Bilingualism: Language and Cognition

cambridge.org/bil

Predicting naming scores from language history: A little immersion goes a long way, and self-rated proficiency matters more than percent use

Anne Neveu and Tamar H. Gollan

Department of Psychiatry, University of California, San Diego, La Jolla, CA 92093, USA

Research Article

Cite this article: Neveu, A., & Gollan, T.H. (2024). Predicting naming scores from language history: A little immersion goes a long way, and self-rated proficiency matters more than percent use. *Bilingualism: Language and Cognition*, 1–15. https://doi.org/10.1017/S1366728924000038

Received: 5 April 2023
Revised: 11 January 2024
Accepted: 12 January 2024

Keywords:
bilingualism; self-report; language dominance; aging; naming

Corresponding author:
Anne Neveu;
Email: aneveu@health.ucsd.edu

Abstract

Language proficiency is a critically important factor in research on bilingualism, but researchers disagree on its measurement. Validated objective measures exist, but investigators often rely exclusively on subjective measures. We investigated if combining multiple self-report measures improves prediction of objective naming test scores in 36 English-dominant versus 32 Spanish-dominant older bilinguals (Experiment 1), and in 41 older Spanish–English bilinguals versus 41 proficiency-matched young bilinguals (Experiment 2). Self-rated proficiency was a powerful but sometimes inaccurate predictor and better predicted naming accuracy when combined with years of immersion, while percent use explained little or no unique variance. Spanish-dominant bilinguals rated themselves more strictly than English-dominant bilinguals at the same objectively measured proficiency level. Immersion affected young more than older bilinguals, and non-immersed (English-dominant) more than immersed (Spanish-dominant) bilinguals. Self-reported proficiency ratings can produce spurious results, but predictive power improves when combined with self-report questions that might be less affected by subjective judgements.

Introduction

Measuring language proficiency, language dominance, and degree of bilingualism is central to psycholinguistic research on bilingualism and in clinical evaluations of bilingual patients (Gasquoine & Gonzalez, 2012; Lorenzen & Murray, 2008; Olson, 2023; Paplikar et al., 2021; Rivera Mindt et al., 2008). Language proficiency typically refers to how quickly, accurately, and easily a person can retrieve words and other linguistic structures, and facility of language use across various communicative contexts (Hulstijn, 2011). Language dominance refers to which language is more proficient, which can change for bilinguals over different points in their lifetime (Birdsong, 2014; Treffers-Daller & Silva-Corvalán, 2016). Numerous different language experiences are thought to influence language proficiency and dominance, including age of acquisition, frequency of use, contexts of use, formal education, time immersed in the language, and many others (Hulstijn, 2011; Schmid & Yılmaz, 2018). Both the variety of factors that can influence proficiency and the unique character of bilinguals' individual experiences make the accurate evaluation of proficiency, whether objective, subjective, or both, a complex endeavor (for a review, see Olson, 2023).

The use of self-report questionnaires to determine proficiency level in each of the bilinguals' languages is ubiquitous in research and clinical settings. Various language history questionnaires have been designed in attempts to gain a more comprehensive picture of a bilingual's language skills and try to uniformize data collection across different labs and clinics (e.g., the Language Experience and Proficiency Questionnaire – LEAP-Q –, Marian et al., 2007; the Language History Questionnaire – LHQ –, Li et al., 2006; the Language Use Questionnaire – LUQ -, Kastenbaum et al., 2018; and the Language and Social Background Questionnaire – LSBQ, Anderson et al., 2018; Luk & Bialystok, 2013 – for a detailed review, see Rothman et al., 2023). However, for a young participant, it takes at least 10 minutes to complete these questionnaires (with the exception of the LUQ which is longer). Moreover, the LEAP-Q has been widely adopted and has contributed to creating more consistency in measuring language history in bilingualism research, but there remains debate as to which aspects of bilingual experience should be used when trying to categorize bilinguals into groups (Kaushanskaya et al., 2020), and development of more time-efficient measures is critical in clinical settings and for encouraging wide use of uniform approaches to measurement.

This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

Self-ratings of proficiency

Perhaps the most widely used item in language history questionnaires is one that asks bilinguals to rate their proficiency level in each language in different modalities (reading, writing, speaking, understanding). However, self-ratings of proficiency often correspond to the participant's own perception of their skill rather than to a reflection of their true performance. In clinical settings it is more common to ask bilinguals to indicate which language is dominant or which language they prefer for testing and evaluation. However, the correlations between self-ratings and objective measures of proficiency have varied from small to moderate (e.g., Marian et al., 2007; Schrauf, 2009), to strong in size (e.g., Ross, 1998), which raises questions about their accuracy and predictive power. In addition, while self-ratings of language dominance are usually accurate, exceptions do occur (in which bilinguals perform better on objective tests in the language they said was less proficient (see Gollan et al., 2012; Tomoschuk et al., 2019), which can have serious unfortunate consequences. Notably, Tomoschuk et al. (2019) reported major discrepancies in how bilinguals of different language combinations (Chinese–English and Spanish–English) and even within the same language combinations bilinguals with a different dominant language (English-dominant vs. Spanish-dominant Spanish–English bilinguals) interpret self-rating scales. This suggests self-ratings should not be used when comparing or collapsing bilinguals of different language combinations, language dominance, and cultural backgrounds since they may not share the same points of reference, which compromises the extent to which self-ratings accurately reflect objective proficiency level.

Among objective measures of proficiency are picture naming (e.g., the Multilingual Naming Test – MINT) and oral proficiency interviews (OPIs; Gollan et al., 2012), and verbal fluency scores (e.g., Ardila et al., 1994; Artiola i Fortuny et al., 1998). One study by Gollan et al. (2012) examined to what extent self-report measures predicted objective measures of proficiency in younger and older Spanish–English bilinguals. Self-ratings of language dominance strongly correlated with measures of proficiency in each language. However, bilinguals who said they were Spanish-dominant tended to be more balanced in proficiency in the two languages, bilinguals who reported being balanced tended to be English-dominant, and bilinguals who rated themselves as English-dominant were the most accurate, although they may have overestimated their English proficiency level (i.e., their OPI and MINT scores were lower than expected given their self-rating). In a more recent study also of Spanish–English bilinguals, self-ratings exhibited low or moderate correlations with objective proficiency measures and were again better at predicting language dominance than absolute proficiency level in each language (Garcia & Gollan, 2022). Similarly, in young adult Mandarin–English bilinguals, objective measures of proficiency revealed that bilinguals who self-reported that they were balanced bilinguals performed better in English than in Mandarin on objective measures (Sheng et al., 2014). Thus, in all three studies, bilinguals were more accurate in determining their dominance than their proficiency level, and though self-rated proficiency was significantly correlated with objectively measured proficiency in both languages, many problematic discrepancies between self-ratings and objective measures were apparent.

