Bilingualism: Language and Cognition 17 (2), 2014, 364–383 © Cambridge University Press 2013 doi:10.1017/S1366728913000424

Assessing language dominance in Mandarin–English bilinguals: Convergence and divergence between subjective and objective measures*

LI SHENG
The University of Texas at Austin
YING LU
The University of Texas at Austin
TAMAR H. GOLLAN
University of California San Diego

(Received: September 19, 2012; final revision received: June 26, 2013; accepted: July 5, 2013; first published online 24 September 2013)

This study examines the convergence and divergence between subjective and objective measures of language proficiency for assessing language dominance in Mandarin–English bilinguals. Sixty-two young adults (Experiment 1) and 27 children (Experiment 2) provided self-ratings of proficiency level (or were rated by their parents), were interviewed for spoken proficiency, and named pictures in the Multilingual Naming Test (MINT) and (in Experiment 1 only) the Boston Naming Test. In Experiment 1, the four measures converged in the number of people classified into different dominance groups but both naming tests indicated greater English dominance than self-report and interview measures. In Experiment 2, parent report and interview measures converged in dominance classifications but the MINT indicated higher degrees of English dominance. To a large extent bilinguals were able to classify themselves (or their children) into dominance groups but some mismatches between measures in dominance classification were observed for all age and dominance groups. These results, together with previous findings with Spanish–English bilingual adults (Gollan et al., 2012), suggest that bilinguals may shift to English dominance in confrontation naming before they do so in conversational fluency, and that dominance shifts persist throughout the lifespan but may be relatively more pronounced in children. These findings caution against the use of self-reports as the sole means of classifying bilinguals into dominance groups and support a multi-measure approach including direct assessment of the relevant linguistic domain.

Keywords:language dominance, proficiency, Mandarin, English, picture naming

Introduction

A common misconception about bilinguals is that they can (or should) be able to speak both languages equally well. The reality is usually far from this ideal, although measuring this is not simple. Even in carefully controlled experimental studies of bilingualism there are challenges in defining who qualifies as "bilingual" (and for that matter also "monolingual") (Grosjean, 1998). Two key concepts in this respect are language dominance and language proficiency. The term PROFICIENCY emphasizes variation between individuals in language abilities, and is often described with reference to monolingual norms (Bedore, Peña, Summers, Boerger, Resendiz, Greene, Bohman & Gillam, 2012; Bialystok, Luk, Peets & Yang, 2010). On the other hand, language DOMINANCE focuses on the relative proficiency of the two languages within the same individual. Cross-linguistic comparisons of a bilinguals' proficiency usually result in one language being more proficient (or dominant) than the other – the less proficient (or non-dominant) language (Gathercole & Thomas, 2009; Kohnert, 2008). To better describe participants in research studies and serve clients in clinical settings, researchers and clinicians have devoted much effort to developing assessment tools for measuring language dominance (Daller, Yıldız, de Jong, Kan & Başbağı, 2011; Dunn & Fox Tree, 2009; Flege, MacKay & Piske, 2002; Lim, Rickard Liow, Lincoln, Chan & Onslow, 2008; Treffer-Daller, 2011) and proficiency (Gutiérrez-Clellen & Kreiter, 2003; Li, Sepanski & Zhao, 2006; Marian, Blumenfeld & Kaushanskaya, 2007; Restrepo, 1998). More recently, researchers also began to examine convergence and divergence across different ways of operationalizing language proficiency and dominance in Spanish–English bilingual adults (Gollan, Weissberger, Runnqvist, Montoya & Cera, 2012) and children (Bedore et al., 2012). The current research builds on a study that introduced the Multilingual Naming Test, a picture-naming test that was designed for English, Spanish, Mandarin Chinese, and Hebrew speakers. The initial study (Gollan

* The authors thank Brooke Lauper for assistance with data collection and Arthur Samuel and Angel Chan for comments on an earlier draft. This research was supported by R01 grants (NIDCD 011492, NICHD 050287 and NICHD 051030).

Address for correspondence:
Li Sheng, Department of Communication Sciences and Disorders, The University of Texas at Austin, 2504-A Whitis Ave., Austin, TX 78705, USA
li.sheng@utexas.edu

et al., 2012) assessed college-aged and aging Spanish–English bilingual adults; the current study considers if the results will generalize to bilinguals of another language combination, specifically, Mandarin–English; and another age group, specifically, young children.

An important question in both experimental and clinical settings is to what extent bilinguals are able to identify which language they speak more proficiently. To address this question Gollan et al. (2012) conducted two experiments. In the first experiment, 52 Spanish–English bilingual young adults (ages 18–36 years) completed a language history questionnaire in which they rated their own listening, speaking, reading and writing proficiency in each language using a 10-point scale (1 = novice low, 10 = superior). The participants also completed an oral proficiency interview (OPI) in which they answered questions designed to elicit different grammatical constructions and tap into conversational fluency on a variety of topics. Based on OPI performance, each participant was assigned a proficiency score in English and Spanish by an examiner using the same 10-point scale as in self-ratings. Participants also completed two naming tests: the Multilingual Naming Test (MINT) and the Boston Naming Test (BNT, Kaplan, Goodglass & Weintraub, 1983). The MINT was developed by Gollan et al. (2012) and includes 68 black-and-white line drawings arranged in order of increasing difficulty. In comparison to the BNT, which was designed for monolingual English speakers, the MINT contains a larger number of items of medium difficulty. On average, the English picture names in the MINT are significantly shorter in length and higher in word frequency than English BNT picture names. The latter three measures (i.e., OPI, MINT, BNT) were grouped under the term objective measures whereas self-rating of speaking proficiency was considered a subjective measure. The authors contrasted subjective and objective measures to examine both convergence and divergence between these measures in assessing bilingual speakers' oral proficiency, language dominance, and degree of bilingualism.

Gollan et al. (2012) found that self-rated speaking proficiency was significantly correlated with each of the three objective measures of proficiency and these correlations were stronger for Spanish than English. To measure language dominance, the authors derived a language dominance score by subtracting the Spanish scores from the English scores. For instance, an individual who rated herself as 10 (i.e., superior) in Spanish and 9 (i.e., advanced high) in English received a dominance score of –1 (or –10% when converted to a percentage scale), indicating Spanish dominance on the self-rating measure. Young adults were fairly good at rating their language dominance, as their self-rated dominance scores correlated significantly with objective scores of language dominance. Despite these significant cross-measure correlations, the four measures diverged on the estimated degree of language dominance. The mean language dominance scores were respectively 8.8%, 9.9%, 16%, and 28.1% for self-ratings, OPI ratings, MINT scores, and BNT scores. Thus, although all four measures indicated English dominance for the participants as a group, the degree of English-dominance was relatively low according to self- and OPI-ratings (which did not differ from each other), significantly higher according to the MINT, and even larger according to the BNT.

While the previous analysis examined language dominance on a continuous scale, the next analysis placed bilinguals into three language dominance categories using a preset cut-off score. Bilinguals who showed a less than 5% between-language difference in either direction (i.e., English better than Spanish or Spanish better than English) on a particular measure were classified as BALANCED bilinguals. Bilinguals who showed a larger than 5% English advantage were classified as English-dominant, and those who showed a larger than 5% Spanish advantage were classified as Spanish-dominant. To illustrate, a person who achieved 78% accuracy in English and 74% accuracy in Spanish on the MINT would be considered a balanced bilingual according to the MINT. Using this approach, three of the measures (i.e., self-ratings, the OPI, and the MINT) did not differ significantly from each other in terms of the number of bilinguals classified into the three groups. In contrast, the BNT differed significantly from self-ratings and OPI, but not from the MINT. Compared to the other measures, the BNT yielded lower estimates of the bilinguals' Spanish proficiency level, and over-classified individuals into the English-dominant group relative to all the other measures. The MINT showed some tendency in this direction as well, but was closer to the self-ratings and the OPI than the BNT.

Further examination of the performance profiles of the three self-rated dominance groups was conducted to better understand the source of divergence in dominance classification. Results revealed that some participants tended to overestimate their Spanish proficiency while others overestimated their English proficiency. Specifically, bilinguals who rated themselves as Spanish-dominant were in fact quite balanced in their Spanish and English proficiency according to the three objective measures. Those who rated themselves as balanced turned out to be English-dominant by objective measures. Finally, bilinguals who rated themselves as English-dominant were relatively more accurate; over 90% of them were rated as English-dominant by objective measures. However, even for these individuals, English OPI ratings were significantly lower than English self-ratings, suggesting that the English-dominant bilinguals may still have overestimated their English proficiency. When averaged across the three objective measures, mismatch rate was respectively 60%, 100%, and 8.6%

for bilinguals who rated themselves as Spanish-dominant, balanced, and English-dominant.

In experimental studies on bilingualism, it is often also important to establish which bilinguals are most balanced in their proficiency levels across languages (Bialystok, Craik, Klein & Viswanathan, 2004; Gollan, Salmon, Montoya & Galasko, 2011; Paap & Greenberg, 2013; Zied, Phillipe, Karine, Valerie, Ghislaine, Arnaud & Didier, 2004). To quantify this, a BILINGUAL INDEX SCORE was calculated using the higher-scoring language as the baseline, and dividing the lower-scoring language by this baseline. For example, a bilingual who was rated as 8 (advanced middle) in English and 10 (superior) in Spanish on the OPI would be classified as "80% bilingual". Young adults were not very accurate in rating the relative proficiency of their two languages in this way; the correlations between self-rated bilingual index scores and objective index scores were small in size and ranged from marginally significant (using the OPI and MINT) to not significant (using the BNT).

