 1 (young bilinguals) and Experiment 2 (older bilinguals). Difference scores are calculated by subtracting percent adjusted Spanish scores from English scores.

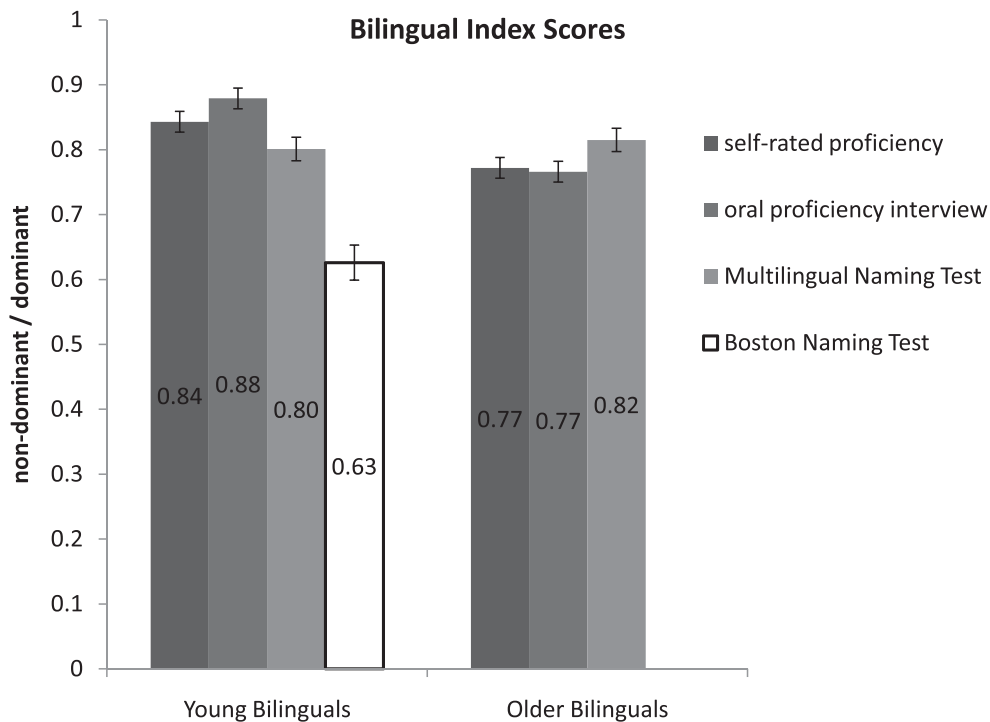


Figure 2. Average bilingual index scores for self-report and objective measures in Experiment 1 (young bilinguals) and Experiment 2 (older bilinguals). Index scores reflect the extent to which proficiency in the two languages is balanced (ignoring direction of dominance), and are calculated by putting whichever language score is lower for each measure in the numerator, and the other language score (the higher scores) for each measure in the denominator.



Table 4. Pearson bivariate correlations between self-rated proficiency and objective measures of proficiency in each language, language dominance difference scores, and bilingual index scores for young bilinguals in Experiment 1 ( $n = 52$ ).

	English			Spanish		
	Self-rating	Oral proficiency interview	Multilingual Naming Test	Self-rating	Oral proficiency interview	Multilingual Naming Test
Oral proficiency interview	.281			.425		
<i>p</i> -value	.043			.002		
Multilingual Naming Test	.460	.397		.520	.518	
<i>p</i> -value	.001	.004		<.001	<.001	
Boston Naming Test	.503	.359	.855	.498	.495	.868
<i>p</i> -value	<.001	.009	<.001	<.001	<.001	<.001
	Language dominance (English minus Spanish)			Bilingual index scores		
Oral proficiency interview	.585			.268		
<i>p</i> -value	<.001			.055		
Multilingual Naming Test	.605	.794		.256	.705	
<i>p</i> -value	<.001	<.001		.067	<.001	
Boston Naming Test	.622	.751	.893	.197	.669	.858
<i>p</i> -value	<.001	<.001	<.001	.161	<.001	<.001

that self-report should not be used to measure degree of balanced bilingualism.

Other correlations shown in Table 4 are of interest. Analyses reported in later sections reveal the BNT as an outlier measure; however, despite these differences the correlation between the BNT and MINT were quite high, ranging from  $r = .855$  to  $r = .893$ . In addition, objective measures of language dominance were strongly correlated with each other, ranging from  $r = .751$  to  $r = .893$  (relative to correlations between self-report and objective measures of language dominance which, as noted above, ranged from  $r = .585$  to  $r = .622$ ). Thus, objective measures of language dominance are probably a better choice than self-report measures.

### **Young bilinguals' ability to self-report language dominance**

#### *Dominance classification into subgroups*

Because dominance classification is often of interest in absolute terms (correct or incorrect) we further investigated correspondence between self-reported and objective measures of language dominance using measure-anchored cut-off scores. Note that we did not ask bilinguals to say which language is dominant (which involves directly comparing the two languages); but dominance ratings can be inferred by inspecting the ratings for each language (and allowing a "balanced"

category). In self-ratings the smallest difference between languages was half a point (a 5% difference) on the 10-point scale we provided (see Appendix). Thus, for balanced bilingualism we allowed any difference of less than 5% in either direction (i.e., English better than Spanish or Spanish better than English) to be classified as objectively balanced, and any difference of 5% or greater in either direction to be classified as objectively dominant in one or the other language (depending on the direction of the difference). Thus cut-offs for Spanish-dominant bilinguals were difference scores of  $-5\%$  and greater; for balanced bilinguals, from  $-4.9\%$  to  $4.9\%$ ; and for English-dominant bilinguals,  $5\%$  and greater. The OPI ratings were on the same 10-point scale as the self-ratings, but MINT and BNT scores were based on a 100-point scale. Thus, for purposes of comparison, naming scores were converted to a 10-point scale by dividing by 10. For example, naming score differences of 5% were considered equivalent to 0.5 points on the 10-point scale used for self-ratings and OPI. Note that these cut-off scores are arbitrary in that there is no sense in which a 5% difference necessarily qualifies as a point in which a significant, measurable, or "true" difference is present. Thus, the scale is consistent across measures and provides a means for comparison but the extent to which misclassifications truly qualify as such could be debated (we return to this in the "General discussion" section below).

Table 5. Percentage of young bilinguals in Experiment 1 whose self-rated language dominance matched or differed from objective measures of dominance. For cases in which self-ratings and objective classifications of dominance do not match the range of discrepancy in scores is indicated in parentheses.