Besides differences in the consistency of self-ratings across language combinations and language dominance, differences have also been found across different age groups. In Gollan et al. (2012), younger and older bilinguals were not compared directly (they were tested in separate experiments). Spanish-dominant older bilinguals, based on their self-ratings, were 36%[1] more proficient in Spanish than in English. Similarly, based on the OPI, they were 32% more proficient in Spanish than in English, but based on the MINT, this value was only 7% (see Gollan et al., 2012, Table 3). By contrast, young Spanish-dominant bilinguals were 20% more proficient in Spanish than in English based on their self-ratings but were more balanced (only 2% more proficient in Spanish) based on the OPI and the MINT (Gollan et al., 2012, Table 3). Thus, young Spanish-dominant bilinguals were more balanced than they realized by both objective measures, whereas the older Spanish-dominant's self-rated dominance matched the gold standard measure (the OPI). In another recent study (Stasenko et al., 2021), older bilinguals on average scored significantly higher on the picture naming test (the MINT) in both languages than young bilinguals, but the same young and older bilinguals classified themselves as having equivalent self-rated proficiency level. These discrepancies might reflect between-group differences in standards of excellence, or older adults' ratings might be lowered by their sense of increasing word-finding difficulties (Burke et al., 1991; Gollan et al., 2008).

Given the inconsistencies in self-ratings and findings of low predictive power in some studies, it would be of interest to determine if the accuracy of self-report can be increased by combining (or perhaps even replacing) self-ratings of proficiency level with other self-report measures.

Percent (frequency) of language use

In line with Grosjean's complementarity principle (1998), where bilinguals are rarely equally fluent in all the languages they know, previous work has shown that frequency of language use is associated with objectively measured proficiency level (Luk & Bialystok, 2013). Luk and Bialystok (2013) examined what factors of bilingual experience matter when examining the consequences of bilingualism on language and cognition. They used exploratory and confirmatory factor analyses and derived two factors: self-reported bilingual language use (self-reported speaking and listening skills in the language used at home together loaded on one factor) and English proficiency (measured with the Peabody Picture Vocabulary Task-III, Form A, Dunn & Dunn, 1997, and the Expressive Vocabulary Task, Williams, 1997; together loaded on a second factor). There was a moderate negative correlation between bilingual language use and English proficiency, suggesting that bilinguals who used a language other than English more often had lower English proficiency than bilinguals who seldom used the other language.

To try to understand the effects of language experience across different social contexts, Gullifer and Titone (2020) derived a measure of distributed language use, called language entropy. Language entropy is computed from composite measures of language use extracted from existing questionnaire data (LEAP-Q, Marian et al., 2007; or LHQ 2.0, Li et al., 2014). A comprehensive examination of patterns of language use across different contexts in daily life has shown that more distributed language use (a higher entropy score) predicted proficiency in the second language, as measured by self-rating questions condensed into one component, over and above second language age of acquisition and exposure (Gullifer & Titone, 2020). However, objective proficiency level was not measured in that study, a gap filled in a subsequent study, which further examined how various aspects of

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

bilingual language use and experience relate to each other (Gullifer et al., 2021). Indeed, in that more recent study, a combination of subjective self-report measures and objective measure of verbal fluency in both languages (category and letter fluency tasks in both English and French) were used to define proficiency (Gullifer et al., 2021). Bilinguals with high scores in L2 verbal fluency also tended to self-rate themselves higher in the L2, although previously, bilinguals have been found to inaccurately judge their performance in the L2 (Tomoschuk et al., 2019). This correlation between objective and subjective ratings in the L2 was however not found in the L1. These findings suggest that self-ratings may depend on the characteristics of the groups sampled, such as their language dominance or language combination (see Tomoschuk et al., 2019).

A caveat for use of detailed questionnaires with older participants and in clinical settings is that administration time would likely be even longer (e.g., when working with cognitively impaired individuals). While it is feasible to dedicate this time to collect language history data in a research setting, it is less so the case in a clinical setting where time is primarily spent towards targeted referral questions (e.g., is there cognitive impairment?). Therefore, to optimize knowledge about proficiency from self-report questions, a "Goldilocks zone" must be found between too many and too few language history questions. At this juncture, the field is ripe for investigating which types of questions provide the most predictive power in the least amount of time to meet the demands in clinical settings, and to encourage widespread use of the best predictors in research settings. Across different labs, investigators would be more likely to adopt a set of commonly used questions and objective measures if the time commitment could be kept to a minimum.

Immersion

An important factor that might introduce differences between younger and older bilinguals is that a lifetime of bilingualism provides more time and therefore longer cumulative use of two languages (Gollan et al., 2008), and a lifetime might also provide more opportunities for extended immersion experience. Time spent immersed in an environment where the nondominant language is spoken increases proficiency of that language in young adults, and even temporary and relatively short-lived immersion can have powerful effects (e.g., Linck et al., 2009; Lynch et al., 2001). Immersion is more efficient for developing skills in a nondominant language compared to learning in a classroom while immersed in the dominant language (Linck et al., 2009). Immersion may often occur due to immigration, which may take place early in life (e.g., among Heritage language speakers), or later in life. In older bilinguals, this could result in extended periods (decades of immersion) in different languages early versus later in life. Research on the effects of long-term immersion (through immigration) in older adults with and without dementia has shown significant and positive correlations between the number of years immersed in their nondominant language and their proficiency in that language (Nanchen et al., 2017). Specifically, parallel decline of both languages across patient and control groups suggests that living immersed in one's nondominant language can help preserve it to a similar extent as the dominant language, regardless of the severity of cognitive decline experienced in older age (Nanchen et al., 2017).

Outcomes in terms of maintenance of the native language after immigration and development of the majority language in the new country depend on many factors including the language spoken at home and the type of school attended (two-way dual-language immersion versus majority language-only). In a study of Welsh–English bilinguals, adults (parents) who continued speaking Welsh at home, compared to using both Welsh and English, maintained their proficiency levels in Welsh to a larger extent (Gathercole & Thomas, 2009). While more exposure to Welsh in childhood tended to yield lower proficiency in English, these gaps closed when reaching adulthood, across all profiles of language use at home and at school (Gathercole & Thomas, 2009). In this and other studies, the conclusion is that maintenance of the native language and development of the later acquired language depend on quantity of language input in each of these languages (e.g., Gathercole & Thomas, 2009; Hoff, 2018; Hurtado et al., 2014; Thordardottir, 2011). Additionally, quality of exposure also matters: specifically, both exposure (from native versus non-native speakers) and speaker variability have been found to support the development of fluency, both in children (De Cat, 2021; Hoff et al., 2014; Unsworth, 2016; but see Carroll, 2017) and adults (Linck et al., 2009; Sinkeviciute et al., 2019).

The current study

Little attention has been given as to which self-report questions, or combined individual short-lists of questions, are more powerful predictors of objectively measured proficiency level, and whether these might vary across bilingual subgroups formed by language dominance and age group. In the current study, we assessed the joint predictive power of self-rated proficiency level, self-reported frequency of use, and years of immersion for predicting picture naming scores in the nondominant language. We focused on the nondominant language which produced stronger correlations between self-rated proficiency and naming scores in previous studies due to broader ranges of skills in that language (e.g., Garcia & Gollan, 2022; Gollan et al., 2012; Sheng et al., 2014; see also Marian et al., 2007), while the dominant language tends to be closer to ceiling levels in both self-ratings and objectively measured performance (and therefore is harder to predict). To this end, we analyzed data gathered in several previous studies (see below for details) on younger and older bilinguals where objective language proficiency was measured with the MINT. Using these data, in Experiment 1, we compared English-dominant to Spanish-dominant older bilinguals, and in Experiment 2, we compared younger and older bilinguals. In both experiments we investigated if groups differed systematically in self-rating measures, and if self-rated proficiency level, years of immersion and percent use together predicted nondominant language naming scores better than each predictor on its own.