Experiment 2 revealed a similar pattern of results for aging bilinguals. Twenty older Spanish–English bilinguals (ages 65–87 years) self-rated their proficiency using a simplified seven-point scale (1 = almost none, 7 = like native speaker), were interviewed with a similar OPI, and named pictures from the MINT in both languages. The BNT was not administered. The participants in the two experiments differed not only in age, but also in their proficiency profiles. Whereas the majority of young bilinguals in Experiment 1 were English-dominant, the older bilinguals in Experiment 2 included a mixture of English-dominant and Spanish-dominant bilinguals. Despite these differences in participant characteristics, results of Experiment 2 replicated those in Experiment 1 in that there were significant correlations between bilinguals' self-rated proficiency and objective proficiency measures in each language, highly robust correlations between self-ratings and objective measures of language dominance, and small and sometimes non-significant correlations between self-ratings and objective bilingual index scores. Also replicating Experiment 1, self-ratings, the OPI, and the MINT did not differ from each other in terms of the number of participants classified into the three dominance groups. In addition, although there was cross-measure agreement in absolute terms, self-report and objective classifications did not always match, and depending on the measure there were total reversals of dominance group in some cases. Nevertheless, older bilinguals in Experiment 2 fared better in estimating their language dominance than did young bilinguals in Experiment 1. In particular, the self-rated Spanish-dominant group demonstrated higher Spanish than English OPI and MINT scores, although the latter difference was not significant. This contrasted with Experiment 1 wherein the self-rated Spanish-dominant young adults overestimated their Spanish proficiency. Gollan et al. (2012) suggested that one reason why older adults might have been better at rating their dominance was that as a group they had a wider range of proficiency levels in their two languages. When one language is more clearly dominant over the other it is easier for bilinguals to classify themselves accurately. Taken together, the results of these two experiments suggest that young and older bilingual adults are capable of classifying their language dominance relatively better than their absolute proficiency in each language, and better than the extent to which their knowledge of the two languages is balanced (i.e., degree of bilingualism).

The Gollan et al. (2012) study yielded three remaining questions that await further investigation. First, the MINT, a test designed with several bilingual groups in mind, elicited higher performance from Spanish–English bilinguals than the BNT, a test designed for monolinguals. However, it is unclear whether the same pattern would hold for other bilingual groups, such as Hebrew–English or Mandarin–English speakers. Second, because the BNT was not designed for use with Spanish speakers, it was not surprising to find that BNT seemed to bias language dominance scores towards English-dominance (in Experiment 1). However, it is not clear why the MINT showed tendencies in the same direction. This could be caused by something specific to the MINT materials, the pairing of languages tested (i.e., Spanish and English), or could reflect something more general to bilingualism. For example, dominance may shift towards the language of immersion (i.e., English-dominance in the USA) in picture naming before it shifts for other aspects of language proficiency. In other words, a person who appears to be a balanced bilingual in an interview might nevertheless be English-dominant for object naming (and therefore also in a confrontation naming test). It could be that confrontation naming, or the retrieval of specific lexical representations in the absence of contextual support requires a high level of lexical activation that is more readily available for the language of immersion. To disentangle the specific effect of language pairing from more general effects of bilingualism, we recruited adult bilinguals who speak Mandarin and English and examined if these participants would demonstrate similar patterns exhibited by the Spanish–English bilinguals.

Having established the extent to which young and older adults are able to accurately classify themselves into language dominance groups, another important question is to ask to what extent caregivers can do the same. This would have relevance on both ends of the lifespan (e.g., bilinguals with Alzheimer's disease might need to rely on caregiver report), but as a starting point we began by asking if parents can reliably report the proficiency level and language dominance of their children. Previous studies indicated that parents are capable of describing their children's current vocabulary and sentence formation

skills (Jackson- Maldonado, Thal, Marchman, Newton, Fenson & Conboy, 2003; Thal, Jackson-Maldonado & Acosta, 2000) and that parent ratings of proficiency correlate with children's performance on experimenter-designed measures of linguistic knowledge (Sheng, Lu & Kan, 2011). However, previous studies have not examined the degree of convergence between parent ratings and objective measures in classifications of language dominance. To address this question, we tested Mandarin–English bilingual children and examined the extent to which their parents' ratings agreed with the children's performance on the OPI and the MINT.

To summarize, the present study aimed to determine if the MINT would elicit similar patterns of performance from speakers of Mandarin–English as previously reported for Spanish–English bilinguals. In particular we were interested in assessing (a) if Mandarin–English bilinguals were similarly accurate (or inaccurate) in classifying themselves into language dominance groups, and (b) if they would exhibit a shift in language dominance in picture naming before connected speech. If so we would find patterns of stronger English dominance in picture naming than in self-ratings and examiner ratings. Finally, we asked if (c) parent ratings of their children's language dominance would be similarly accurate to adult's self-ratings.

Experiment 1: Young adult bilinguals

Method

Participants

Sixty-two young bilinguals (ages 18–26 years) participated. Most were undergraduates at the University of California, San Diego (UCSD), and participated in exchange for course credit or monetary compensation. Participants' characteristics are presented in Table 1. Using a 5% between-language difference as cut-off, the bilinguals were classified into three groups: Mandarin-dominant bilinguals, who rated their Mandarin as more proficient than their English ($n = 14$); balanced bilinguals, who selected the same rating for each language ($n = 6$), and English-dominant bilinguals, who rated their English as more proficient than their Mandarin ($n = 42$). Participants were protected according to the guidelines of the Institutional Review Board of UCSD.

Materials and procedure

The materials and procedures were identical to those used in Gollan et al. (2012) except replacing Spanish materials with Mandarin translations. Participants completed a Language History Questionnaire, followed by an English vocabulary test (the Shipley Vocabulary Test, Shipley, 1946) and a test of nonverbal reasoning (the Matrices subtest of the Kaufman Brief Intelligence Test, Second Edition, KBIT-2; Kaufman & Kaufman, 2004). Raw Shipley and Matrices scores are presented in Table 1. Next the participants answered interview questions from the OPI, and named pictures from the Boston Naming Test (BNT, Kaplan et al., 1983) and the MINT with test order (BNT, MINT) and language of testing (English, Mandarin) counterbalanced across participants. To minimize language switching, the OPI and naming tests were administered in succession in one language, followed by the OPI and naming tests in the other language. A proficient Mandarin–English experimenter administered the entire test battery and assigned each participant a proficiency score for each language based on the OPI. The MINT was presented on a Macintosh computer with a 17-inch color monitor using the PsyScope software version 1.2.5 (Cohen, MacWhinney, Flatt & Provost, 1993). A second Mandarin–English bilingual who did not administer the tasks listened to recordings of both sessions and assigned OPI ratings for both languages. The correlation between the initial and the second ratings was significant for both English, $r = .51$, $p < .001$, and Mandarin, $r = .78$, $p < .001$. For better consistency, the experimenter's ratings were used in subsequent analyses. The average difference between the two ratings was low (less than one point on the 10-point scale) for both English (M(ean difference) $= .89$, $SD = 1.16$) and Mandarin ($M = .05$, $SD = 1.18$).

The testing protocol took about 90 minutes on average and no more than two hours to complete. Details about the language proficiency self-rating, the OPI, and the MINT can be found in Gollan et al. (2012).

Results

Table 2 presents the means and standard deviations of bilingual young adults' self-rated spoken language proficiency, the OPI ratings, and proportion correct on the MINT and the BNT as a function of language and self-rated dominance groups. Although the OPI ratings are to some extent more similar to self-ratings than to naming test scores, to be consistent with Gollan et al. (2012), we grouped the OPI, MINT and BNT under the term OBJECTIVE MEASURES because the OPI rating was objective in the sense that it did not originate from the speakers themselves. To contextualize the present findings with existing evidence of multi-measure assessment of bilingual language dominance, as we describe the results, we make comparisons between the current findings and those in Gollan et al.'s previous study of Spanish–English bilinguals.

Correlations between measures of proficiency, dominance, and degree of bilingualism

First we examined the correlations among the proficiency measures in each language. We also calculated a language

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

Table 1. *Mean (M), standard deviation (SD), and range of participant characteristics in Experiment 1.*

	Mandarin-dominant bilinguals (n = 14)			Balanced bilinguals (n = 6)			English-dominant bilinguals (n = 42)		
	M	SD	Range	M	SD	Range	M	SD	Range
Age	21.14	2.35	19–26	20.33	1.37	18–22	20.19*	1.53	18–25
% Female	0.43	N/A	N/A	0.67	N/A	N/A	0.57	N/A	N/A
Education	14.07	1.38	12–16	14.50	1.05	13–16	14.00	1.36	12–17
Age of first exposure to English	9.36	4.60	0–15	9.33^{eee}	4.32	3–13	3.80***	3.65	0–12
Age of first exposure to Mandarin	0.39	0.74	0–2.5	0.83	1.17	0–3	0.78	2.23	0–14
% Current English use	65.5	28.2	20–98	78.83^{ee}	22.54	50–98	89.86***	9.60	60–100
% English use growing up	26.4	22.1	0–70	42.50^{e}	26.03	0–75	61.10***	21.81	2–99
How often speak to bilinguals[1]	4.86	1.96	2–7	4.33	1.97	2–7	3.21***	1.47	1–7
How often speak to bilinguals growing up[1]	4.07	2.37	1–7	4.33	2.73	1–7	5.36**	1.54	2–7
How often switch languages[2]	3.43	1.45	1–5	3.00	1.10	2–4	2.50***	0.99	1–5
How often switched growing up[2]	2.43	1.28	1–5	1.83^{e}	0.98	1–3	2.93	1.33	1–5
Primary parent education level	15.71	2.20	12–20	14.33	2.66	12–18	15.62	2.87	6–20
Secondary parent education level	15.85	4.68	0–20	14.67	2.07	12–16	16.00	3.00	8–20
Shipley Vocabulary Test	24.64††	3.73	19–30	28.67	1.63	26–30	31.17***	3.59	25–39
Matrices Reasoning Subtest	41.00	2.66	34–46	37.83	8.13	22–45	39.17	3.96	29–46
Self-ratings of proficiency[3]									
English speaking	7.50	0.94	6–9	8.33^{ee}	1.37	7–10	9.26***	0.86	7–10
English listening	8.14	1.03	6–10	8.83	1.17	7–10	9.36***	0.88	7–10
English writing	7.21††	0.80	5–8	8.50	1.22	7–10	9.05***	1.15	6–10
English reading	7.57	1.02	5–9	8.50	1.38	7–10	9.14***	1.07	6–10
Mandarin speaking	9.50††	0.65	8–10	8.33^{eee}	1.37	7–10	6.98***	0.98	5–9
Mandarin listening	9.57††	0.65	8–10	8.67^{e}	1.21	7–10	7.74***	1.04	4–9
Mandarin writing	8.29††	1.64	4–10	5.33	3.33	1–10	4.11***	1.93	0–7
Mandarin reading	9.14†††	1.10	7–10	6.17	2.93	3–10	4.90***	2.55	0–9

[1] 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.