	Young bilinguals		
	self-rated as Spanish-dominant (n = 10)		
	Objectively Spanish-dominant	Objectively balanced	Objectively English-dominant
Oral proficiency interview	40%	50% (0%)	10% (10.0%)
Multilingual Naming Test	40%	20% ((-4.4%) – (-2.9%))	40% (5.9% – 11.8%)
Boston Naming Test	40%	0%	60% (6.7% – 21.7%)
		Self-rated as balanced (n = 7):	
Oral proficiency interview	0%	0%	100% (5.0% – 15.0%)
Multilingual Naming Test	0%	0%	100% (7.3% – 25.0%)
Boston Naming Test	0%	0%	100% (16.7% – 41.7%)
		Self-rated as English-dominant (n = 35):	
Oral proficiency interview	3% (-5.0%)	11% (0%)	86%
Multilingual Naming Test	3% (-7.4%)	6% (1.5%)	91%
Boston Naming Test	0%	3% (1.7%)	97%

With this method of classifying bilinguals into three groups (Spanish-dominant, balanced, English-dominant), self-classifications did not differ from OPI classifications and MINT scores, (both  $ps \geq .22$ ), but self-classifications were significantly different from BNT classifications,  $\chi^2(2, n = 52) = 8.92, p = 0.01$ . Similarly, OPI ratings did not differ from MINT classifications ( $p = .33$ ), but were significantly different from BNT classifications,  $\chi^2(2, N = 52) = 7.46, p = 0.02$ . Thus, the BNT stands out as significantly different from self-ratings and OPI, though the MINT and BNT classifications did not differ significantly from each other;  $p = .35$ . Table 5 illustrates the percentage of bilinguals in each self-rating group (i.e., Spanish-dominant, balanced, and English-dominant) whose self-ratings seemed to match objective dominance classifications.

#### *Dominance along a continuum*

On average as a group, bilinguals obtained higher scores in English than in Spanish in self-ratings, OPI (oral proficiency interview) ratings, MINT scores, and BNT

scores (all  $ps < .001$ ). However, as shown in Figure 1 above, the extent of English dominance varied across measures (see also Bedore et al., 2011); for self-ratings it was by 8.8% ( $SD = 16.4$ ), for OPI ratings by 9.9% ( $SD = 10.8$ ), for MINT scores by 16.0% ( $SD = 15.6$ ), and for BNT scores by 28.1% ( $SD = 21.4$ ). Six paired t-tests comparing all possible two-way comparisons of these difference scores were all significant ( $ps \leq .001$ ), with one exception, which was that self-ratings and OPI ratings were not significantly different from each other ( $p = .54$ ). Thus, self-ratings agreed with OPI ratings, but not with naming tests and the BNT in particular seemed to stand out in this regard.

Comparing the two naming tests, the degree of English dominance appeared to be considerably greater for the BNT than the MINT. A  $2 \times 2$  ANOVA with test (MINT, BNT) and language (English, Spanish) as within-subject factors, and proportion correct as the dependent variable revealed this interaction to be highly robust statistically. There were main effects of language such that scores were higher in English than in Spanish [ $F(1,51) = 78.010$ ,

$MSE = 0.032$ ,  $\eta_p^2 = .605$ ,  $p < .001$ ], a main effect of test such that scores were higher on the MINT than on the BNT [ $F(1,51) = 352.563$ ,  $MSE = 0.004$ ,  $\eta_p^2 = .874$ ,  $p < .001$ ], and a significant interaction such that English appeared to be more dominant with BNT than with MINT scores [ $F(1,51) = 73.182$ ,  $MSE = 0.003$ ,  $\eta_p^2 = .589$ ,  $p < .001$ ]. Thus, the test not designed for use with Spanish or bilinguals seemed to bias classifications towards English dominance.

*What is the source of discrepancy between subjective and objective measures of language dominance?*

Beginning with the middle of Table 5, of the bilinguals who classified themselves as balanced, none seemed to be balanced by objective measures. Instead, all were classified as English-dominant. Some of these misclassifications were very small (i.e., only 5% and therefore possibly not true misclassifications); however, others appeared to have misclassified themselves much more obviously (e.g., a difference of up to 41.7%). Bilinguals who rated themselves as Spanish-dominant matched objective classifications a bit better; however, here too the match between self-report and objective measures was only 40%. For example, one bilingual who said s/he was Spanish-dominant was classified as English-dominant on the oral proficiency interview (OPI), and six bilinguals who said they were Spanish-dominant obtained higher naming scores on the BNT in English than in Spanish. Finally, in English-dominant bilinguals the match between self-report and objective measures seemed to be better, but even here, one bilingual scored better in Spanish than in English on the OPI, another (a different person) scored better in Spanish than in English on the MINT, and a handful more seemed to be relatively balanced bilinguals on objective measures.

Table 3 illustrates that bilinguals who reported being Spanish-dominant seemed to be the most balanced bilinguals by objective measures, and those who reported being balanced bilinguals tended to be English-dominant. For example, bilinguals who reported being Spanish-dominant on average rated their Spanish to be about 1.5 points better than English, but objective measures revealed very small differences between languages, and suggested that these bilinguals may have overestimated their abilities in Spanish (e.g., they rated their Spanish at 9.5 on average but scored only an 8.5 on the Spanish OPI, and named about 84% of pictures on the MINT). Other studies have also found that the most objectively balanced bilinguals were also those who reported being dominant in, and also have a later age of acquisition for, their second-learned language (see Flege, MacKay & Piske, 2002 for a similar result with Italian–English bilingual immigrants to Canada). Bilinguals who rated themselves as balanced had higher self-ratings overall (over 9.5 on average in both languages) but like self-rated Spanish-

dominant bilinguals also seemed to overestimate their abilities in Spanish (on average scoring between 12% and 29.8% better in English than in Spanish depending on the measure). Bilinguals who reported being English-dominant had virtually the same average rating values as Spanish-dominant bilinguals (just reversed by language; 9.5 for language chosen as dominant and about 7.8 for language chosen as non-dominant), but were more accurate given that objective measures seemed to confirm their English dominance.

Additional subgroup comparisons confirmed that bilinguals who rated themselves as balanced bilinguals resembled English-dominant bilinguals in their objective scores (see also Gollan & Ferreira, 2009). For example, balanced bilinguals rated their abilities in Spanish as higher ( $p < .001$ ), but did not score significantly higher, than English-dominant bilinguals in Spanish on the OPI ( $p = .21$ ), the MINT ( $p = .29$ ) or the BNT ( $p = .26$ ). This lack of differences (between test scores in each language in self-reported Spanish-dominant bilinguals) could not be attributed to lack of sensitivity in the measures given that self-rated balanced bilinguals did score significantly higher than self-rated Spanish-dominant bilinguals in English ( $p = .04$  on OPI; both  $ps < .01$  on MINT and BNT). Similarly, although self-rated Spanish-dominant bilinguals rated their ability in Spanish as significantly higher than their ability in English ( $p < .01$ ), their performance on objective measures was not different between languages (all  $ps \geq .34$ ). Other significant differences of note were that self-rated English-dominant bilinguals were significantly different from those of Spanish-dominant bilinguals in both languages on all measures (all  $ps \leq .01$ ) with the exception of OPI scores in English, which only trended in the expected direction ( $p = .18$ ). Finally, self-rated English-dominant bilinguals did not rate their spoken English proficiency as lower than monolinguals, but named significantly fewer pictures on both the MINT and the BNT ( $ps < .01$ ) confirming previous reports of bilingual disadvantages (e.g., Gollan et al., 2007; 2008; Roberts et al., 2002), and demonstrating sensitivity in the MINT to differences between bilinguals and monolinguals as well as to proficiency differences within bilinguals.