Experiment 1 – Language Dominance Effects in Older Bilinguals

Methods

Participants

Sixty-eight cognitively healthy older Spanish–English bilinguals for whom item level data on the picture naming test in both languages were readily available from two previous studies (Gollan et al., in press; Gollan & Goldrick, 2016) were selected for analysis. Fifty-nine were tested on the MINT during their yearly evaluation as part of their participation in a longitudinal study at the

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

University of California, San Diego (UCSD) Alzheimer's Disease Research Center (ADRC), and nine were part of a separate research study (Gollan & Goldrick, 2016). The study procedures were approved by the UCSD Institutional Review Board. Participants characteristics are presented in Table 1. Participants were living in San Diego, which is about 15-20 miles from the Mexican border, and in which both English and Spanish are used frequently. About half the bilinguals were English-dominant and therefore were immersed in their dominant language, whereas half were Spanish-dominant bilinguals and immersed in their nondominant language. Seven participants reported having some immersion in a non-English- or non-Spanish-speaking country, ranging from 2 months to 6 years. Classification of language dominance was derived from the average of self-ratings of proficiency level in English and Spanish in four modalities (reading, writing, speaking and listening comprehension) on a 1 to 7 scale. Self-rated proficiency was averaged across the four modalities and whichever average was higher determined which language is dominant. For four bilinguals the average score was the same for the two languages. For two of these we used self-reported percent of current English use to determine dominance; two reported using English more often than Spanish and were thus classified as English-dominant and one reported using Spanish more often than English and was classified as Spanish-dominant. For the remaining balanced participant, percent current English use was at 50%, so we looked at number of years immersed in a non-English speaking country: this number corresponded to the participant's age (89), with zero years immersed in the nondominant language, and therefore we classified this bilingual as English-dominant.

When compared to English-dominant bilinguals, the Spanish-dominant bilinguals had significantly lower education level, and lower picture naming scores in the dominant language (see Materials and Procedure below), but a higher percent use of the nondominant language, and more years of immersion in the nondominant language. Average self-rated proficiency level in the dominant language was at ceiling for both groups, and for the English-dominant group, average self-rated proficiency in the nondominant language tended to be slightly higher than for the Spanish-dominant group, although this difference was not significant (see Table 1).

Materials

Participants named pictures from the MINT, in each of their languages (English and Spanish). The MINT comprises 68 black-and-white pictures that are increasing in difficulty level from beginning to end. If a participant had difficulty recognizing the picture, a semantic cue was provided. If the correct name was produced before or after the semantic cue, it was coded as correct. If the name was not produced at that point, a phonetic cue was provided, and the item was coded as incorrect.

Procedure

The testing session was conducted by a proficient Spanish–English bilingual at the ADRC or in participants' homes. The MINT was administered towards the end of testing session in which other (unrelated) tasks were administered. For participants tested at the ADRC (n = 59), testing was discontinued after 6 items could not be named, but note that the 9 bilinguals who were tested on the complete test (without discontinuation after 6 failed items) named on average less than 1 additional picture correctly after the point where the discontinuation rule would have been applied (on average just .67 points; *SD* = 1): thus, we assume this small difference in procedure likely had minimal or no effect. Naming accuracy was recorded simultaneously while testing.

Correlations between the language history variables and MINT naming scores in the nondominant language are summarized in Table 2. The full correlation tables separated by language dominance groups are available in Supplementary Materials, in Appendix SA, Tables SA.1 and SA.2.

Table 1 Participant characteristics for Experiment 1, sample of older bilinguals divided by language dominance

	Language dominance					
Characteristic	English[a] *(n = 36)*		Spanish[a] *(n = 32)*		t-test *t*	*p*-value
Gender (female/male)	24/12		23/9		0.04[b]	.84
	M	*(SD)*	*M*	*(SD)*		
Age at MINT	72.5	(9.5)	73.8	(7.8)	−0.60	.55
Education in years	14.7	(2.6)	12.5	(3.6)	2.81	.01
MINT score						
Dominant language	64.4	(3.3)	61.4	(3.5)	3.53	.001
Nondominant language	45.4	(14.5)	47.1	(15.4)	−0.49	.63
Percent current use nondominant language	12.2	(13.4)	25.8	(22.0)	−3.03[c]	.004
Years immersed nondominant language	3.3	(5.3)	37.2	(19.9)	−9.35[c]	<.001
Self-rated proficiency						
dominant language	6.8	(0.5)	6.8	(0.4)	−0.27	.79
nondominant language	4.9	(1.4)	4.5	(1.7)	1.11	.27

[a]Defined by average self-rated dominance (across reading, speaking, writing and understanding, on a scale from 1 (lowest) to 7 (highest).
[b]For the variable of gender, Pearson's Chi-squared test was run instead as the data are categorical.
[c]Welch's t-test was used as variances across groups were unequal.

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

Table 2 Pearson's correlations between language history variables and nondominant MINT scores for Experiment 2

	Multilingual Naming Test – Nondominant Language			
		Older Bilinguals		
	English-dominant *(n = 36)*	*p*-value	Spanish-dominant *(n = 32)*	*p*-value
Self-rated proficiency	.79	<.001	.74	<.001
Years immersed	.49	.003	.44	.01
Percent current use	.37	.03	.61	<.001
Age at MINT	.20	.23	−.30	.09
Education (yrs)	.16	.35	.43	.01

Of the three predictors of primary interest, the correlations between self-rated proficiency and MINT scores in the nondominant language were strongest (in both English- and Spanish-dominant bilinguals). Immersion and percent current use were also significantly and positively correlated with nondominant MINT scores in both groups. Age was not a significant predictor, and years of education was significantly correlated with MINT scores, in both dominance groups. However, education was collinear with self-rated proficiency in the nondominant language in English-dominant bilinguals ($r = .34$, $p < .05$) and Spanish-dominant bilinguals ($r = .61$, $p < .001$), and as such we did not include it in our analyses.

Analyses

We examined the extent to which average self-rated proficiency in the nondominant language, years immersed in the nondominant language, and percent current use of the nondominant language predicted the likelihood of naming a picture accurately in the nondominant language. We examined the joint power of the three predictors in the full sample first, and next each predictor separately (to avoid running overly complex models), with each interacting with language dominance group. As naming accuracy is a binary outcome variable (1 or 0), we analyzed the data using logistic mixed-effects models (lme4 package, Bates et al., 2015) in R Studio, version 4.2.3 (2023-03-15). To control for the closer similarity of answers within- compared to between-participants, as each provided 68 answers, models included a by-subject random intercept. We also included by-item random intercepts and slopes for each of the between-subjects variables and simplified the random-effects structure when applicable to resolve convergence and singularity issues (Brauer & Curtin, 2018). Model assumptions were tested with the DHARMa package (Hartig, 2022), and were satisfied.