[2] 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 = a lot or sometimes even constantly.

[3] 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.

†† Significant *t*-test comparing Mandarin-dominant to balanced bilinguals ($p < .05$).

††† Significant *t*-test comparing Mandarin-dominant to balanced bilinguals ($p < .01$).

^{e} Marginally significant *t*-test comparing balanced bilinguals to English-dominant ($p < .10$).

^{ee} Significant *t*-test comparing balanced bilinguals to English-dominant ($p < .05$).

^{eee} Significant *t*-test comparing balanced bilinguals to English-dominant ($p < .01$).

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

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

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

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

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

	Self- or parent-rated spoken proficiency				Oral proficiency interview				Multilingual Naming Test (MINT) [a]				Boston Naming Test (BNT) [a]			
	English		Mandarin		English		Mandarin		English		Mandarin		English		Mandarin	
Self- or parent-rated dominance	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Young adults																
Mandarin-dominant	7.50	0.94	9.50	0.65	7.00	1.04	8.75	1.68	0.80	0.09	0.87	0.05	0.51	0.13	0.70	0.09
Balanced	8.33	1.37	8.33	1.37	8.58	0.86	8.17	1.63	0.87	0.09	0.81	0.06	0.73	0.14	0.55	0.12
English-dominant	9.26	0.86	6.98	0.98	8.58	0.82	6.63	1.50	0.92	0.05	0.62	0.18	0.83	0.10	0.33	0.19
Children																
Mandarin dominant	8.16	1.40	9.44	0.46	4.80	1.48	5.40	1.82	0.66	0.11	0.45	0.22				
Balanced	9.20	0.80	9.20	0.80	6.00	1.73	5.67	2.08	0.76	0.18	0.50	0.24				
English-dominant	9.31	0.88	7.39	1.19	5.84	1.21	4.47	1.26	0.78	0.07	0.35	0.14				

[a] MINT and BNT scores are reported as proportion correct for ease of comparison across tests given that the MINT contains 68 items and the BNT contains 60 items. Children's proficiency ratings were adjusted from a five-point scale to a 10-point scale for ease of comparison. Actual ratings for English were 4.55 (SD = 0.54) and for Mandarin were 3.88 (SD = 0.77).

dominance score and a bilingual index score for each of the four measures (self-rating, OPI, MINT, and BNT). As in the previous study, the language dominance score was derived by subtracting the Mandarin scores from the English scores, with negative scores indicating Mandarin dominance and positive scores indicating English dominance (see Figure 1 for mean dominance scores). The bilingual 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 language that produced the higher score (see Figure 2 for mean index scores). Dominance and index scores were then analyzed to examine the degree of association among the four dominance scores and among the four index scores.

Results of the correlation analyses are presented in Table 3. Self-rated proficiency measures showed statistically significant correlations with all three objective proficiency measures, ps $< .001$. To illustrate, the inter-correlations among self-rated and objective English proficiency measures ranged from .58 to .67; and the inter-correlations among self-rated and objective Mandarin proficiency measures ranged from .65 to .67. Further, the Mandarin–English bilinguals appeared to be equally good – or even better – at reporting their language dominance, with correlations among self-rated and objective dominance scores varying between .81 and. 87. Finally, bilinguals were also able to estimate the extent to which their knowledge of the two languages was balanced – or degree of bilingualism, with correlations among self-rated and objective bilingual index scores varying between .62 and .64, and these correlations were statistically significant (ps $< .001$). Finally, correlations among the three objective measures were strong and were largely comparable to the correlations among self-rated and objective measures (range: .53 to .93, all ps $< .001$). The highest correlations were found between the MINT and the BNT, which ranged between .90 and .93; these correlations indicate consistency between picture-naming assessments, but importantly do not confirm their validity for assessing bilingualism (see below and Gollan et al., 2012).

The current findings were similar to those in the previous study in that in both studies, the strength of association among the three objective measures (OPI, MINT, and BNT) was robust and in particular, the correlations between the MINT and the BNT were very high (greater than .90 in the current study and greater than .85 in the previous study). However, there were three main differences in the correlation results. First, the correlations in the current study tended to be stronger than those found in the previous study. For instance, only one of the 12 correlation coefficients (r) between self-rated and objective measures in the current study was smaller than .60 (mean r across the 12 analyses = .69), but 10 of

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

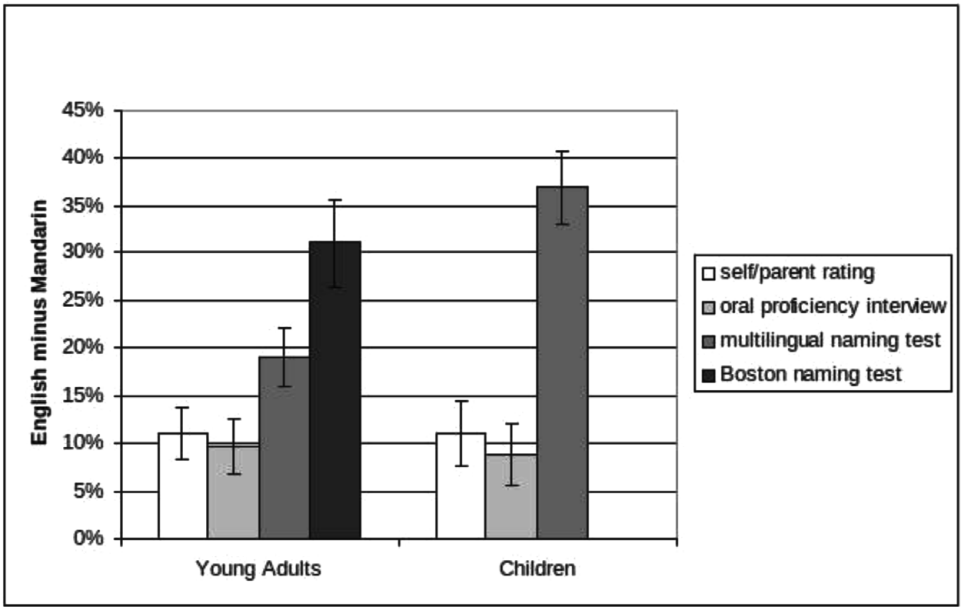

Figure 1. Average degree of English dominance for self-report (or parent report) and objective measures in Experiment 1 (young adult bilinguals) and Experiment 2 (children). Difference scores are calculated by subtracting percent adjusted Mandarin scores from English scores.

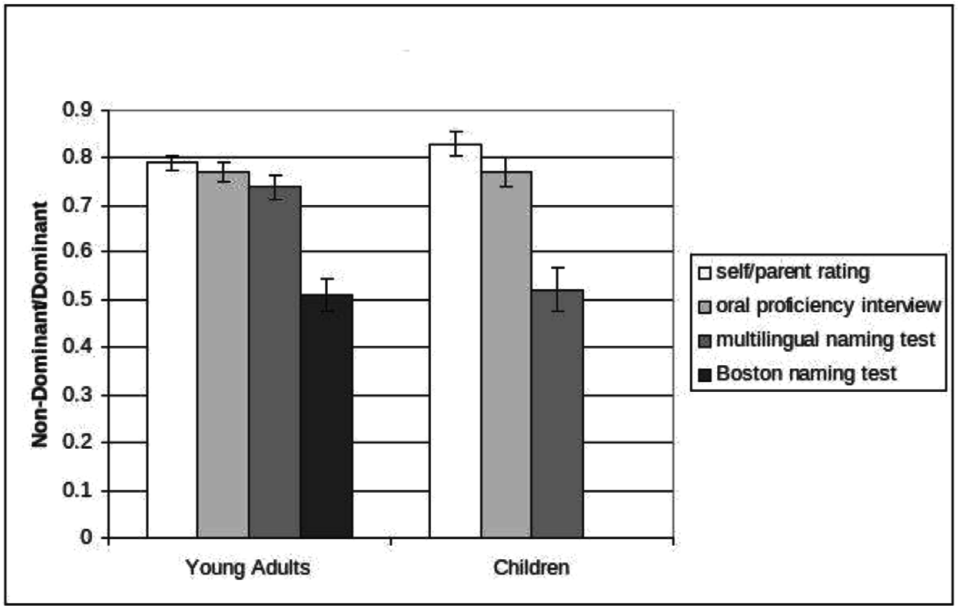

Figure 2. Average bilingual index scores for self-report (or parent report) and objective measures in Experiment 1 (young adult bilinguals) and Experiment 2 (children). Index scores reflect the extent to which proficiency in the two languages is balanced (ignoring direction of dominance), and are calculated 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.