*Degree of balanced bilingualism*

Figure 2 above illustrates the index score means. The self-ratings, proficiency interviews, and the MINT all classify bilinguals as between 80% and 88% bilingual. In contrast, the BNT seems to underestimate the degree of bilingualism, classifying them as only 63% bilingual. The BNT index scores were significantly lower than all other index scores (all  $ps < .001$ ). MINT index scores were only marginally different from self-rating index scores ( $p = .06$ ), though like the BNT, the MINT index scores were significantly lower than oral proficiency interview (OPI)

index scores ( $p < .001$ ). Finally, self-rating index scores were only marginally lower than OPI index scores ( $p = .06$ ).

### Discussion

Experiment 1 revealed significant correlations between measures of bilingual language proficiency. As a group young bilinguals were best able to predict their own language dominance, and could also predict their level of proficiency in each language (especially the non-dominant language). In contrast, bilinguals were relatively unable to predict the extent to which they were balanced bilinguals (i.e., self-rated index scores were not significantly correlated with objectively measured index scores in Experiment 1, and not consistently in Experiment 2). For predicting purposes, degree of language dominance, self-ratings and the oral proficiency interview (OPI) ratings agreed with each other, and also with the MINT in absolute classification into groups. However, considering degree of language dominance, both naming tests indicated greater English dominance than self-report and interview measures (see Figure 1). Although bilinguals were fairly good at classifying themselves into three dominance groups (without considering degree of dominance), in all self-assigned dominance groups (English-dominant, balanced, Spanish-dominant) some bilinguals seemed to make classification errors, and these errors seemed to be driven in part by self-rated Spanish-dominant and balanced bilinguals' overestimating their abilities in Spanish, and English-dominant bilinguals overestimating their ability in English. Importantly, the BNT stood out as an outlier in several analyses; it was most likely to classify bilinguals as English-dominant, classified the group as much more English-dominant than any other measure (Figure 1), and also seemed to underestimate the extent of balanced knowledge of the two languages (Figure 2), relative to all the other measures. Before considering the implications of these results, in Experiment 2 we further investigated bilinguals' ability to estimate their own language dominance by testing a group of older Spanish–English bilinguals.

### Experiment 2: Older bilinguals

#### Method

##### Participants

Table 6 shows the characteristics of the 20 older Spanish–English bilinguals who participated in Experiment 2. The majority of older bilinguals ( $n = 15$ ) were recruited for participation from a cohort of healthy bilingual controls at the Alzheimer's Disease Research Center (ADRC) at the University of California, San Diego (UCSD), and were diagnosed as cognitively intact by two senior staff

neurologists using criteria developed by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA; McKhann, et al., 1984) and based on medical, neurological, and neuropsychological evaluations and a number of laboratory tests (to rule out dementia). Five additional Spanish–English bilinguals were recruited from the San Diego area and were assumed to be cognitively intact based on high levels of reported functioning in daily life.

##### Materials and procedure

These were the same as in Experiment 1 with two exceptions. First, the BNT, Shipley vocabulary, and Matrices subtest were not administered. Participants not from the ADRC were tested with the Dementia Rating Scale (DRS; Mattis, 1988), and Mini Mental State Examination (MMSE; Folstein, Folstein & McHugh, 1975) in their self-reported dominant language. For ADRC participants the DRS and MMSE scores were obtained from the most recent annual testing session at the ADRC. In addition, a shorter version of the language history questionnaire was used with self-ratings on a simpler scale ranging from 1 to 7. This simpler scale may be more practical for use in clinical settings.

The OPI ratings were all completed by the same multilingual experimenter who assigned OPI ratings in Experiment 1 (with the exception of two English scores for which recordings were missing and thus scores were taken from the experimenter who administered the interview instead). The correlation between the final and initial ratings for English was  $r = .69$ ,  $p < .01$ , and for Spanish  $r = .86$ ,  $p < .01$ . These correlations are a bit higher than the analogous correlations in Experiment 1, possibly reflecting the slightly broader range of proficiency levels in Experiment 2. All bilinguals had at least some moderate proficiency in both languages and the range was broader in Experiment 2 than in Experiment 1 (based on the one rater who rated all speakers in both studies these ranged from 5.5 to 10 in both languages in Experiment 2, but only from 6.5 to 10 in English and from 6 to 10 in Spanish in Experiment 1). As in Experiment 1, the average difference in rating between the final and initial ratings was low; in this case, under half a point of difference on average between raters for both languages ( $M = 0.19$ ;  $SD = 1.11$  for English, and  $M = 0.43$ ;  $SD = 0.91$  for Spanish). Thus, on average, the ratings matched each other within a difference of less than half of a point on the 10-point scale used to assign OPI ratings in both languages (see Appendix).

On average as a group, older bilinguals were relatively balanced exhibiting comparable English and Spanish self-ratings and OPI ratings (both  $F_s < 1$ ), although MINT scores exhibited some tendency towards English dominance overall [ $F(1,19) = 2.97$ ,  $MSE = 0.018$ ,

Table 6. Means (Ms), standard deviations (SDs), and range of participant characteristics in Experiment 2. Means reported separately based on self-rated language dominance. Statistical comparisons reported in this table are to be interpreted with caution given the small number of participants.

	Spanish-dominant bilinguals (n = 10)			Balanced bilinguals (n = 3)			English-dominant bilinguals (n = 7)		
	M	SD	Range	M	SD	Range	M	SD	Range
Age	75.9	7.2	65–84	82.0	4.6	77–86	77.0	9.9	66–87
% Female	80.0	n/a	n/a	67.0	n/a	n/a	57.0	n/a	n/a
Education	12.5	2.6	9–18	13.3	1.2	12–14	13.9	2.5	11–18
Age 1st Exposure to English	8.9	3.3	4.5–13	1.7	2.9	0–5	3.0***	2.8	0–6
Age 1st Exposure to Spanish	0.0	0.0	n/a	0.0	0.0	n/a	1.1	2.8	0–7.5
% Currently Using Spanish	53.4	30.0	20–99	33.3	15.3	20–50	15.7***	13.7	0–40
% Used Spanish growing up	85.4	19.1	50–100	50	0	50	35.7***	28.8	0–80
How often speak bilinguals <sup>1</sup>	2.5	1.0	1–4	2.7	0.6	2–3	2.4	1.3	1–5
How often speak to bilinguals growing up <sup>1</sup>	0.5	1.3	0–4	2.0	1.0	1–3	2.7***	1.3	1–4
Dementia Rating Scale	137.8	4.3	131–142	131.3	26.7	130–133	137	29.6	131–140
Mini Mental Status Exam	28.3	1.8	25–30	26.7	2.5	24–29	29.6*	0.5	29–30
Self Ratings of Proficiency <sup>2</sup>									
English speaking	4.9	0.9	3–6	5.3	0.6	5–6	6.7***	0.5	6–7
English listening	5.3	0.8	4–7	5.7	0.6	5–6	6.7***	0.8	5–7
English writing	4.4	1.8	1–7	5.3	0.6	5–6	6.1**	1.2	4–7
English reading	5.3	0.8	4–7	6.0	0.0	6	6.4**	0.8	5–7
Spanish speaking	6.6	0.7	5–7	5.3	0.6	5–6	4.8***	0.9	3–6
Spanish listening	6.5	0.9	4–7	4.7	0.6	4–5	3.6***	1.6	2–6
Spanish writing	6.4	0.9	4–7	4.7	0.6	4–5	3.4***	1.8	1–6
Spanish reading	6.5	0.9	4–7	6.3	0.6	6–7	5.2**	0.8	4–6