Results

Joint predictive power of the three self-report measures

We ran a mixed-effects logistic regression model including the main effects of self-rated proficiency in the nondominant language, years immersed in the nondominant language and percent current use of the nondominant language. The model converged and no singularity issues emerged with the full random-effects structure, including a by-subject and by-item random intercept, and by-item random slopes for each of the three independent variables. The model included 4624 observations. There was a main effect of self-rated proficiency such that the odds of accurately naming an item on the MINT increased by a factor of 3.13 – corresponding to medium effect size of $d = .5$ (Chen et al., 2010) per each additional unit of self-rated proficiency ($b = 1.14$, $SE = 0.15$, $z = 7.48$, $OR = 3.13$, $p < .001$). There was also a main effect of years immersed such that the odds of accurately naming an item on the MINT increased by a factor of 1.02 (small effect size) per each additional year immersed ($b = 0.02$, $SE = 0.01$, $z = 3.01$, $OR = 1.02$, $p = .003$). The main effect of percent current use was not significant ($p = .17$, see Supplementary Materials, Appendix SB, Table SB.1a for full results).

To examine whether effects of percent current use of the language might have been obscured by years of immersion, we further explored this variable along with self-rated proficiency level in a subset of the data that included only individuals who reported having no immersion in the nondominant language. Only English-dominant bilinguals fit this profile. The model included 1156 observations and convergence and singularity issues were resolved by removing random slopes and retaining by-item and by-subject random intercepts. The model showed a significant main effect of average self-rated proficiency in the nondominant language such that the odds of accurately naming an item on the MINT increased by a factor of 2.59 for every one unit increase in self-rated proficiency (medium effect size) ($b = 0.95$, $SE = 0.36$, $z = 2.60$, $OR = 2.59$, $p = .009$). The main effect of percent current use of the nondominant language was still not significant ($p = .57$, see Supplementary Materials, Appendix SB, Table SB.1b for full results). In summary, average self-rated proficiency was the strongest predictor of naming accuracy, followed by immersion, and percent current use was not a significant predictor of naming accuracy, even in bilinguals with zero years of immersion. These results did not change when adding education as a covariate in the models.

Next, to determine whether there were significant differences between groups in terms of which variables predicted nondominant MINT accuracy, we looked for interactions between participant group and each of the self-report questions. We ran three mixed-effects logistic regression models including the interaction of participant group (coded as -0.5 for Spanish-dominant and 0.5 for English-dominant) with self-rated proficiency, immersion and percent current use of the nondominant language. Models converged and singularity issues were resolved by including by-subject and by-item random intercepts and removing lower-order random effect terms. Briefly summarized, the models contrasting self-rated proficiency level by group, and years of immersion by group showed significant differences between English-dominant and Spanish-dominant bilinguals, while the

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

model examining self-reported current percent use of the nondominant language showed no difference between groups. These results did not change when adding education as a covariate in the models.

Self-rated proficiency level by language dominance group

The model predicting by-item MINT accuracy from the interaction of self-rated proficiency by participant group included 4624 observations (see Figure 1). There was a main effect of self-rated proficiency such that the odds of accurately naming an item on the MINT increased by a factor of 3.46 (medium effect size) per each additional unit of self-rated proficiency ($b = 1.24$, $SE = 0.13$, $z = 9.16$, $OR = 3.46$, $p < .001$). There was also a main effect of participant group such that at each level of self-rated proficiency (which ranged from 1-7 in these bilinguals), Spanish-dominant bilinguals named more pictures correctly in their nondominant language than did English-dominant bilinguals, by a factor of 0.34 (small effect size) ($b = -1.07$, $SE = 0.40$, $z = -2.67$, $OR = 0.34$, $p = .008$). The interaction was not significant ($p = .54$, see Supplementary Materials, Appendix SB, Table SB.2).

Years of immersion by language dominance group

The model predicting by-item MINT accuracy from the interaction of years of immersion by participant group included 4624 observations. There was a main effect of immersion such that the odds of accurately naming an item on the MINT increased by a factor of 1.15 (small effect size) per each additional year immersed ($b = 0.14$, $SE = 0.04$, $z = 3.88$, $OR = 1.15$, $p < .001$). Given the same average number of years immersed, English-dominant bilinguals named more pictures correctly in the nondominant language compared to Spanish-dominant bilinguals, by a factor of 63.43 (very large effect size) ($b = 4.15$, $SE = 1.25$, $z = 3.31$, $OR = 63.43$, $p < .001$). The interaction of participant group and years of immersion was significant, such that for both groups, each additional year of immersion increased accuracy, but the effect was stronger for English-dominant than it was for Spanish-dominant bilinguals ($b = 0.17$, $SE = 0.07$, $z = 2.37$, $OR = 1.19$, $p = .02$) (see Supplementary Materials, Appendix SB, Table SB.3). Follow-up comparisons suggested that the odds of accurately naming an item on the MINT for the English-dominant group increased by a factor of 1.25 for any additional year of immersion, while they increased by a slightly smaller factor of 1.06 for the Spanish-dominant group (see Figure 2) (both small effect sizes).

To consider if the interaction was robust to a control for between-group differences in education level (see Table 1), we ran a model where only participants with 12 years or more of education were included (which included 35 English-dominant and 20 Spanish-dominant bilinguals who did not differ significantly in education level, $t = -0.36$, $p = 0.72$). Twelve years corresponds to completing high school, a level more commonly shared between undergraduate college students and the older adults in this sample. This matched analysis revealed highly similar point estimates, standard errors and z-statistics as in the model with all participants (see Supplementary Materials, Appendix SB, Table SB.4).

Percent current language use by language dominance group

The model predicting by-item MINT accuracy from the interaction of percent current language use by participant group included 4624 observations. There was a main effect of percent current language use such that the odds of accurately naming an item on the MINT increased by a factor of 1.07 (small effect size) per each additional percentage point of use ($b = 0.07$, $SE = 0.02$, $z = 4.25$, $OR = 1.07$, $p < .001$). The main effect of language dominance ($p = .47$) and the interaction of percent current language use and language dominance ($p = .50$) were not significant (see Supplementary Materials, Appendix SB, Table SB.5).