the 12 in the previous study was smaller than .60 (mean r across the 12 analyses = .44). Second, the Spanish–English bilinguals in the previous study appeared to be unable to estimate their degree of bilingualism, which was reflected by the consistently low and non-significant correlations among self-rated and objective bilingual index scores. By contrast, our participants were fairly good at estimating the extent to which their knowledge of the two languages was balanced. Third, Mandarin–English bilinguals also produced numerically larger correlations between measures of language dominance: the mean correlation coefficient between subjective (self-rating)

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

Table 3. *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 adult bilinguals in Experiment 1 (n = 62).*

	English				Mandarin		
	Self-rating	OPI	MINT		Self-rating	OPI	MINT
OPI	0.67			OPI	0.68		
p-value	< .001			*p*-value	< .001		
MINT	0.58	0.53		MINT	0.65	0.79	
p-value	< .001	< .001		*p*-value	< .001	< .001	
BNT	0.67	0.64	0.90	BNT	0.67	0.77	0.93
p-value	< .001	< .001	< .001	*p*-value	< .001	< .001	< .001
	Language dominance (English minus Mandarin)				Bilingual index scores		
	Self-rating	OPI	MINT		Self-rating	OPI	MINT
OPI	0.81			OPI	0.62		
p-value	<.001			*p*-value	<.001		
MINT	0.83	0.83		MINT	0.64	0.65	
p-value	<.001	<.001		*p*-value	<.001	<.001	
BNT	0.87	0.85	0.93	BNT	0.62	0.60	0.91
p-value	<.001	<.001	<.001	*p*-value	<.001	<.001	<.001

and objective dominance scores was .84 in the present study and .60 in the previous study.

Young bilinguals' ability to self-report language dominance

Dominance classification into subgroups

Similar to the previous study, we allowed a between-language difference of less than 5% (in either direction) in self-ratings to classify bilinguals as balanced, and any difference of 5% or greater in either direction to be classified as dominant in one or the other language. MINT and the BNT were converted to the same 10-point scale to allow for cross-measure comparisons. It is important to note that balanced bilingualism is not a binary concept but could rather be defined along a continuum. The current cutoff, although narrow and arbitrary, does align with the observation that the vast majority of bilinguals are not truly balanced bilinguals (Genesee, Paradis & Crago, 2004; Grosjean, 1998), and also matches the standard deviation of monolingual naming scores on the MINT (see Gollan et al., 2012).

Using this method, the number of people classified as Mandarin-dominant, balanced, and English-dominant was respectively 14, 6, and 42 by self-rating; 19, 6, and 37 by OPI rating; 10, 10, and 42 by MINT scores; and 13, 3, and 46 by BNT scores. Chi-square analyses indicated that self-classifications did not differ from OPI, MINT, and BNT classifications, all χ^2s $(2, n = 62) \leq 1.67, ps \geq .43$. In addition, OPI classifications did not differ from either MINT or BNT classifications, χ^2s $(2, n = 62) \leq 4.11, ps \geq .13$. Finally, MINT and BNT classifications did not differ from each other, χ^2 $(2, n = 62) = 4.34, p = .11$. Table 4 illustrates the percentage of bilinguals in each self-rating group whose self-ratings matched objective dominance classifications.

To summarize, similar to the previous study, classifications based on self-rating, OPI, and the MINT were roughly consistent with each other. Also like the previous study, the MINT and the BNT were generally consistent with each other in dominance classification. Contrary to the previous study, in which the BNT stood out as significantly different from self-ratings, OPI, and MINT, the BNT measure achieved comparable classification as the other measures in the current study. However, as revealed in subsequent analyses, the BNT did stand out as an outlier in other aspects.

Language dominance along a continuum

As a group, the bilingual young adults obtained higher scores in English than in Mandarin on self-ratings, OPI ratings, MINT scores, and BNT scores, $ts > 3.28, df = 61$, $ps < .002$. However, the degree of this English advantage varied between measures. For self-ratings, it was by 11% ($SD = 21\%$), for OPI ratings by 9.7% ($SD = 23\%$), for MINT scores by 19.3% ($SD = 24\%$), and for BNT scores by 31.1% ($SD = 37\%$). Paired-sample *t*-tests revealed that

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

Table 4. *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.*

	Self-rated as Mandarin-dominant (n = 14)		
	Objectively Mandarin-dominant	Objectively balanced	Objective English-dominant
OPI	93%	0%	7% (35.0%)
MINT	64%	29% (0%–4.4%)	7% (7.4%)
BNT	86%	7% (–3.3%)	7% (5.0%)
	Self-rated as balanced (n = 6)		
OPI	17% (–15.0%)	33%	50% (15.0%–20.0%)
MINT	17% (–20.6%)	17%	67% (10.3%–17.6%)
BNT	17% (–26.7%)	0%	83% (16.7%–41.7%)
	Self-rated as English-dominant (n = 42)		
OPI	12% ((–10.0%)–(–5%))	10%	79%
		0%	
MINT	0%	12% ((–4.4%)–4.4%)	88%
BNT	0%	5% ((–3.3%)–1.7%)	95%

five of the six two-way comparisons of these difference scores were significant, ts > 2.56, ps < .02. The only non-significant comparison was between self-rating and OPI rating, $p = .46$. Thus, self-ratings of dominance agreed with OPI ratings, but not with naming tests and the BNT stood out most obviously in these comparisons. To further examine the continuum of dominance as a function of naming test, a 2 × 2 ANOVA with test (MINT, BNT) and language (English, Mandarin) as repeated measures, and proportion correct as the dependent variable was conducted. There were a main effect of language such that scores were higher in English than in Mandarin ($F(1,61) = 44.45$, $MSE = .088$, $p < .001$, $\eta_p^2 = .42$), a main effect of test such that scores were higher on the MINT than the BNT ($F(1,61) = 1149.03$, $MSE = .002$, $p < .001$, $\eta_p^2 = .96$), and an interaction such that the English–Mandarin performance gap was greater with BNT than with MINT ($F(1,61) = 28.62$, $MSE = .008$, $p < .001$, $\eta_p^2 = .32$).

To summarize, analyses regarding the degree of language dominance revealed similar results to the previous study. That is, relative to the other three measures, the BNT appeared to be biased towards greater English dominance, and was significantly more difficult than the MINT – but both naming tests indicated greater English dominance than the self-ratings and OPI ratings.

The source of discrepancy between subjective and objective measures of language dominance

Although the correlations between self-rated and objective dominance scores were quite high (ranging from $r = .81$ to $r = .87$), there were some discrepancies. As shown in Table 4, of the six participants who rated themselves as balanced, only two were also rated as balanced by the OPI, only one by MINT scores, and none was considered balanced by BNT scores. For this group, the mismatch rate in classification was very high (83.3%). Some of the inconsistencies were attributed to relatively lower estimates of their Mandarin proficiency by 15% to 27%, depending on the objective measure; others were due to relatively lower estimates of their English proficiency by between 10% and 42%. In comparison, the mismatch rate for the Mandarin-dominant (16.7%) and the English-dominant (12.7%) groups were much lower. A third of the discrepancies in dominance group classification (eight out of 23) resulted in a reversal of dominance category (from Mandarin-dominant to English-dominant or vice versa).

To summarize, there was better agreement across measures for the classification of Mandarin-dominant and English-dominant groups than for the balanced group. This elevated level of disconcordance should be interpreted with caution given the relatively small number of balanced bilinguals, but note that Gollan et al. (2012) reported similar findings. All of the self-rated balanced Spanish-speaking bilingual young adults in the previous study turned out to be English-dominant according to objective measures and two thirds of the self-rated balanced Mandarin-speaking bilingual young adults in the current study turned out to be English-dominant by objective measures. Other similarities and differences in results were noted between the current and the previous studies. Like the previous study, bilinguals were relatively accurate in classifying themselves as English-dominant.

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

Different from the previous study, in which the self-rated Spanish-dominant bilinguals overestimated their Spanish proficiency, the Mandarin-dominant bilinguals were rather accurate in self-rating.

Degree of balanced bilingualism

Figure 2 shows the bilingual index scores. The self-ratings, OPI, and the MINT classified bilinguals as 79%, 77%, and 74% bilingual. By contrast, the BNT yielded lower estimates of the degree of bilingualism, classifying the group as only 51% bilingual. The BNT index scores were significantly lower than all other index scores, *t*s > 9.33, $df = 61$, *p*s $< .001$. MINT index scores were lower than self index scores ($t = 2.36$, $df = 61$, $p = .02$) but comparable to OPI index scores, $p = .17$. Self-rated and OPI index scores were comparable to each other, $p = .25$. Similar to the previous study, the BNT appeared to provide lower estimates of the degree of bilingualism, whereas the MINT clustered closely with self-ratings and examiner ratings. This similarity in study results emerged even though the Mandarin–English participants appeared to be slightly less balanced than the Spanish–English bilinguals in the previous study, with mean bilingual index scores being 5% to 12% lower depending on the measure.