\*Marginally significant t-test comparing Spanish-dominant to English dominant ( $p < .10$ )

\*\*Significant t-test comparing Spanish-dominant to English dominant ( $p < .05$ )

\*\*\*Significant t-test comparing Spanish-dominant to English-dominant ( $p < .01$ )

<sup>1</sup>The following seven-point scale was used: 1 = rarely or never, 2 = less than one hour per day, 3 = about one hour per day, 4 = about 2 hours per day; 5 = about 3–4 hours per day, 6 = about 5 hours per day, 7 = 6 or more hours per day.

<sup>2</sup>Self-ratings were based on a seven-point scale: 1 = almost none, 2 = very poor, 3 = fair, 4 = functional, 5 = good, 6 = very good, 7 = like native speaker.

$\eta_p^2 = .14, p = .10$ ]. The relatively more balanced profile in the overall means (compared with English dominance for younger bilinguals in Experiment 1) reflects the lower proportion of self-reported English-dominant participants in Experiment 2 (seven out of 20 or 35%) relative to Experiment 1 (35 out of 52 or 67%; compare Tables 1 and 7).

#### Correlations between measures

Table 7 shows the correlations between measures, difference scores, and index scores. As in Experiment 1, there were significant correlations between bilinguals' self-rated proficiency in each language and objective measures, ranging from  $r = .690$  to  $r = .786$ . Also as in Experiment 1, correlations between self-ratings and objective measures of language dominance tended to be larger, ranging from  $r = .794$  to  $r = .876$ ,

whereas correlations between self-ratings and objective index scores tended to be smaller, ranging from  $r = .396$  to  $r = .586$ . Finally, objective measure index scores were correlated with each other, ranging from  $r = .473$  to  $r = .874$ . These analyses confirm those reported in Experiment 1, and demonstrate that older bilinguals can also predict their language dominance, in this case using a simpler rating scale (for details see bottom of Table 6).

#### Older bilinguals' ability to self-report language dominance

##### Dominance classification into subgroups

Using the same measure-anchored cut-off system as in Experiment 1, in Experiment 2, self-classifications did not differ from OPI-classifications or from MINT score classifications, and OPI and MINT scores classifications also did not differ from each other

Table 7. Pearson bivariate correlations between self-rated and objective measures of ability in each language, of language dominance and bilingual index scores for older bilinguals in Experiment 2 (n = 20).

	English		Spanish	
	Self-rating	Oral proficiency interview	Self-rating	Oral proficiency interview
Oral proficiency interview	.690		Oral proficiency interview	.770
<i>p</i> -value	.001		<i>p</i> -value	.000
Multilingual Naming Test	.786	.649	Multilingual Naming Test	.775
<i>p</i> -value	.000	.002	<i>p</i> -value	.000
	Language dominance (English minus Spanish)		Bilingual index scores	
Oral proficiency interview	.794		Oral proficiency interview	.396
<i>p</i> -value	.000		<i>p</i> -value	.084
Multilingual Naming Test	.876	.864	Multilingual Naming Test	.586
<i>p</i> -value	.000	.000	<i>p</i> -value	.007

Table 8. Percentage of older bilinguals in Experiment 2 whose self-rated language dominance matched or differed from objective measures of dominance. For cases in which self-ratings and objective classifications of dominance do not match the range of discrepancy in scores is indicated in parentheses.

	Older bilinguals self-rated as Spanish-dominant (n = 10)		
	Objectively Spanish-dominant	Objectively balanced	Objectively English-dominant
Oral proficiency interview	80%	10% (0%)	10% (5.0%)
Multilingual Naming Test	50%	30% ((-2.9%) – 0%)	20% (14.7% – 16.2%)
	Self-rated as balanced (n = 3)		
Oral proficiency interview	0%	33%	67% (20.0% – 30.0%)
Multilingual Naming Test	0%	33%	67% (5.9% – 25.0%)
	Self-rated as English-dominant (n = 7)		
Oral proficiency interview	0%	0%	100%
Multilingual Naming Test	0%	0%	100%

(*ps* ≥ .26). These results replicate those reported for young bilinguals. Further replicating Experiment 1, self-report and objective classifications did not always match, and depending on which measure was considered there were some total reversals of dominance group. Table 8 illustrates the percentage of older bilinguals of each type (self rated Spanish-dominant, balanced, English-dominant) whose self-ratings seemed to match objective

dominance classifications, and Table 3 illustrates some of the source of discrepancy between self-report and objective measures. Of the three bilinguals who classified themselves as balanced, one was confirmed to be balanced by the OPI, but this same bilingual scored 5.9% better on the MINT in English than in Spanish. Another was classified as relatively balanced by the MINT (scoring 4.4% better in Spanish than in English), but was rated as

20% better in English than in Spanish on the OPI (a rating of 8.5 for English and only 6.5 for Spanish). Among the 10 bilinguals who rated themselves as Spanish-dominant, two scored about 15% better on the MINT in English than in Spanish. Finally, as in Experiment 1, in English-dominant bilinguals the match between self-ratings and objective measures seemed to be better (all seven were classified as English-dominant in all measures).

#### *Dominance along a continuum*

On average, difference scores (English minus Spanish) were relatively balanced (see Figure 1); for self-ratings the scores averaged slightly in the direction of Spanish dominance by 2.5% ( $SD = 27.2$ ), and in OPI ratings by 1.3% ( $SD = 27.0$ ), whereas MINT scores averaged in the direction of English dominance by 7.3% ( $SD = 19.1$ ). As in Experiment 1, paired t-tests revealed significant differences between self-ratings and MINT difference scores, and between OPI and MINT difference scores (both  $ps = .01$ ), but self-rating and OPI based difference scores were not significantly different from each other ( $p = .75$ ).