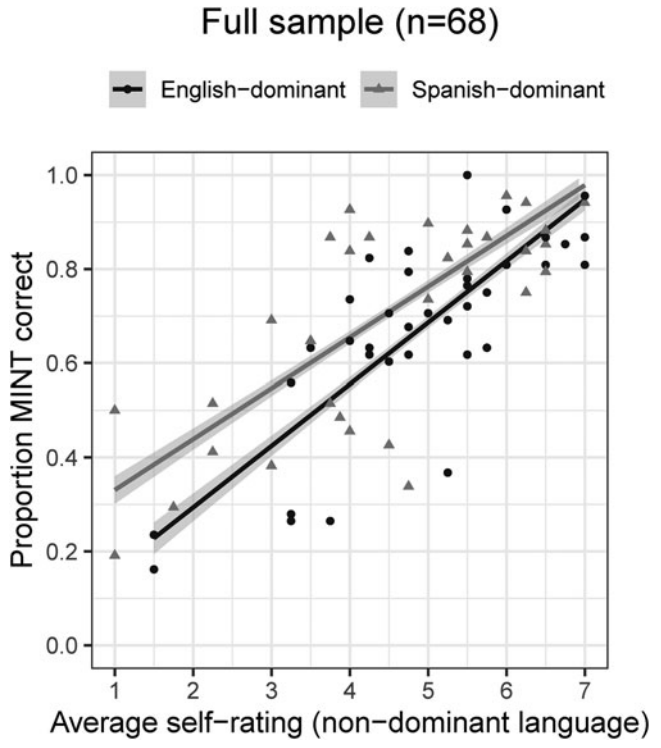

Figure 1. Experiment 1 - Predicted proportion correct on MINT accuracy in the nondominant language by average self-rated proficiency level in the nondominant language, across language dominance groups. Raw and predicted data are superimposed; grey ribbons show standard error.

Discussion

The results of Experiment 1 revealed several key findings, including some significant differences between English-dominant and Spanish-dominant older bilinguals. First, average self-rated proficiency level was the most powerful predictor of objectively measured proficiency in the nondominant language using picture naming scores in the overall sample. We also found significant correlations between years of immersion and naming scores, and relative to Spanish-dominant bilinguals, each year of immersion had a stronger effect on English-dominant bilinguals, who had relatively few years of immersion (and many English-dominant bilinguals had no immersion experience). Although self-reported percent use of the nondominant language was also correlated with naming scores, it did not explain additional variance when jointly predicting naming scores along with self-rated proficiency level (even when considering individuals with no immersion experience).

In addition, self-rated proficiency level seemed to be a powerful predictor, but at each level of self-rated proficiency, Spanish-dominant bilinguals outperformed English-dominant bilinguals in naming scores in the nondominant language, i.e., Spanish-dominant bilinguals scored higher in English than

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

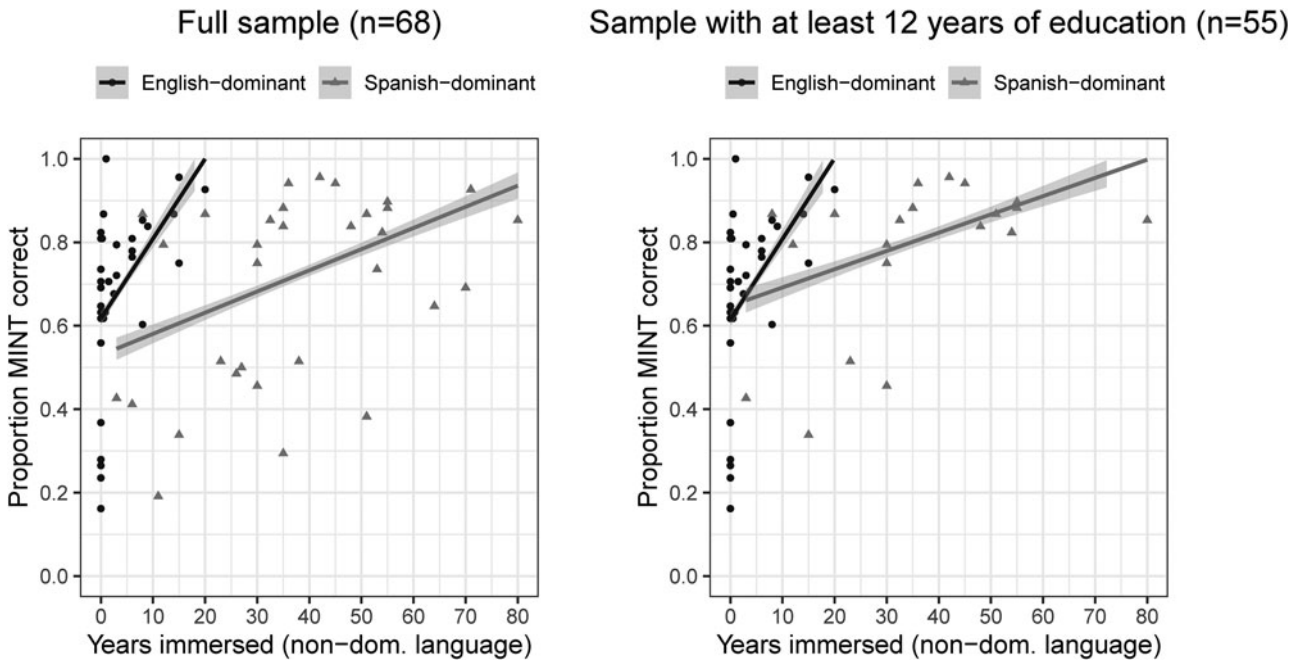

Figure 2. Experiment 1 - Predicted proportion correct on MINT accuracy in the nondominant language by the number of years immersed in the nondominant language, across dominance groups. Raw and predicted data are superimposed; grey ribbons show standard error.

English-dominant bilinguals scored in Spanish (see Figure 1). This outcome could reflect differences in how different groups interpret the self-rating scale, or a systematic bias in the MINT (e.g., if the test were objectively easier in English than in Spanish). The latter seems unlikely given that the MINT was developed from the ground up as a naming test for multiple languages (with English and Spanish among them). We defer further discussion of these possibilities to the General Discussion but note that other clear inaccuracies in the rating scales were apparent. For example, seven bilinguals rated their proficiency in the nondominant language between 1 ("very poor") and 2.25 (with 2 corresponding to "poor"), thus considering themselves functionally monolingual. Four of these participants named between 11 and 20 pictures in the nondominant language, but three named more than a third of the pictures correctly, a number that clearly exceeds what most would consider "monolingual" and revealing another inaccuracy in the use of self-rated proficiency level.

In Experiment 2, we investigated age group differences while also further examining the effects of immersion in a larger number of bilinguals (adding young bilinguals allowed us to include many more participants from a larger set of available previous studies that did not include older bilinguals).

Experiment 2 – Aging Effects

Methods

Participants

One hundred and twenty-three bilinguals were selected for analysis based on available proficiency matched young bilinguals. Forty-one were older bilinguals, a subset from Experiment 1 with two younger bilinguals matched for MINT scores in English and Spanish, and also for self-reported percentage of current English use. We used the case control matching function in SPSS (version 28.0.0.1) and allowed a three-point fuzzy match per MINT score, and a 20-percent fuzzy match for percent current English use. Younger bilinguals' data came from eleven different studies. These bilinguals were undergraduates at the University of California, San Diego. The study procedures were approved by the UCSD Institutional Review Board.

Participants characteristics were collected through a language history questionnaire and are presented in Table 3. Language dominance was derived in the same way as in Experiment 1; 17 bilinguals had the same average self-rated proficiency level in both languages (two older adults and 15 younger adults). Of these, 14 had a higher percent current use of English, so these were classified as English-dominant, and one reported greater current use of Spanish, so this bilingual was classified as Spanish-dominant. The remaining two had zero years spent in a non-English speaking country, and therefore we also classified these as English-dominant. Three of the older participants from Experiment 1 who had reported some brief immersion in a country where English or Spanish is not spoken were also present in this sample, and two younger bilinguals reported immersion in non-English and non-Spanish-speaking countries, ranging from one month to three years.