Summary and discussion

Experiment 1 revealed significant correlations between measures of bilingual language proficiency, between measures of language dominance, and between measures of degree of bilingualism. The young adult bilinguals were best able to predict their own dominance, and could also predict their level of proficiency in each language, and how bilingual they were. Interestingly, the four measures agreed with each other in absolute classification (i.e., the number of people assigned) into dominance groups, but not in the relative extent of English dominance. With the exception of the small group of balanced bilinguals, the Mandarin–English bilinguals were fairly good at classifying themselves into dominance groups. In the relatively few cases of classification errors, these were in part driven by the self-rated balanced bilinguals giving higher estimates of their Mandarin ability. Similar to the previous study, the BNT seemed to exaggerate the degree of English dominance (Figure 1), and to underestimate the extent of balanced knowledge of the two languages (Figure 2) relative to all the other measures. At the same time, both naming tests seemed to indicate greater English dominance than the other two measures – perhaps implying that language-dominance shifts in picture naming before it does in connected speech.

Before further considering the implications of these results, in Experiment 2 we investigated parents' ability to estimate their children's language dominance by testing a group of Mandarin–English bilingual children.

Experiment 2: Children

Method

Participants

Table 5 shows the characteristics of the 27 Mandarin–English bilingual children (ages 4;6–9;7) who participated in Experiment 2. The children resided in Austin, Texas at the time of testing. All children were typically developing according to parent reports. Two of the 27 children were raised in families where the father was a monolingual native speaker of English and the mother was a native speaker of Mandarin. These two children were exposed to both Mandarin and English from birth. All the other children were raised in families where both parents were native speakers of Mandarin or another Chinese dialect but spoke Mandarin fluently. For these children, systematic English exposure began when the child was enrolled in daycare or preschool. All the parents except for the two monolingual English speakers reported speaking Mandarin only or a mixture of Mandarin and English to their children. The child participants were protected according to the guidelines of the Institutional Review Board of the University of Texas at Austin.

Materials and procedure

There were four main differences in materials and procedure between Experiments 1 and 2. First, the children participated in a battery of examiner-designed tasks that measured vocabulary, grammar, and narrative abilities with the MINT and the OPI embedded. Second, the BNT, the Shipley vocabulary test, and the Matrices subtest were not administered to the children. Third, a different language history questionnaire (Sheng et al., 2011) was given to the primary caregiver (a parent) of each child and as a part of the questionnaire, the parent was asked to rate the proficiency of their child in the domains of vocabulary, grammar, sentence length, listening comprehension, and pronunciation using a five-point scale. For example, for English vocabulary, parents were asked "How much English vocabulary does your child use from the words she/he learns at home (e.g., food, clothing) or school (e.g., science terms)?" and had to choose from five options: 1 = a few words, 2 = a limited range of words, 3 = some words, 4 = many words, and 5 = extensive vocabulary. Ratings in different domains were averaged for each child to obtain an overall proficiency rating for each language. Children were assigned to dominance groups based on the overall proficiency rating.

Fourth, the OPI interview was modified to include questions that were appropriate for children (see Appendix). We used 22 interview questions in each language. These questions were related to various aspects of children's daily life (e.g., family members, food, school, animals, play, and holidays) and were designed to elicit

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

Table 5. *Mean (M), standard deviation (SD), and range of participant characteristics in Experiment 2.*

	Mandarin-dominant bilinguals (n = 5)			Balanced bilinguals (n = 3)			English-dominant bilinguals (n = 19)		
	M	SD	Range	M	SD	Range	M	SD	Range
Age (in months)	68.80	18.01	57–99	86.33	28.36	54–107	85.89*	17.87	57–115
% Female	0.40	N/A	N/A	0.33	N/A	N/A	0.53	N/A	N/A
Maternal Education	17.20†	2.68	13–19	21.00ee	1.73	19–22	16.84	2.83	12–22
Age of first exposure to English (in months)	28.80	6.57	24–36	22.33	12.66	11–36	24.32	12.48	0–36
% Current English input	0.56	0.10	0.48–0.72	0.59e	0.13	0.45–0.7	0.69**	0.09	0.5–0.8
% Current English output	0.62	0.14	0.48–0.78	0.64ee	0.16	0.45–0.7	0.81***	0.12	0.62–1
Parent ratings of proficiency									
English vocabulary	4.00	0.71	3–5	4.33	0.58	4–5	4.47	0.70	3–5
English listening comprehension	4.00	0.71	3–5	4.67	0.58	4–5	4.63**	0.50	4–5
English grammar	3.80	1.30	2–5	4.33	0.58	4–5	4.53	0.77	2–5
English sentence length	4.60	0.89	3–5	5.00	0.00	5–5	5.00**	0.00	5
English pronunciation	4.00	0.71	3–5	4.67	0.58	4–5	4.63*	0.68	3–5
Mandarin vocabulary	4.20	0.45	4–5	4.33e	0.58	4–5	3.11**	1.05	1–5
Mandarin listening comprehension	4.80	0.45	4–5	4.67e	0.58	4–5	3.79***	0.71	3–5
Mandarin grammar	4.80	0.45	4–5	4.33e	0.58	4–5	3.37***	0.83	2–5
Mandarin sentence length	5.00	0.00	5–5	5.00	0.00	5	3.95**	1.08	1–5
Mandarin pronunciation	4.80	0.45	4–5	4.67	0.58	4–5	4.26	0.73	3–5

[e]Marginally significant *t*-test comparing balanced bilinguals to English-dominant ($p < .10$).
[ee]Significant *t*-test comparing balanced bilinguals to English-dominant ($p < .05$).
†Marginally significant *t*-test comparing Mandarin-dominant to balanced bilinguals ($p < .10$).
*Marginally significant *t*-test comparing Mandarin-dominant to English-dominant ($p < .10$).
**Significant *t*-test comparing Mandarin-dominant to English-dominant ($p < .05$).
***Significant *t*-test comparing Mandarin-dominant to English-dominant ($p < .01$).

different grammatical structures and constructions. For example, some questions require answers with present tense (e.g., "What do you like to do after school?"), past tense (e.g., "Who did you play with yesterday after school and what did you do together?") or future tense (e.g., "What will you do if your mom doesn't allow you to have as much chocolate as you want?"). Some questions prompted the children to describe an object (e.g., "What is your favorite toy? Tell me something about your favorite toy."), narrate an event (e.g., "When is your birthday? What do you do on your birthday?"), or provide reasoning (e.g., "Which holiday do you like best? Why?"). These questions were divided into several blocks and were asked in between the battery of experimental tasks. Given these changes in the interview protocol, we also modified the instruction of the OPI rating scale accordingly. Specifically, we took out wordings referring to the use of professional language and events of work, current, and public relevance. We added descriptions that prompt the examiner to judge children's ability to tell coherent personal stories and use complex sentences and conjunctions. Finally, the rating scale ranged from one (novice low) to eight (advanced middle) to reflect the limited range of topics assessed.

Children were tested on two different days with the language of testing counterbalanced. A monolingual speaker of English conducted the English OPI and the English MINT; a bilingual experimenter who is a native speaker of Mandarin administered the Mandarin OPI and MINT. The examiners later listened to recordings of the OPI and assigned ratings to each child. A second Mandarin–English bilingual, who did not administer the tasks, listened to recordings of both sessions and assigned OPI ratings for both languages. The correlation between the initial and the second ratings was significant for both English, $r = .86$, $p < .001$, and Mandarin, $r = .78$, $p < .001$. For better consistency, the second bilingual's ratings were used in subsequent analyses. The average difference between the initial and the final ratings was low for both English ($M = .19$, $SD = .82$) and Mandarin ($M = –.19$, $SD = 1.00$).

Results

Correlations between measures of proficiency, dominance, and degree of bilingualism

Before presenting the results, it is worth noting that the parent ratings were higher than OPI ratings for both English and Mandarin, ts > 11.40, ps $< .001$ (see Table 2 for means). The parent rating was originally on a five-point scale but was converted to a 10-point scale. These differences may reflect content differences between the parent and OPI rating scales. The parent questionnaire prompted parents to compare their own children to other children they know as a point of reference. In contrast, the OPI rating scale describes absolute proficiency levels with ideal anchor points ranging from lowest to highest possible proficiency level. Only one child achieved the highest possible rating (eight) in either language on the OPI (range = 3–8 in both languages), however, parents often rated their children as 5 on the five-point scale.

The correlations between the three proficiency measures (parent rating, OPI rating, and MINT), language dominance scores, and bilingual index scores are presented in Table 6. For English, parent ratings of proficiency were correlated with children's MINT scores but not with the OPI ratings. For Mandarin, parent ratings of proficiency were correlated with both MINT scores and the OPI. As in Experiment 1, language dominance scores were highly correlated across measures. Correlations between parent ratings and objective measures of dominance (OPI ratings, MINT scores) were respectively .71 and .72, ps $< .001$. As for bilingual index scores, parent ratings of children's degree of balanced bilingualism showed significant correlation with examiner ratings; the correlation between parent rating and MINT scores approached significance, $p = .06$. Finally, three of the four correlations between the two objective measures were significant; the only non-significant correlation between OPI ratings and MINT scores was for bilingual index scores.

To summarize, similar to self-ratings in Experiment 1, we found that using a simplified rating scale, parents were better able to predict their children's language dominance than proficiency level or degree of balance. Parents were also better able to predict their children's proficiency in Mandarin than English. Furthermore, parents and examiner agreed with each other fairly well on the children's degree of balanced bilingualism. It is also notable that the correlational coefficients between subjective and objective measures were generally lower in Experiment 2 (overall mean $r = .53$) than Experiment 1 (overall mean $r = .69$).