#### *What is the source of discrepancy between subjective and objective measures of language dominance?*

As in Experiment 1, the three self-reported balanced bilinguals seemed to be English-dominant on objective measures (both OPI and the MINT). Though cross-experiment comparisons are to be exercised with caution (young and older participants were not matched for language proficiency and other characteristics, and were tested with slightly different procedures), in other respects older bilinguals in Experiment 2 seemed to fare better in estimating their language dominance than did young bilinguals in Experiment 1. For example, instead of exhibiting a balanced profile as in Experiment 1, self-rated Spanish-dominant older bilinguals seemed to score significantly better in Spanish than in English on the OPI ( $p = .01$ ), and their MINT naming scores were 5.9% higher in Spanish than in English (instead of just 1.7% higher in Experiment 1; though the 5.9% difference still was not significant,  $p = .27$ ). Finally, as in Experiment 1, older bilinguals who reported being English-dominant had average rating values very similar to those of Spanish-dominant bilinguals (again just reversed by language), but were more accurate given that objective measures confirmed their English dominance (both  $ps < .01$ ).

#### *Assessment of degree of bilingualism*

Figure 2 above illustrates the index score means. The self-ratings, proficiency interviews, and the MINT, classified older bilinguals as between 77% and 82% bilingual and there were no significant differences in index scores across measures (all  $ps \geq .14$ ).

## General discussion

The results of the current study simultaneously validate, and illustrate the limitations of, self-report measures of language proficiency and language dominance. The approach taken here assumes that no single measure will provide a complete assessment of bilingual language proficiency which can vary from domain to domain, and will reflect different aspects of knowledge and skill. A bilingual who is classified as dominant in one language by objective measures but nevertheless rates herself as dominant in the other language is not necessarily “wrong” in this self assessment. Instead, this bilingual may be focusing on something that is not measured by naming tests and proficiency interviews (or other objective tests).

The proficiency interviews in the current study provided an objective measure of language proficiency that is relatively naturalistic, and more similar to self-ratings in a number of ways. Perhaps most notably, interview scores were likely influenced by a range of abilities including lexical retrieval ability, formulation of syntactic structures, perhaps knowledge of colloquial expressions, range of registers, accent, and other skills. In contrast, MINT scores reflect only the ability to retrieve picture names. As such, it might be expected that the interviews would be more strongly correlated with self-ratings which probably also are based on a wide range of abilities (i.e., it is unlikely that bilinguals consider only their ability to produce object names when providing a rating of their ability to speak each language). Moreover, in Experiment 1 both self-ratings and proficiency interview scores were based on the same scale and detailed descriptions of the skills associated with each scale level (see Appendix).

Indeed, self-ratings and interview scores did not differ from each other in determining degree of language dominance (see Figure 1 above), and both differed significantly from dominance classifications derived from naming tests (in both Experiment 1 and Experiment 2). However, Tables 4 and 7 do not confirm this expectation; instead, the correlations between self-ratings and interviews were often smaller than correlations between self-ratings and naming tests, and between interview-ratings and naming tests. Without the proficiency interviews, it might seem that self-ratings and naming tests do not produce perfect correlations because naming tests do not measure a variety of skills, and because the scale of measurement is not the same across these two measures. Instead, it seems that there may be some real differences in language dominance across different domains (Bedore et al., 2011; Grosjean, 2008) – and perhaps also some degree of true error – in self-ratings.

***Can bilinguals tell which language is dominant, and if not why not?***

The current findings begin to provide an answer to the question “Can bilinguals accurately tell which language is dominant?” The answer to this question appears to be yes to some degree – particularly if degree of dominance does not matter (see also Dunn & Fox Tree, 2009). However, bilinguals may still perform relatively better on objective measures in the language they report is not dominant, particularly if measures were not designed for use with bilinguals (i.e., BNT). Moreover, the consequences of classification error will be so great in many circumstances that it would be very wise not to rely exclusively on self-report. Tables 5 and 8 illustrate an estimation of the percent of bilinguals who seemed to have slightly or greatly misclassified their own language dominance in their own self-ratings. Some of the misclassifications include cases of complete dominance reversals (i.e., saying one language is dominant but then performing better in the other language). These were observed in both Experiment 1 and Experiment 2, sometimes with very large discrepancies. Subtler differences were also found and it might be debated as to whether or not they truly qualify as true misclassifications, but could nevertheless have important consequences for conclusions drawn in both clinical settings and for shaping models of bilingual language processing (more on this below).

Our method of classifying bilinguals into groups could be criticized. For example, our 5% cut-off point was anchored to the self-rating scores, and the fact that half a point of difference on the 10-point scale was the smallest distinction chosen by any of the bilinguals. This approach is somewhat arbitrary and not necessarily defensible in its application across measures. For example, in Experiment 2 we used only a seven-point scale and there too a half a point of difference was the smallest distinction used in self-ratings even though half a point corresponds to a greater percentage of difference on a seven-point than on a 10-point scale (which in turn implies that bilinguals’ ratings were influenced to some extent by the scale they were provided with and not exclusively by actual proficiency levels). Having acknowledged this limitation in our approach there are also reasons to believe that a 5% difference constitutes a reasonable cut-off point for misclassifications. For example, a 5% difference on the BNT corresponds to a standard deviation of monolinguals’ naming scores (see Table 3). In terms of cognitive assessment and also in terms of theoretical interpretation, a standard deviation would be considered a significant difference in many (if not most) cases.

The data reported here do not provide a definitive answer as to why some bilinguals seem to misclassify their language dominance but the participant characteristics tables (Tables 1 and 6) as well as the self-reported sub-

group means (in Table 3) provide some clues. First, note several significant differences between subtypes in a range of self-report characteristics. Spanish-dominant bilinguals reported learning English at a later age, and using Spanish relatively more often both currently and when growing up, relative to both English-dominant and balanced bilinguals. In Experiment 1 self-reported balanced bilinguals also had significantly higher non-verbal reasoning scores (this skill was not measured in Experiment 2). Thus, one could speculate that people with higher intellectual ability might be more willing to give themselves a very high rating in both languages (even if such a rating is not warranted!). Looking at the subgroup means (Table 3), one might have expected that bilinguals immersed in a language that is not their self-reported dominant language could be more likely to underestimate the extent to which they have become dominant in the language dominant to the environment. This seemed to be the case for balanced bilinguals (both young and older in Experiments 1 and 2), who rated their abilities as equal in the two languages but then performed better in English on objective measures (proficiency interviews and naming tests). But the means in Table 3 tell a slightly different story especially for young Spanish-dominant bilinguals who underestimated their abilities in English only slightly, but seemed to overestimate their abilities in their dominant language (i.e., Spanish) to a larger extent. Similarly, English-dominant bilinguals (again especially young bilinguals in Experiment 1) seemed to overestimate their abilities in English. Thus, overestimation of abilities in the dominant language seems to be part of the reason why self-report and objective measures of dominance do not match perfectly. The presence of an effect in the same direction for Spanish-dominant and English-dominant bilinguals suggests a locus of discrepancy that is not specific to maintenance of a minority language (e.g., see Hakuta & D’Andrea, 1992, which presented evidence that positive attitude towards maintenance of Spanish proficiency in an English-dominant environment influences proficiency ratings).