Older bilinguals acquired the nondominant language on average about 2.9 years later compared to younger bilinguals, but both young and older bilinguals reported acquiring the nondominant language early in childhood (on average before age 5 even for older bilinguals). While self-rated proficiency in the dominant language was equivalent (and close to ceiling) in both groups, older bilinguals scored significantly higher in their dominant language on the MINT compared to younger bilinguals (see also Stasenko et al., 2021). An additional apparent inaccuracy in the self-ratings of proficiency was that younger bilinguals self-rated their proficiency level in the nondominant language as significantly higher than that of the older bilinguals, but MINT scores in the nondominant language were not significantly different

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

Table 3. Participant characteristics for older and younger bilinguals matched in Experiment 2.

Characteristic	Age Group				Welch's t-test *t*	*p*-value
	Older bilinguals *(n = 41)*		Younger bilinguals *(n = 82)*			
Gender (female/male) ns.	30/11		58/24		0.00[a]	.94
English-dominant/Spanish-dominant[b]	29/12		64/18		0.45[a]	.50
	M	*(SD)*	*M*	*(SD)*		
Age at MINT testing	73.2	(9.3)	20.2	(1.9)	36.07	<.001
Education in years	14.3	(3.0)	13.9	(1.8)	0.90	.37
Age of acquisition nondominant language[c]	4.6	(7.4)	1.7	(2.9)	2.36	.02
MINT score						
Dominant language	63.0	(3.7)	60.9	(5.2)	2.58	.01
Nondominant language	50.9	(9.7)	51.2	(9.4)	−0.15[d]	.88
English	62.3	(4.6)	61.2	(4.3)	1.32[d]	.19
Spanish	51.6	(10.1)	50.9	(9.5)	0.39[d]	.70
Percent current use nondominant language	22.0	(21.5)	24.8	(20.8)	−0.69[d]	.49
Years immersed nondominant language	17.9	(24.3)	4.4	(6.4)	3.49	.001
Proportion of life immersed	25.8	(32.8)	21.9	(32.4)	0.60[d]	.55
Self-rated proficiency						
dominant language	6.8	(0.4)	6.7	(0.5)	1.23	.22
nondominant language	5.2	(1.0)	5.8	(0.8)	−3.32	.001

[a]For these variables, Pearson's Chi-squared test was run instead as the data is categorical.
[b]Dominance is derived from the highest average self-rating (across speaking, understanding, reading and writing skills) between English and Spanish.
[c]Data available for 41 older adults.
[d]A two-sample t-test was used as variances were equal between variables.

across groups. Finally, older bilinguals had significantly more years of immersion in the nondominant language, but the proportion of lifetime immersed were not different across groups.

Correlations between the language history variables and nondominant language MINT scores are summarized in Table 4. The full correlation tables separated by age groups are shown in Supplementary Materials, in Appendix SC, Tables SC.1 and SC.2.

The correlations showed that average self-rated proficiency again predicted naming scores in the nondominant language, but to a more moderate extent than in Experiment 1, which included more bilinguals with lower proficiency level than we had in Experiment 2 (compare Figures 1 & 3). Immersion and percent current use were also again significantly correlated with nondominant language MINT scores. Age and education were weak predictors only in young bilinguals, tended not to vary much (as would be expected in college undergraduates), but we briefly considered their effects in the analysis of young bilinguals below.

Materials

The materials were the same as in Experiment 1.

Procedure

The procedure was the same for the older bilinguals as in Experiment 1, young bilinguals were tested without the stopping rule (but see in Table 3, this had little or no effect on performance for most older participants). Out of a sample of 25 older bilinguals and 25 younger bilinguals selected randomly, 96% of older bilinguals gained 0 or just 1 point without applying the stopping rule (only one gained 3 points) and 80% of younger bilinguals gained 0 or just 1 point, and only 12% (3 bilinguals) gained 4 or 5 points. Therefore, the slight difference in administration procedure likely did not significantly impact the results reported below.

Table 4 Pearson's correlations between language history variables and nondominant MINT scores for Experiment 2

	Multilingual Naming Test – Nondominant Language			
	Age Group			
	Older bilinguals *(n = 41)*	*p*-value	Younger bilinguals *(n = 82)*	*p*-value
Self-rated proficiency	.52	<.001	.39	<.001
Years immersed	.56	<.001	.50	<.001
Percent current use	.41	.01	.45	<.001
Age at MINT	.02	.90	.23	.04
Education (yrs)	.18	.25	.23	.04

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

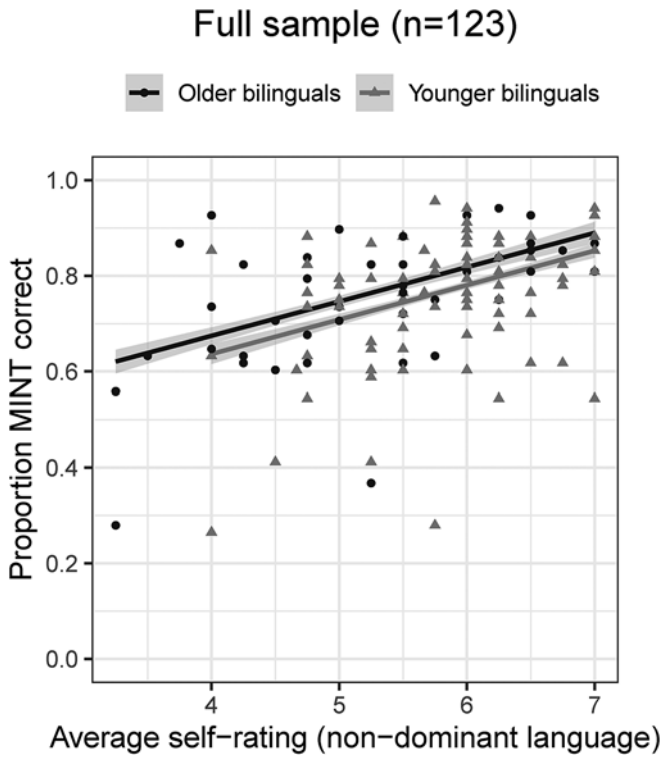

Figure 3. Experiment 2 - Predicted proportion correct on MINT accuracy in the nondominant language by average self-rated proficiency level in the nondominant language, across age groups. Raw and predicted data are superimposed; grey ribbons show standard error.

Analyses

We ran the same set of analyses as in Experiment 1 but contrasting age groups instead of language dominance groups. Model assumptions were also satisfied.