Parents' ability to report their children's language dominance

Dominance classification into subgroups

Using the same measure-anchored cut-off scores as in Experiment 1, the number of children classified as Mandarin-dominant, balanced, and English-dominant was respectively 5, 3, and 19 by self-rating; 6, 5, and 16 by OPI rating; and 0, 2, and 25 by MINT scores. Chi-square analyses indicated that parent classification did not differ from the OPI, $p = .65$. However, classification by MINT differed significantly from both parent and OPI classifications, χ^2s $(2, n = 27) \geq 6.02$, ps $\leq .05$. These results differed from Experiment 1, in which the four measures were not significantly different from each other in absolute classification of young adults into dominance subgroups.

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

Table 6. *Pearson bivariate correlations between parent-rated and objective measures of proficiency in each language, language dominance difference scores, and bilingual index scores for bilingual children in Experiment 2 (n = 27).*

	English			Mandarin	
	Self-rating	OPI		Self-rating	OPI
OPI	.31		OPI	.48	
p-value	.11		*p*-value	.01	
MINT	.40	.56	MINT	.48	.72
p-value	.04	.002	*p*-value	.01	< .001
	Language dominance (English minus Mandarin)			Bilingual index scores	
	Rating	OPI		Rating	OPI
OPI	.71		OPI	.73	
p-value	< .001		*p*-value	< .001	
MINT	.72	.62	MINT	.37	.31
p-value	< .001	< .001	*p*-value	= .06	= .11

Language dominance along a continuum

As a group, the bilingual children had stronger English than Mandarin skills based on parent ratings *t*s ≥2.76, *df* = 26, *p*s ≤ .01. However, the degree of this English advantage varied by measure. For parent ratings, it was by 11% (*SD* = 17%), for OPI ratings by 8.9% (*SD* = 17%), and for MINT scores by 37% (*SD* = 20%). As in Experiment 1, paired-sample *t*-tests indicated that parent ratings and OPI ratings did not differ from each other, *p* = .38; but MINT scores differed significantly from both rating scores, *t*s > 8.99, *df* = 26, *p*s < .001. Thus, self-ratings of dominance agreed with OPI ratings, but not with MINT. Whereas the two rating scales indicated a similar degree of English dominance in these bilingual children as in the young adults, the MINT implied a much stronger degree of English dominance in Experiment 2 like the BNT did for the adults in Experiment 1 (see Figure 1).

The source of discrepancy between subjective and objective measures of language dominance

Although the correlations between parent-rated and objective dominance scores were quite high, there were some discrepancies. Table 7 shows the percentage of children in each subgroup (Mandarin-dominant, balanced, and English-dominant) whose parent ratings showed matches and mismatches with objective dominance classifications (see Table 2 for actual ratings and accuracy scores). Of the five children who were rated by their parents as Mandarin-dominant, two were rated as balanced and one was rated as English-dominant by the OPI. Based on MINT scores, two of these five children were classified as balanced and three were English-dominant. Among the three children who were classified by their parents as balanced, one scored 10% higher in English than Mandarin on the OPI and all three scored higher in English than Mandarin on the MINT. Finally, in the 19 English-dominant children, four received higher OPI ratings in Mandarin than English and one received the same OPI ratings in both languages. The match between parent ratings and MINT scores were better as the MINT classified all 19 children as English-dominant.

To summarize, the mismatch rate for classifying the child participants in Experiment 2 was comparable to the rate for young adults in Experiment 1 for the balanced and English dominant subgroups but much higher for the Mandarin-dominant subgroup.

Degree of balanced bilingualism

Figure 2 above illustrates the index score means. The parent ratings, OPI ratings, and the MINT respectively classified children as 83% (*SD* = 13%), 77% (*SD* = 16%), and 52% (*SD* = 24%) bilingual. Paired-sample t-tests revealed significant differences between all three measures, *t*s > 3.00, *df* = 26, *p*s < .006. Thus, unlike Experiment 1, in which the MINT clustered closely with self- and examiner ratings, the MINT provided lower estimates of the children's degree of bilingualism relative to the other measures.

General discussion

The results of the current study agree in several key ways with Gollan et al.'s (2012) conclusion that bilinguals' ratings of their own language proficiency (and parent ratings) are best for determining language dominance, and are relatively less accurate in determining absolute proficiency level, the extent to which one language is dominant over the other, and degree of balanced

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

Table 7. *Percentage of bilingual children in Experiment 2 whose parent-rated language dominance matched or differed from objective measures of dominance. For cases in which parent ratings and objective classifications of dominance do not match, the range of discrepancy in scores is indicated in parentheses.*

	Rated by parents as Mandarin-dominant (n = 5)		
	Objectively Mandarin-dominant	Objectively balanced	Objective English-dominant
OPI	40%	40% (0%)	20% (10%)
MINT	0%	40% ((–1.5%) – 4.4%))	60% (26%–38%)
	Rated by Parents as Balanced(n = 3)		
OPI	0%	67%	33% (10%)
MINT	0%	0%	100% (7%–60%)
	Rated by Parents as English-dominant(n = 19)		
OPI	21% (–10.0%)	5% (0%)	74%
MINT	0%	0%	100%

bilingualism (the bilingual index). The greater apparent utility of self-ratings for determining language dominance than absolute proficiency of the two languages likely reflects the fact that dominance ratings only require bilinguals to indicate which of their two languages is more proficient than the other (and bilinguals have ready access to considerable information along these lines). In contrast, absolute proficiency ratings require greater degrees of precision in assessment and require bilinguals to compare themselves to other bilinguals (and bilinguals have less information about other bilinguals than they do about themselves).

Summarizing the results a bit more specifically, Mandarin–English bilinguals, both college-age (Experiment 1) and young children (Experiment 2) demonstrated significant correlations between self-ratings (or parent ratings) of language proficiency and objective measures of language proficiency, language dominance, and degree of balanced bilingualism. In young adults these ranged from moderate to large in size and were statistically quite robust. Parent ratings of their children's proficiency level and objective measures tended to be less strongly correlated, and in one case was not significant, but patterned similarly to young adults' self-ratings in some aspects. For example, in both experiments, correlations between language-dominance measures were highest, and measures of proficiency level in the dominant language were least correlated. Importantly, although both studies reported robust correlations between measures, classifications were far from perfect, and cases of divergence also occurred in both, implying that objective measures are needed when assessing bilinguals.

Furthermore, both the current study and the previous study indicated relatively low cross-measure convergence for the bilingual index score. Different approaches have been utilized to define balanced bilinguals ranging from self-ratings (Marian et al., 2007; Paap & Greenberg, 2013), naming performance on the BNT (Gollan et al., 2011; Zied et al., 2004), and self-reported daily use of the two languages (Bialystok et al., 2004). The current result underscores the notion that perfectly balanced bilinguals may be more of a fantasy than reality and that for studies in which it is critical to have relatively balanced bilinguals, self-ratings should not be used to classify bilinguals as such (Dunn & Fox Tree, 2009).

Self-ratings of proficiency in young adult bilinguals of different language pairs

With regard to the young adult bilinguals, there were three main differences between the current findings and those in Gollan et al.'s previous study. First, the strength of correlation between self-ratings and objective measures was higher in the current (mean $r = .69$) than in the previous study (mean $r = .44$). In fact, the overall strength of correlation for the present Mandarin–English young bilinguals was comparable to that of the older Spanish–English bilinguals (mean $r = .71$) in Experiment 2 of Gollan et al. (2012). Second, the self-classifications of language-dominance by Mandarin-dominant young adult bilinguals in the current study agreed more with objective measures (average mismatch rate with object measures = 16.7%) than both the Spanish-dominant young adult bilinguals (mismatch rate = 60%) and older adult bilinguals (mismatch rate = 35%) in the previous study. Third, in the current study, Mandarin–English bilingual young adults also exhibited significant correlations between measures for degree of balanced bilingualism, whereas in Gollan et al. these correlations were not significant. Thus, in at least some respects, namely, in terms of agreement with objective measures used in the two studies, Mandarin–English bilinguals were

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

relatively more accurate in their ability to rate their own proficiency levels than Spanish–English bilinguals.

Gollan et al. (2012) suggested that the reason the older adults fared better than young adults in self-rating was because the older adults had a wider range of proficiency. However, this did not seem to explain the current finding as the Mandarin–English young adults demonstrated similar proficiency ranges (self-ratings ranged from 5–10 for Mandarin and 6–10 for English) as the Spanish–English young adults (5–10 for Spanish and 6–10 for English) in the previous study. Nevertheless, several differences in participant characteristics may explain the better cross-measure convergence in the current study. First, as noted earlier, participants in the current study appeared to be less balanced in their bilingual proficiency than the young adults in the previous study. Specifically, the bilingual index score for the current participants was respectively 5%, 11%, 6%, and 12% lower by self-rating, OPI rating, the MINT, and the BNT. Accurate self-assessment of language dominance should be easier to achieve when there is clearer separation between the dominant and non-dominant languages. Second, the current participants also had a later age of first exposure to English (M = 5.60, range = 0–15) than the young adults in the previous study (M = 3.64, range = 0–10). Delayed onset of English exposure again could have led to clearer separation between the bilinguals' two languages. Last but not least, differences in linguistic structures could have also attributed to better separation and enhanced self-evaluation. Mandarin and English are structurally very different and these differences are easily noticeable even by naïve speakers at the phonological, prosodic, and orthographic levels. By contrast, Spanish and English are both alphabetic languages and share many cognates, words that are similar in form and meaning (Marinova-Todd & Uchikoshi, 2011). It is likely that structural distance could affect the accuracy and ease of self-ratings of language proficiency and dominance.