The term “overestimation” is used here on the assumption that the objective measures capture an aspect of proficiency that should be included in an ideal measure of proficiency but that self-ratings somehow fail to capture. An alternative possibility is that the self-ratings are more accurate and the objective measures are all flawed, but even if so the correspondence between them is important given that objective measures must be used in testing situations (where the goal will often be to test in whichever language produces a better performance). There is also an assumption of proportional correspondence between measures in scales. As noted above, the extent to which this correspondence is justified could be debated. However, some degree of confidence in the correspondence can be drawn from



the significant correlations between objective measures in these comparisons. Having noted these, it is also important to discuss some of the differences found between objective measures in the extent to which one language was dominant over the other (for the same bilinguals).

### ***Limitations of the BNT for bilingual assessment***

Particularly notable in this regard in the current study was the bias in favor of English on the BNT. For all bilinguals, the BNT seemed to underestimate Spanish proficiency, provided an inaccurate measure of the degree of bilingualism, and distorted language dominance classifications relative to all three other measures (including some complete reversals of dominance classification). For Spanish-dominant bilinguals, the BNT produced the largest proportion of completely reversed classifications of language dominance (see Table 5; i.e., 60% of bilinguals who said they are Spanish-dominant were actually able to name more pictures in English than in Spanish on the BNT). For self-rated balanced bilinguals and English-dominant bilinguals the BNT likely overestimates the extent to which English is dominant over Spanish. The BNT is likely inadequate for assessing bilingual language proficiency because it was not designed for use with bilinguals or with Spanish speakers. (e.g., Allegri et al., 1997; Gollan et al., 2007; Kohnert et al., 1998; Patricacou, Psallida, Pring & Dipper, 2007; Weintraub et al., 2009), and thus the items may be relatively more difficult in Spanish than in English (for discussion see de la Plata et al., 2007, and Peña-Casanova et al. (2009, p. 350); the latter suggest that “more studies about the suitability of each item for assessment of naming ability in Spanish” are needed).

The BNT seemed to be an outlier in terms of both index scores and dominance classifications (see Figure 1 and Table 5). Nevertheless, performance on the two naming tests was highly correlated (see Table 4; the BNT was not used in Experiment 2 with older bilinguals). The correlations indicate that the extent to which the BNT is biased in favor of English (and against Spanish) is relatively uniform across subjects (the direction of difference between languages on the two tests is similar between individuals). Thus, although we caution against using the BNT to assess language dominance and degree of bilingualism, in other respects the BNT may provide a useful measure (e.g., for tracking changes in ability in each language over time; or for determining how bilinguals perform in English). Despite its potential flaws in this context, the BNT remains commonly used both in clinical settings and in experimental research with bilinguals (e.g., Gollan et al., 2010; Rosselli et al., 2000; Silverberg & Samuel, 2004), thus it is important to qualify interpretation of scores with a detailed understanding of

specifically how the test may distort bilingual language assessment.

### ***Implications for research and clinical use***

To facilitate future use of naming tests for these purposes, detailed information about which items on both tests were more difficult in Spanish than in English for different types of bilinguals can be found in Supplementary Materials. In addition, to provide a measure of difficulty level for each item in each language, these Supplementary Materials tables include two columns that show naming accuracy for bilinguals who were rated at the highest possible proficiency level in the OPI (a Superior rating; there were 11 young and two older bilinguals who received this score for English, and two young and three older bilinguals who received this score for Spanish). Finally, Table 9 provides mean (and SD) naming test scores for each language at each self-rated proficiency level. These means may be useful in clinical settings for asking more specific questions relating self-rated proficiency level to performance (e.g., given a rating of X on language Y, what is the range of normal performance?). Note that means go down with each rating level for both naming tests and in both languages, again validating self-ratings (with some exceptions, where the n is small); however, the standard deviations also become larger as the means become smaller (scanning from the top to the bottom, where lower proficiency levels are represented). This suggests greater variability in performance, and reduced reliability of ratings at lower proficiency levels. In addition, with few exceptions, standard deviations tend to be larger in the BNT than the MINT, especially in Spanish; thus, for diagnostic purposes, the MINT may be more useful than the BNT.

Previous studies which claimed that bilinguals are not able to indicate which language is dominant may have drawn this conclusion because of limitations in the choice of measures used to evaluate self-ratings. As an example, in lieu of self-report, Dunn and Fox Tree (2009) developed and recommend the use of a language dominance scale which includes questions about each language for age of acquisition, extent to which bilinguals feel “comfortable” speaking, location of language use, language used for math, presence of foreign accent, schooling, language dominant to the environment, and questions about language loss (including loss of knowledge and forced choice of which language is more important). They reported that bilinguals who were classified as relatively balanced on this scale translated words more slowly than bilinguals with one clearly dominant language, thus demonstrating utility of their measure for predicting performance on an objective measure. In addition, they found no correlation between self-reported degree of language dominance and translation speed, and therefore

Table 9. Mean (*M*) and standard deviation (*SD*) proportion correct naming scores on the Multilingual Naming Test (MINT) and Boston Naming Test (BNT) by self-rated ability in each language.

		MINT–English		BNT–English	
English self-rating		M	SD	M	SD
Young monolingual	Superior (n = 28)	.958	.028	.903	.053
	Advanced high (n = 2)	.941	.042	.900	.024
	Advanced low/mid (n = 6)	.939	.044	.881	.062
Young bilingual	Superior (n = 27)	.911	.043	.842	.087
	Advanced high (n = 15)	.885	.057	.770	.117
	Advanced mid (n = 5)	.821	.043	.663	.072
	Advanced low (n = 4)	.875	.050	.746	.088
Older bilinguals	Intermediate high (n = 1)	.794	n/a	.600	n/a
	Native (n = 5)	.944	.032		
	Very good (n = 5)	.932	.038		
	Good (n = 7)	.855	.063		
	Fair-to-functional (n = 3)	.652	.209		
		MINT–Spanish		BNT–Spanish	
Spanish self-rating		M	SD	M	SD
Young bilingual	Superior (n = 12)	.833	.070	.669	.156
	Advanced high (n = 10)	.737	.148	.477	.199
	Advanced mid (n = 18)	.713	.121	.501	.130
	Advanced low (n = 8)	.700	.075	.415	.125
	Intermediate high/mid (n = 4)	.540	.164	.358	.119
Older bilinguals	Native (n = 6)	.882	.051		
	Very good (n = 5)	.859	.097		
	Good (n = 7)	.702	.074		
	Fair-to-functional (n = 2)	.676	.125		

concluded that self-ratings are not reliable. They also concluded that balanced bilinguals translate more slowly because they suffer from more interference between languages than unbalanced bilinguals.