Results

Joint predictive power of the three self-report measures

We ran a mixed-effects logistic regression model including the main effects of self-rated proficiency in the nondominant language, years immersed in the nondominant language and percent current use of the nondominant language. The model converged and singularity issues emerged with the full random-effects structure, which we simplified by removing covariances among random effects. The random-effects structure included by-subject and by-item random intercepts, and by-item random slopes for each of the three independent variables. The model included 8364 observations. There was a main effect of self-rated proficiency such that the odds of accurately naming an item on the MINT increased by a factor of 1.92 (small effect size) per each additional unit of self-rated proficiency ($b = 0.65$, $SE = 0.14$, $z = 4.61$, $OR = 1.92$, $p < .001$). There was also a main effect of years immersed such that the odds of accurately naming an item on the MINT increased by a factor of 1.05 per each additional year immersed ($b = 0.05$, $SE = 0.01$, $z = 5.01$, $OR = 1.05$, $p < .001$). The main effect of percent current use was also significant, such that the odds of accurately naming an item on the MINT increased by a factor of 1.02 per each additional percentage point ($b = 0.02$, $SE = 0.01$, $z = 2.84$, $OR = 1.02$, $p = .004$) (see Supplementary Materials, Appendix SD, Table SD.1). Adding education as a covariate did not affect results.

We next ran three mixed-effects logistic regression models including the interaction of age group (coded as -0.5 for older adults and 0.5 for younger adults) with each of the three predictors: i.e., self-rated proficiency level, immersion, and percent current use. Models converged and singularity issues were resolved by including by-subject and by-item random intercepts and removing lower-order random effect terms. Adding education as a covariate again did not affect results.

Self-rated proficiency level by age group

The model predicting by-item MINT accuracy from the interaction of self-rated proficiency by age group included 8364 observations. There was a main effect of average self-rated proficiency such that the odds of accurately naming an item on the MINT increased by a factor of 2.34 (small to medium effect size) per each additional unit of self-rated proficiency ($b = 0.85$, $SE = 0.16$, $z = 5.33$, $OR = 2.34$, $p < .001$). Older bilinguals tended to score higher than younger bilinguals at each level of self-rated proficiency, but this main effect of age group ($p = .12$) and the interaction ($p = .94$) was not significant (see Figure 3 and Supplementary Materials, Appendix SD, Table SD.2).

Years of immersion by age group

The model predicting by-item MINT accuracy from the interaction of nondominant language immersion by age group included 8364 observations. There was a main effect of immersion such that the odds of accurately naming an item on the MINT increased by a factor of 1.11 (small effect size) per each additional year immersed ($b = 0.10$, $SE = 0.01$, $z = 7.40$, $OR = 1.11$, $p < .001$). Given the same average number of years immersed, younger adults named more pictures correctly in the nondominant language compared to older adults, by a factor of 2.72 (medium effect size) ($b = 1.00$, $SE = 0.30$, $z = 3.36$, $OR = 2.72$, $p < .001$). The interaction of age group and immersion was significant, such that for both groups, each additional year immersed improved accuracy, but the effect was stronger for younger bilinguals so that age effects were strongest when approaching 20 years of immersion (which was not far from the average lifespan of younger bilinguals) ($b = 0.08$, $SE = 0.03$, $z = 2.61$, $OR = 1.08$, $p = .009$) (see Supplementary Materials, Appendix SD, Table SD.3). Indeed, a follow-up of this interaction revealed that the odds of being accurate on the MINT for the younger adults increased by a factor of 1.16 for every additional year immersed, while they increased by a factor of 1.05 for the older adults (see Supplementary Materials, Appendix SD, Figure SA).

Given that many English dominant bilinguals had no immersion experience we repeated the age-contrast after excluding Spanish-dominant bilinguals i.e., in English-dominant participants only ($n = 93$). The model included 6324 observations and converged without singularity issues with the full random-effects structure. As before, the model revealed a main effect of immersion such that the odds of accurately naming an item on the MINT increased by a factor of 1.22 per each additional year immersed ($b = 0.20$, $SE = 0.04$, $z = 4.61$, $OR = 1.22$, $p < .001$). Importantly, given the same average number of years immersed, younger bilinguals still named more pictures correctly in the nondominant language compared to older bilinguals, by a factor of 2.44 (small to medium effect size) ($b = 0.89$, $SE = 0.39$, $z = 2.26$, $OR = 2.44$, $p = .024$). The interaction of age group and immersion was however no longer significant; instead younger bilinguals

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

named more pictures than older bilinguals at every level of years of immersion (see Figure 4) (see Supplementary Materials, Appendix SD, Table SD.3a). By contrast, when analyzing the interaction of immersion with age group in the Spanish-dominant group only (2040 observations and removing lower-order random effects terms to solve singularity issues), none of the main effects or the interaction were significant (see Figure 4, and Supplementary Materials, Appendix SD, Table SD.3b).

Percent current language use by age group

The model predicting MINT accuracy from the interaction of percent current use by age group included 8364 observations. There was a main effect of percent current use such that the odds of accurately naming an item on the MINT increased by a factor of 1.04 (small effect size) per each additional percent of use ($b = 0.04$, $SE = 0.01$, $z = 5.48$, $OR = 1.04$, $p < .001$). The main effect of age group ($p = .81$) and the interaction ($p = .89$) were not significant (see Supplementary Materials, Appendix SD, Table SD.4).

We last examined the role of percent use in the much larger dataset of Tomoschuk et al. (2019), where both percent current use and percent use of the nondominant language while growing up was available, but no data on years immersed in the nondominant language was available. The data were coded at the subject-level, therefore we ran linear models. The model predicting nondominant MINT accuracy from percent current use included 979 observations. There was a significant but very small main effect of percent current use such that for every percent increase in usage, accuracy increased by .08 units ($b = .08$, $SE = .01$, $t = 6.78$, $p < .001$). There was also a main effect of self-rated proficiency such that for every one unit increase in rating, accuracy increased by 4.08 units ($b = 4.08$, $SE = .44$, $t = 9.30$, $p < .001$).

The model predicting nondominant MINT accuracy from percent use while growing up included 979 observations. There was a main effect of percent use while growing up such that for every percent increase in usage, accuracy increased by .08 units ($b = .08$, $SE = .02$, $t = 4.97$, $p < .001$). There was also a main effect of self-rated proficiency such that for every one unit increase in rating, accuracy increased by 4.36 units ($b = 4.36$, $SE = .44$, $t = 9.91$, $p < .001$).

Discussion

Self-rated proficiency had the largest effect on naming accuracy, and immersion and percent current use had similar, small effect sizes. We note that compared to Experiment 1, self-rated proficiency had a smaller effect on naming accuracy in this experiment. This could be because younger bilinguals tend to be overly generous with their self-ratings of proficiency in the nondominant language compared to older bilinguals (see Table 3). In Experiment 2, younger and older bilinguals did not score significantly differently on the MINT in their nondominant language, but younger bilinguals rated their proficiency in that language as significantly higher compared to older bilinguals. Comparing Figures 1 and 3, the range and variability of naming scores in the nondominant language was similar, but bilinguals in Experiment 2 did not use ratings below 3.25.