Despite these cross-group differences, we noted that the young adults in Gollan et al. (2012) and the current study were quite similar in MINT performance. A comparison between the current Table 2 and Gollan et al.'s Table 3 showed that differences in accuracy level between the Mandarin-speaking groups and the Spanish-speaking groups on the MINT were quite small, varying from 1% to 7% depending on language and subgroup. By contrast, the BNT yielded greater performance differences that ranged from 10% to 18% between the Spanish–English and Mandarin–English bilingual groups. It is interesting to note that in all these cases, the BNT favored the Spanish-speaking groups over the Mandarin-speaking groups. In addition, cross-group comparisons on the OPI revealed similar patterns with the Spanish-speaking groups receiving higher ratings in both languages than their Mandarin-speaking counterparts. The fact that the BNT and the OPI showed the same trends suggests that there may be some real differences in proficiency level between the Spanish–English and Mandarin–English bilinguals. Nevertheless, this pattern of results may also suggest that as inappropriate as the BNT is for assessing Spanish, it may be even less valid as a measure of Mandarin Chinese proficiency. In contrast, the MINT, which elicited the most comparable cross-group performance, may be culturally and linguistically more appropriate for both groups.

In relation to our question "Do bilinguals shift dominance in picture naming before they do so in connected speech?", the current study and the previous study revealed strikingly similar patterns in that picture naming tests – especially the BNT but true for both naming tests – classified bilinguals as relatively more English-dominant than the other measures. Also, in both studies, self-ratings and OPI ratings agreed with each other but not with scores from the two naming tests; both naming tests indicated greater English dominance than self-report and interview measures. Together, findings from two groups of bilingual young adults who speak structurally different first languages and have different background characteristics suggest that bilinguals may indeed shift towards English dominance in picture naming before they do so in connected speech. Undoubtedly, rating scales and naming tasks focus on very different set of skills. When rating themselves or others, the rater likely considers many different aspects such as vocabulary, grammar, comprehension, accent, and fluency. Picture naming, on the other hand, is a more circumscribed skill that may be modifiable within a relatively shorter period of time than linguistic competence in other domains. To illustrate, one could imagine someone learning to name 1,000 pictures in Mandarin without "being bilingual" by most definitions of what it means to be bilingual. But the reverse is not true – building conversational fluency in a second language entails grammar learning and semantic network formation and takes a protracted period of time. Fluent conversational ability on a large number of topics would certainly qualify a person as "bilingual" by many standards.

Comparing bilingual children and bilingual young adults

In comparison to Experiment 1, Experiment 2 revealed cross-measure correlations that were weaker in strength. Moreover, mismatch rates between self-ratings and objective measures for bilingual children in Experiment 2 and young adults in Experiment 1 seemed about comparable for balanced bilinguals (67% for children and 83.3% for adults), and for English-dominant bilinguals (13% for children and 12.7% for adults), but were higher for Mandarin-dominant children than for young adult

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

bilinguals (80% versus 16.7%). Cases of mismatches in classification were sometimes due to parents indicating greater levels of Mandarin proficiency in their children than indicated by objective measures, and other times due to lower parent ratings of the children's Mandarin proficiency than objective measures. It is not surprising to find that parent ratings of their children's proficiency level corroborated less with objective measures than did self-ratings in young adult bilinguals. This could be due to the fact that bilingual children have distributed language use in different contexts and parents may not have the opportunity to observe their children's language skills in all of these contexts (Bedore et al., 2012; Gutiérrez-Clellen & Kreiter, 2003; Kohnert, 2010). In addition, language proficiency is under dynamic change in young children and parents may be using quite different reference points when making judgments about their children. As noted earlier, the parent-rating scale differed from the OPI rating scale in that the parents were prompted to use other children they were familiar with as reference whereas young adults used the same anchor point descriptions of proficiency as OPI examiners. Also, the examiners not only had more experience rating language performance but also had a better comparison set than the parents. These differences in rating scales and raters' experience could have contributed to the lowered cross-measure correlations for the child bilinguals.

Another important difference in results between Experiments 1 and 2 was that while the MINT clustered quite closely with the rating scales in Experiment 1, it functioned much like the BNT in Experiment 2 and classified the children as relatively more English-dominant than rating measures (see Figures 1 and 2). This finding provides additional evidence that bilinguals may shift towards English dominance earlier in picture naming than in connected speech and could also suggest that this tendency may be especially strong in young children who are just beginning to learn the language dominant to the culture. In support of this view, Kohnert, Kan and Conboy (2010) reported that typically developing preschoolers learning Hmong (L1) and English (L2) used longer utterances and a greater diversity of words to retell stories in Hmong than English. However, on receptive vocabulary tests in which these same children were asked to point to pictures of named objects, English performance was comparable to performance in Hmong. Here the Hmong–English preschoolers may be perceived as relatively balanced in a single-word receptive vocabulary task but L1-dominant when considering their discourse abilities. In a similar vein, Golberg, Paradis and Crago (2008) found that children were able to meet native-speaker expectations on a single-word English receptive vocabulary test after an average of 34 months of English exposure. But the same children also displayed overuse of general-all-purpose verbs (e.g., *do*) in language samples, indicating that they had to stretch their lexical resources in more demanding communicative context. As argued earlier, single-word vocabulary tasks place very different demands than more integrative measures such as story retell, conversational samples, and rating scales. Bilingual children who are in the early stages of English learning may be more focused on gaining new vocabulary and may have lots of opportunities in their school environment to practice confrontation naming or picture identification, which may have attributed to a heightened degree of English dominance on single-word retrieval tasks. The presence of a shift towards English-dominance in picture naming for both children and adults supports the notion that this result is a reflection of gradual shift towards English-dominance (rather than some idiosyncratic aspect of picture-naming, or the particular pictures in the naming test). The similarity in results obtained with young children, young adults, and also older adults, demonstrates language dominance as an ever-changing aspect of bilingual performance that continues to shift in qualitatively similar ways across the life span.

These findings have methodological, clinical, and theoretical implications. Methodologically, the results illustrate how different measures can lead to different classifications of bilinguals into dominance groups (Bedore et al., 2012, MacSwan & Rolstad, 2006; Pray, 2005). For instance, Bedore et al. (2012) demonstrated that the concordance between semantic and morphosyntatic measures in classifying children into one of five dominance groups was 49%. Bedore et al also suggested that children will appear to have switched from heritage language-dominance to English-dominance earlier when the tests load heavily on semantically-based items. The same children might appear to be rather balanced or even dominant in the heritage language when the tests load on morphosyntax. Thus, care is needed when selecting objective proficiency measures that are most relevant to the targeted knowledge domain. Researchers and clinicians should also be cognizant of the underlying reasons of divergent patterns of classification and be mindful of the content and format of proficiency measures when interpreting test results. That picture naming patterns more similarly with semantic than with syntactic measures in the Bedore et al. study strengthens the conclusions that the first locus of dominance shift is in connecting meaning with lexical representations, and that syntax and linking together strings of lexical representations into grammatical utterances shifts later.

This order of dominance shift has interesting parallels to the literature on both first language acquisition and adult second language acquisition. The literature on first language acquisition suggests that vocabulary learning is less constrained by a biologically determined sensitive period than syntactic learning (Curtiss, 1977; Neville,

Mills & Lawson, 1992) and that lexical development is more susceptible to environmental influences than grammatical learning (Hoff-Ginsberg, 1998; Zhang, Jin, Shen, Zhang & Hoff, 2008). Studies of adult second language acquisition also attest to the relatively greater plasticity of word learning over grammar learning (Birdsong, 2006; McLaughlin, Osterhout & Kim, 2004; Ullman, 2001). It is not unusual to encounter fluent L2 speakers who speak rapidly in their second language but with many grammatical errors such as in gender agreement or tense marking. The greater difficulty of grammatical versus lexical learning is in some ways arguably surprising given that speakers must master a relatively limited number of syntactic structures when compared with the number of individual words that must be learned, and from this perspective a priori one might have predicted that syntax should shift dominance first. These parallels across areas of study (late L1 and L2 learners and early bilinguals) imply that important insights about bilingualism, and about the mechanisms that underlie proficient language production in all speakers (bilinguals and monolinguals alike), could be gained by investigating more specifically why grammar is harder to learn than individual words. That said, it would be important to first confirm this conclusion with additional data. For example, confrontation naming was our only measure of lexical knowledge, and different results might be found with a different measure (e.g., lexical diversity in discourse; see Kohnert et al., 2010). Additionally, grammatical errors in speech of fluent but late learners of an L2 might simply be more apparent than difficulties with lexical access, which can be avoided in overt speech by circumlocution.

To conclude, the present investigation, together with Gollan et al. (2012), demonstrate both the validity of, and limitations in, self-ratings and parent ratings of language proficiency, and suggest that the MINT may be a culturally and linguistically appropriate tool for various groups of adult bilinguals. The two studies also highlight the limitations and pitfalls of existing measures in assessing proficiency, dominance, and balanced bilingualism. Our direct measures of language proficiency (OPI, MINT, BNT) took the format of an interview to tap into global spoken proficiency (in the tradition of assessing second language learners) and confrontation naming tasks that assess expressive vocabulary. Both formats are widely used in the assessment literature but there are many other alternatives that meet the needs of researchers who are interested in domains such as phonological memory (Windsor, Kohnert, Lobitz & Pham, 2010), semantic depth (Sheng, Bedore, Peña & Fiestas, 2013), syntactic ability (Birdsong, 2006), and narrative skills (To, Stokes, Cheung & T'sou, 2010). Future studies may explore the utility of these other measures in assessing bilingual proficiency and dominance and to further test the observation that dominance shifts more rapidly in some language domains than in others. To elucidate the mechanisms that drive the shift toward dominance in the language of immersion, future studies may also test bilinguals with different kinds of immersion experience (Spanish–English bilinguals in Spain or Mandarin–English bilinguals in China). More practically, very much in agreement with other recently published work in the field (e.g., Bedore et al., 2012) the findings we reported caution against simple classifications of bilinguals into dominance groups given that this dimension can vary with assessment measure. Multi-measure assessment is needed for classifying bilinguals and particularly bilingual children, who vary even more from measure to measure and who can't provide self-ratings.