The Bilingual Dominance Scale is compelling in many ways, and it would be interesting to see if it improved on self-ratings in classifying bilinguals into dominance groups. However, the analyses presented here reveal a number of problems with the interpretations offered therein. First, the way Dunn and Fox Tree (2009) assessed self-report as a predictor did not measure if self-reported and objective classifications of language dominance match or not. Their analyses asked if dominance ratings predict translation times. The current data indicate that bilinguals can be fairly accurate in indicating which language is dominant, but are less able to assess the extent to which their knowledge of the two languages is balanced. The distinction between these is quite subtle but could nevertheless have tremendous significance in

terms of the conclusions drawn. In particular, bilinguals are certainly not completely useless at indicating which language is dominant; the data in Figure 1 suggest that bilinguals' self-ratings of degree of language dominance align quite well with those determined by proficiency interviewers. Bilinguals do not exclusively imagine themselves translating single words, or naming pictures when they provide self-ratings of proficiency. Thus, the measure used to assess accuracy of self-ratings is of critical importance.

To illustrate, the same balanced bilinguals who translated more slowly than other bilinguals at the single word level in Dunn and Fox Tree (2009), also translated with fewer hesitations (*ums* and *uhs*) and elongations than less balanced bilinguals when given a more difficult task (translation of sentences). In this second task, no analysis was reported to assess if self-ratings were correlated with translation fluency (presumably because by that point they had already abandoned self-ratings as a flawed

measure given results of analyses of the single-word task). However, a closer look at the methods and results reveals that apparently items in the single-word translation task included words with multiple translations, and more proficient bilinguals might have therefore been slower to translate because they were choosing between multiple alternative possible translations (with balanced bilinguals having “difficulty choosing the most accurate translation” (Dunn & Fox Tree, 2009 p. 282). If so, the theoretical implications of finding that bilinguals translated single words more slowly in the single-word task could have nothing to do with interference between languages, but rather with greater proficiency and a need to select WITHIN A SINGLE TARGET-LANGUAGE the best translation (an issue completely orthogonal to the possibility of between-language interference).

To conclude, bilinguals are largely pretty good at reporting which of their two languages is dominant, but the extent of difference between languages can vary with domain (and with different measures), and some bilinguals completely miss the mark thus sole reliance on self-report is not advised. Although we did not set out to compare young and older bilinguals, the data we presented also appear to be largely comparable across age-groups. In cases where bilinguals perform relatively better in the language they report is not dominant, this may occur because their level of ability is better in some domains in their otherwise less-dominant language, because the test is biased towards their non-dominant language, because dominance varies with domain (Bedore et al., 2011), or for other reasons (e.g., overestimating ability in the dominant language). In clinical settings, bilinguals who report balanced ability in both languages should be questioned and it should not be assumed that they could be tested in either language. English-dominant bilinguals can be tested in English, but should not be expected to perform like monolinguals.

Although we have focused here largely on measurement of bilingual language proficiency and the accuracy of self-report measures, it is important to consider the possibly far-reaching implications of the results reported here for developing theoretical models of bilingual language processing. There has been some focus recently on whether a non-dominant language can influence processing in a dominant language, both in research on visual word recognition (van Assche, Duyck, Hartsuiker & Diependaele, 2009; van Hell & Dijkstra, 2002), and in research on language production and verbal fluency (e.g., Costa et al., 2000; Ivanova & Costa, 2008; Sandoval, Gollan, Ferreira & Salmon, 2010). In such investigations, it would be wise to establish dominance using objective measures rather than relying on self-report. In addition, such assessment should be reported for each individual included in the analysis rather than for the group as a whole. For example, in Experiment 1 the overall

means suggest English dominance in the group as a whole; however, 40% of these bilinguals are classified as Spanish-dominant by objective measures. In looking for effects of a non-dominant language on a dominant language, it is extremely important to exclude participants who might be incorrectly self-classifying their dominance. Bilinguals with a relatively balanced profile should also be excluded from analysis to allow strong conclusions to be drawn. Similar approaches should be taken in studies that wish to distinguish between balanced and unbalanced bilinguals. Self report measures seemed to be least accurate for this type of classification. Future attempts to draw theoretical conclusions about the effects of language dominance, or balanced versus unbalanced bilingualism, should take into consideration the limitations in self-report and objective measures, and temper conclusions accordingly while also taking extra measures to ensure that misclassifications are very unlikely.

#### **Appendix. Language proficiency scale for self-ratings in Experiment 1, and by oral proficiency interviewers in Experiments 1 and 2**

- 1 = Novice Low = No real functional ability. Given lots of time and cues may be able to exchange greetings, give identity and name a number of familiar objects. Cannot participate in a true conversational exchange.
- 2 = Novice Middle = Can communicate only very minimally and with great difficulty using a number of isolated words and memorized phrases.
- 3 = Novice High = Can communicate with some success about simple topics only. Heavy reliance on memorized phrases, or on words provided by person speaking with. Speaks in short or incomplete sentences, and frequent miscommunications occur.
- 4 = Intermediate Low = Can successfully handle a limited number of uncomplicated communicative tasks by combining and recombining into short statements what they know and what the person speaking with says.
- 5 = Intermediate Middle = Can successfully handle a variety of uncomplicated communicative tasks about simple topics (food, travel, family, daily activities and personal preferences). Speaks in full sentences and even with some strings of sentences.
- 6 = Intermediate High = Can successfully handle many uncomplicated tasks and social situations requiring an exchange of basic information related to work, school, recreation, particular interests and areas of

competence. Some hesitation, errors, and gaps in communication may still occur.

- 7 = Advanced Low = Can participate actively in most informal and a limited number of formal conversations on activities related to school, home, and leisure activities and, to a lesser degree, those related to events of work, current, public, and personal interest or individual relevance. Can rarely function at the level of formal or professional language, and cannot speak at a professional level for an extended period of time.
- 8 = Advanced Middle = Can handle with ease and confidence a large number of communicative tasks such as informal and some formal exchanges on a variety of concrete topics relating to work, school, home, and leisure activities, as well as to events of current, public, and personal interest or individual relevance. Can sometimes function at a formal or professional level of language but not consistently and not with a broad range of topics.
- 9 = Advanced High = Can participate fully and effectively in conversations on a variety of topics in formal and informal settings from both concrete and abstract perspectives. Can speak at a formal or professional level of language usually without difficulty. When speaking at a formal or professional level some patterns of errors may still appear but these do not interfere with communication.
- 10 = Superior = Speaks like a highly educated native speaker. Can participate fully and effectively in conversations on a variety of topics in formal and informal settings from both concrete and abstract perspectives with accuracy and fluency using formal and professional quality language. Occasional errors may still occur but these do not interfere with communication.