Appendix. Children's version of the oral proficiency interview and modified rating scale

1. What is your name? Do you have a nickname? Why do people call you ___ (nickname)?
2. How many people are there in your family? Who are they?
3. Do you like going to school? Why or why not?
4. What do you like to do after school?
5. Who did you play with yesterday after school and what did you do together?
6. What is your favorite toy? Tell me something about your favorite toy.
7. What is your favorite game? Tell me something about your favorite game.
8. What kind of food do you like to eat? (Or what is your favorite food?) What kind of food do you dislike? (Or: What is your least favorite food?)
9. Do you like chocolates? Can you eat a lot of chocolates? What will you do (what will you say to your mom) if your mom doesn't allow you to have as much chocolates as you want? [If the child doesn't like chocolate, replace chocolate with the name of the child's favorite food.]
10. What did you eat for breakfast today?
11. When is your birthday? What do you do on your birthday?
12. What was your birthday wish?
13. Tell me a story you like best.
14. What animal do you like best?
15. What is your favorite color? Why do you like _____ (color)?
16. What is your day like? (When do you get up? What do you do all day? When do you go to sleep? Etc.)

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

17. Which book do you often read? Tell me something about it.
18. Do you have a pet? Tell me something about your pet.
19. What are you going to do tomorrow? Where will you go?
20. Which holiday do you like best? Why?
21. What will you be for Halloween this year?/What costume will you wear for Halloween this year?
22. Do you have a lot of friends? Can you tell me something about your best friend?

1. 你叫什么名字? 你有小名么? (你妈妈叫你什么呢?)为什么叫这个名字?
2. 你家里一共有几个人? 告诉我他们是谁?
3. 你喜欢上学么? 为什么喜欢/不喜欢?
4. 你放学以后喜欢做什么事情?
5. 你昨天放学以后和谁一起玩? 你们玩了什么?
6. 你最喜欢的玩具是什么? 告诉我这个玩具是什么样的?
7. 你最喜欢的游戏是什么? 告诉我这个游戏是什么样的?
8. 你喜欢吃什么? 你不喜欢(讨厌)吃什么?
9. 你喜欢巧克力么? 你能吃很多巧克力么? 如果你想吃巧克力的时候你妈妈不让你吃, 你怎么办呢? (你会做什么呢?)
10. 你今天早餐吃什么了?
11. 你什么时候过生日? 你生日的时候做些什么事情呢?
12. 你上次的生日愿望是什么?
13. 讲一个你最喜欢的故事给我听。
14. 你最喜欢什么动物?
15. 你最喜欢什么颜色? 为什么喜欢这种颜色呢?
16. 告诉我你的一天是怎么度过的好么? (什么时候起床。做什么事情。什么时候睡觉。)
17. 你最经常看哪本书? 告诉我这本书讲了什么。
18. 你有宠物么? 告诉我关于它的一些事情。
19. 你明天有什么要做的事情么? (你明天要做什么?) 会去什么地方么?
20. 你最喜欢过什么节日? 为什么?
21. 你这个万圣节要打扮成什么?
22. 你有很多朋友么? 你最好的朋友是谁?

Modified OPI rating scale

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 conversational 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 conversational tasks about simple topics (food, family, daily activities and personal preferences). Speaks mostly in full sentences but rarely uses conjoined or complex sentences. Grammatical errors are still common. Ability to produce a monologue (narrative) without listener support is still limited.

6 = Intermediate High = Can successfully handle many uncomplicated conversational tasks and social situations requiring an exchange of basic information related to school, recreation, and particular interests. Some hesitation, errors, and gaps in communication may still occur. Can tell short personal stories although stories may be incomplete or incoherent.

7 = Advanced Low = Can participate actively in most conversations on activities related to school, home, and leisure activities. Can tell complete and coherent stories of a personal nature with few errors. Uses many complex sentences.

8 = Advanced Middle = Can handle with ease and confidence a large number of communicative tasks such conversational exchanges on a variety of concrete topics relating to school, home, and leisure activities, as well as to narrate stories of both a personal and fictional nature. Uses diverse vocabularies and sentence structures including proper use of conjunctions and complex sentences.

References

Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K., Resendiz, M. D., Greene, K., Bohman, T., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism: Language and Cognition, 15,* 616–629.

https://doi.org/10.1017/S1366728913000424 Published online by Cambridge University Press

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19,* 290–303.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13,* 525–531.

Birdsong, D. (2006). Age and second language acquisition. *Language Learning, 56 (Supplement 1),* 9–49.

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.

Curtiss, S. (1977). *Genie: A psycholinguistic study of a modern-day "wild child"*. New York: Academic Press.

Daller, M. H., Yıldız, C., de Jong, N. H., Kan, S., & Başbaĝi, R. (2011). Language dominance in Turkish–German bilinguals: Methodological aspects of measurements in structurally different languages. *International Journal of Bilingualism, 15,* 215–236.

Dunn, A., & Fox Tree, J. (2009). A quick and gradient Bilingual Dominance Scale. *Bilingualism: Language and Cognition, 12,* 273–289.

Flege, J., McKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics, 23,* 567–598.

Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition, 12,* 213–237.

Genesee, F., Paradis, J., & Crago, M. B. (2004). *Dual language development and disorders*. Baltimore, MD: Paul H. Brookes.

Golberg, H., Paradis, J., & Crago, M. B. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics, 29,* 41–65.

Gollan, T. H., Salmon, D. P., Montoya, R. I., & Galasko, D. R. (2011). Degree of bilingualism predicts age of diagnosis of Alzheimer's disease in low-education but not in highly educated Hispanics. *Neuropsychologia, 49,* 3826–3830.

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

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1,* 131–149.

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition usingparent and teacher reports. *Applied Psycholinguistics, 24,* 267–288.

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics, 19,* 603–629.

Jackson-Maldonado, D., Thal, D. J., Marchman, V., Newton, T., Fenson, L., & Conboy, B. (2003). *El Inventario del Desarrollo de Habilidades Comunicativas: User's guide and technical manual* (MacArthur Foundation: Communicative Development Inventory). Baltimore, MD: Paul H. Brookes.

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. (2008). *Language disorders in bilingual children and adults*. San Diego, CA: Plural.

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders, 43,* 456–473.

Kohnert, K., Kan, P. F., & Conboy, B. (2010). Links between words and grammar in sequential bilingual preschoolers. *Journal of Speech, Language, and Hearing Research, 53,* 684–698.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A Web-based interfacefor bilingual research. *Behavior Research Methods, 38,* 202–210.

Lim, V., Rickard Liow, S., Lincoln, M., Chan, Y. H., & Onslow, M. (2008). Determining languagedominance in English–Mandarin bilinguals: Development of a self-report classification tool for clinical use. *Applied Psycholinguistics, 29,* 389–412.

MacSwan, J., & Rolstad, K. (2006). How language proficiency tests mislead us about ability: Implications for English language learner placement in special education. *Teachers College Record, 108,* 2304–2328.

Marian, V., Blumenfeld, H., & 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.

Marinova-Todd, S. H., & Uchikoshi, Y. (2011). The role of first language in oral language development in English: The case of both alphabetic and nonalphabetic languages. In A. Y. Durgunoğlu & C. Goldenberg (eds.), *Language and literacy development in bilingual settings*, pp. 29–60. New York: The Guilford Press.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7,* 703–704.

Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex, 2,* 244–258.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66,* 232–258.

Pray, L. (2005). How well do commonly used language instruments measure English oral language proficiency? *Bilingual Research Journal, 29,* 387–409.

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1398–1411.

Sheng, L., Bedore, L. M., Peña, E. D., & Fiestas, C. E. (2013). Semantic development in Spanish–English bilingual children: Effects of age and language experience. *Child Development, 84,* 1034–1045.

Sheng, L., Lu, Y., & Kan, P. (2011). Lexical development in Mandarin–English bilingual children. *Bilingualism: Language and Cognition, 14,* 579–587.

Shipley, W. C. (1946). *Institute of Living Scale*. Los Angeles, CA: Western Psychological Services.

Thal, D. J., Jackson-Maldonado, D., & Acosta, D. (2000). Validity of a parent-report measure of vocabulary and grammar for Spanish-speaking toddlers. *Journal of Speech, Language, and Hearing Research, 43,* 1087–1100.

To, C. K.-S., Stokes, S. F., Cheung, H.-T., & T'sou, B. (2010). Narrative assessment for Cantonese-speaking children. *Journal of Speech, Language, and Hearing Research, 53,* 648–669.

Treffers-Daller, J. (2011). Operationalizing and measuring language dominance. *International Journal of Bilingualism, 15,* 147–163.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Learning and Cognition, 4,* 105–122.

Windsor, J., Kohnert, K., Lobitz, K., & Pham, G. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology, 19,* 298–310.

Zhang, Y., Jin, X., Shen, X., Zhang, J., & Hoff, E. (2008). Correlates of early language development in Chinese children. *International Journal of Behavioral Development, 32,* 145–151.

Zied, K. M., Phillipe, A., Karine, P., Valerie, H.-T., Ghislaine, A., Arnaud, R., & Didier, L. G. (2004). Bilingualism and adult differences in inhibitory mechanisms: Evidence from a bilingual Stroop task. *Brain and Cognition, 54,* 254–256.