## References

- Allegri, R. F., Mangone, C. A., Fernandez Villavicencio, A., Rymberg, S., Taragano, F. E., & Baumann, D. (1997). Spanish Boston Naming Test norms. *The Clinical Neuropsychologist*, *11*, 416–420.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (CD-ROM). Philadelphia, PA: Philadelphia Linguistic Data Consortium, University of Pennsylvania.
- Bedore, L. M., Peña, E. D., García, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Language, Speech, and Hearing Services in Schools*, *36*, 188–200.
- Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K., Resendiz, M. D., Greene, K., Bohman, T., & Gillam, R. B. (2011). The measure matters: Language dominance profiles across measures in Spanish/English bilingual children. Ms., University of Texas at Austin.
- Byrd, D. A., Sanchez, D., & Manly, J. J. (2005). Neuropsychological test performance among Caribbean-born and U.S.-born African American elderly: The role of age, education and reading level. *Journal of Clinical and Experimental Neuropsychology*, *27*, 1056–1069.
- Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphical system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instrumentation and Computation*, *25*, 257–271.
- Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1283–1296.
- Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain & Language*, *94*, 94–103.
- Daller, M. (2011). The measurement of bilingual proficiency: Introduction. Special Issue of *International Journal of Bilingualism*, *15*, 123–127.
- Davies, Mark. (2002). Corpus del Español (100 million words, 1200s–1900s). <http://www.corpusdelespanol.org> (retrieved January, 2011).
- Davies, Mark. (2008) The Corpus of Contemporary American English (COCA): 410+ million words, 1990–present. <http://www.americancorpus.org> (retrieved January, 2011).
- Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*, 65–70.
- Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, *37*, 665–671.
- de la Plata, C. M., Vicioso, B., Hynan, L., Evans, H. M., Diaz-Arrastia, R., Lacritz, L., & Cullum Munro, C. (2007). Development of the Texas Spanish Naming Test: A test for Spanish speakers'. *The Clinical Neuropsychologist*, *22*, 288–304.
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test – Revised*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test – Third Edition*. Bloomington, MN: Pearson Assessments.
- Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient Bilingual Dominance Scale. *Bilingualism, Language, and Cognition*, *12*, 273–289.
- Dunn, L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986). *Examiner's manual for the Test De Vocabulario en Imágenes Peabody: Adaptación Hispanoamericana*. Circle Pines, MN: American Guidance Service.
- Flege, J. E., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*, 567–598.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

- Gollan, T. H., & Acenas, L. A. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish–English and Tagalog–English bilinguals. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*, 246–269.
- Gollan, T. H., & Brown, A. S. (2006). From tip-of-the-tongue (TOT) data to theoretical implications in two steps: When more TOTs means better retrieval. *Journal of Experimental Psychology: General*, *135*, 462–483.
- Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The bilingual effect on Boston Naming Test performance. *The 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, & Cognition*, *35*, 640–665.
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*, 787–814.
- Gollan, T. H., Salmon, D. P., Montoya, R. I., & da Pena, E. (2010). Accessibility of the nondominant language in picture naming: A counterintuitive effect of dementia on bilingual language production. *Neuropsychologia*, *48*, 1356–1366.
- Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition*, *4*, 63–84.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*, 3–15.
- Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*, 131–149.
- Grosjean, F. (2008). *Studying bilinguals*. Oxford: Oxford University Press.
- Hakuta, K., & D'Andrea, D. (1992). Some properties of bilingual maintenance and loss in Mexican background high-school students. *Applied Linguistics*, *13*, 72–99.
- Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*, 277–288.
- Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test*. Philadelphia, PA: Lea & Febiger.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test – Second Edition*. Circle Pine, MN: American Guidance Service.
- 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.
- Kroll, J. F., & de Groot, A. M. B. (eds.) (2005). *The handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Li, P., Sepanski, S., & Zhao, X. (2006). Language History Questionnaire: A web-based interface for bilingual research. *Behavior Research Methods*, *38*, 202–210.
- Lim, V. P. C., Rickard Liow, S. J., Lincoln, M., Chan, Y. H., & Onslow, M. (2008). Determining language dominance in English–Mandarin bilinguals: Development of a self-report classification tool for clinical use. *Applied Psycholinguistics*, *29*, 389–412.
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940–967.
- Mather, N., & Woodcock, R. W. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities: Examiner's Manual*. Itasca, IL: Riverside Publishing Company.
- Mattis, S. (1988). *Dementia Rating Scale: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS–ADRDA Work Group. *Neurology*, *34*, 939–944.
- Paradis, M., & Libben, G. (1987). *The assessment of bilingual aphasia*. Hillsdale, NJ: Lawrence Erlbaum.
- Patricacou, A., Psallida, E., Pring, T., & Dipper, L. (2007). The Boston Naming Test in Greek: Normative data and the effects of age and education on naming. *Aphasiology*, *21*(12), 1157–1170.
- Pearson, B. Z., Fernández, S., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language and Learning*, *43*, 93–120.
- Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, *78*, 1255–1264.
- Peña-Casanova, J., Quiñones-Úbeda, S., Gramunt-Fombuena, N., Aguilar, M., Casas, L., Molinuevo, J. L., Robles, A., Rodríguez, D., Sagrario Barquero, M., Antúnez, C., Martínez-Parra, C., Frank-García, A., Fernández, M., Molano, A., Alfonso, V., Sol, J. M., Blesa, R., & the NEURONORMA Study Team (2009). Spanish Multicenter Normative Studies (NEURONORMA Project): Norms for Boston Naming Test and Token Test. *Archives of Clinical Neuropsychology*, *24*, 343–354.
- Roberts, P. M., & Deslauriers, L. (1999). Picture naming of cognate and non-cognate nouns in bilingual aphasia. *Journal of Communication Disorders*, *32*, 1–23.
- 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.
- Rosselli, M., Ardila, A., Araujo, K. Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solis, F. (2000). Verbal fluency and repetition skills in healthy older Spanish–English bilinguals. *Applied Neuropsychology*, *7*, 17–24.
- Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency: The dual-task analogy. *Bilingualism: Language and Cognition*, *13*, 231–252.

- Sebastián-Gallés, N., Martí, M. A., Cuetos, F., & Carreiras, M. (2000). *LEXESP: Léxico informatizado del español*. Barcelona: Edicions de la Universitat de Barcelona.
- Shipley, W. C. (1946). *Institute of Living Scale*. Los Angeles, CA: Western Psychological Services.
- Silverberg, S., & Samuel, A. G. (2004). The effect of age of second language acquisition on the representation and processing of second language words. *Journal of Memory & Language*, *51*, 381–398.
- Treffers-Daller, J. (2011). Operationalising and measuring language dominance. *International Journal of Bilingualism*, *15*, 147–163.
- Umbel, V. M., Pearson, B. Z., Fernández, M. C., & Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development*, *63*, 1012–1020.
- van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science*, *20*, 923–927.
- van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*, 780–789.
- Weintraub, S., Salmon, D., Mercaldo, N., Ferris, S., Graff-Radford, N. R., Chui, H., Cummings, J., DeCarli, C., Foster, N. L., Galasko, D., Peskind, E., Dietrich, W., Beekly, D. L., Kukull, W. A., & Morris, J. C. (2009). The Alzheimer's Disease Centers' Uniform Data Set (UDS): The neuropsychologic test battery. *Alzheimer Disease and Associated Disorders*, *23*, 91–101.
- Woodcock, R. W., & Muñoz-Sandoval, A. F. (1996). *Bateria Woodcock – Muñoz Pruebas de Aprovechamiento – Revisada*. Chicago, IL: Riverside.