By contrast, immersion effects seemed similar across experiments. In Experiment 1, naming accuracy in older bilinguals with little immersion experience benefitted more from each additional year of immersion, than in bilinguals with many years of immersion. In Experiment 2, we saw an immersion by age

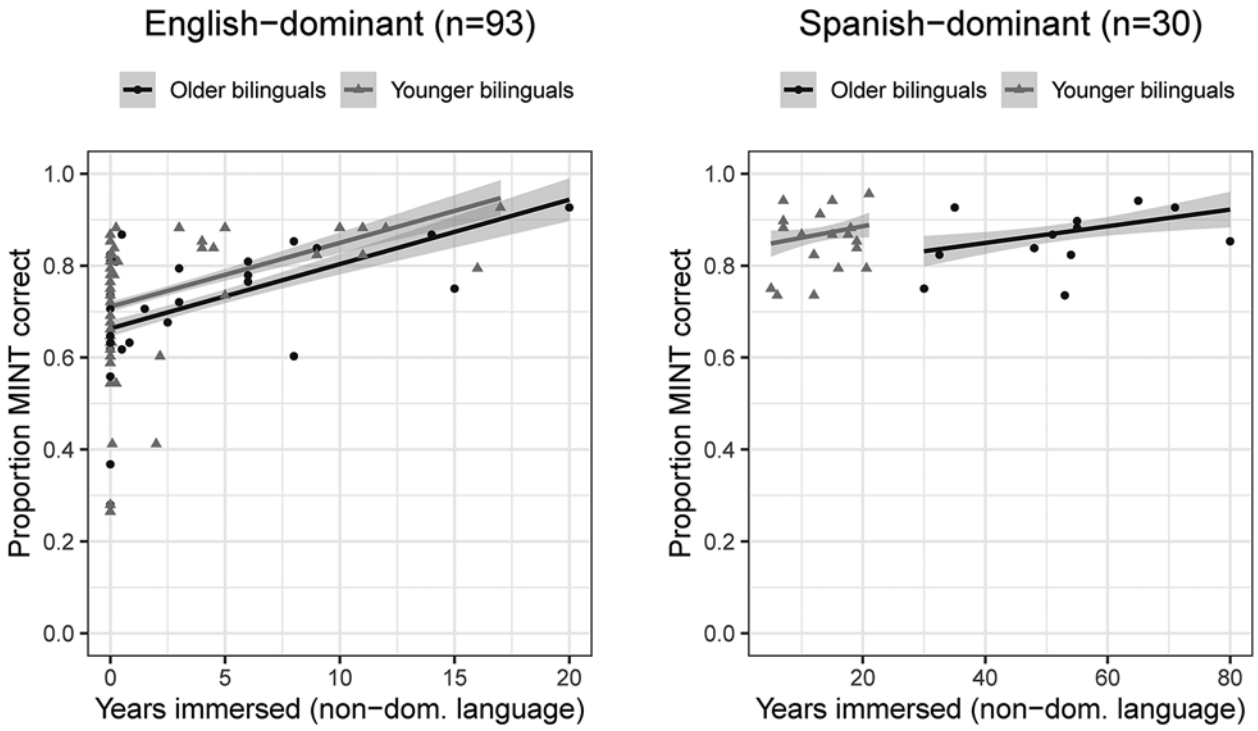

Figure 4. Experiment 2 - Predicted proportion correct on MINT accuracy in the nondominant language by the number of years immersed in the nondominant language, across age groups and for English-dominant bilinguals (left) and Spanish-dominant bilinguals (right). Raw and predicted data are superimposed; grey ribbons show standard error.

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

group interaction. Specifically, young bilinguals, who had relatively fewer years of immersion and more variability in naming scores (including some quite low naming scores), benefitted more from each year of immersion over time than older bilinguals. This could imply that small amounts of immersion have large effects that later asymptote with increased immersion. However, when separating by language dominance groups, the interaction was not significant in English-dominant bilinguals; instead, there was a main effect such that young bilinguals benefitted more than older bilinguals from each year of immersion. By contrast, in Spanish-dominant bilinguals, both young and older bilinguals tended to have high naming scores in the nondominant language (in which they were immersed), and years of immersion did not overlap (which made it impossible to consider age-group by immersion differences). Thus while initially it seemed as if bilinguals with fewer years of immersion benefitted more from each year of immersion than bilinguals with many years of immersion, the interaction in Experiment 2 was likely spurious (an artifact of collapsing across English-dominant and Spanish-dominant bilinguals). We consider alternative explanations in the General Discussion.

Percent current use was a more powerful predictor in Experiment 2. This could be because we had more participants than in Experiment 1. This finding was confirmed when examining the much larger dataset of Tomoschuk et al. (2019). This suggests that more power is needed to detect the smaller effect of percent current use on naming accuracy, compared to self-rated proficiency and immersion.

General Discussion

We observed some systematic differences between groups differing on language dominance and age that revealed inherent inaccuracies based on individual differences in self-rated proficiency levels, and some important differences as to which predictors were important across bilinguals, which we now examine in turn.

Overall, self-rated proficiency in the nondominant language was the strongest predictor of picture naming accuracy in that language, but this predictor also showed variability at each self-rating level, and stricter ratings across age and dominance groups. Self-rated proficiency was especially powerful for predicting MINT accuracy when a small number of functionally monolingual speakers were included (Experiment 1), and relatively less predictive when examining bilinguals alone (Experiment 2). The effects of immersion in the nondominant language on proficiency in that language were stronger initially (in the first years of immersion), and the effect of each immersion year ebbed after many years of immersion. Related to this point, younger bilinguals benefitted more from immersion than older bilinguals. Finally, percent current use of the nondominant language explained unique variance in bilinguals with relatively higher self-rated proficiency level in the nondominant language, but this was a relatively weak predictor perhaps because bilinguals may not be able to accurately report their language usage.

Self-rated proficiency: predictive power and notable inaccuracies

The fact that self-rated proficiency was a stronger predictor of accuracy in both experiments compared to years of immersion and percent current use of the language was surprising considering previous results from Tomoschuk et al. (2019) which showed that within Spanish–English bilinguals, accuracy of self-rated proficiency in the nondominant language varied depending on whether the nondominant language was English or Spanish. However, although self-rated proficiency was the most powerful predictor of naming accuracy in both experiments here, there were also several notable discrepancies between self-rated proficiency and MINT accuracy, reminiscent of the concerns raised by Tomoschuk et al. (2019).

First, both Figures 1 and 3 show tremendous variability in naming scores attained for each self-rating level. Second, older bilinguals had significantly higher naming scores than young bilinguals in the dominant language, although the groups rated their proficiency as comparable (and close to ceiling; see Table 3). Third, this tendency was reversed in the nondominant language, where performance across age groups on the MINT was similar but younger bilinguals' self-rated proficiency was significantly higher than that of older bilinguals. Most notably, Spanish-dominant older bilinguals seemed to self-rate their proficiency level more strictly than English-dominant older bilinguals at the same naming proficiency level; Spanish-dominant bilinguals, who were immersed in English at the time of testing, named more pictures in their nondominant language than did English-dominant bilinguals in Spanish, with the same self-rated proficiency level. This finding was surprising considering that Spanish-dominant bilinguals had a lower education level on average than English-dominant bilinguals (see Table 1), which should have – if anything – decreased naming scores in both languages.

Above, we argued against the possibility that this result might be an artifact of the MINT being easier in English than in Spanish because the MINT was developed specifically for testing multilinguals in these languages. For example, the Boston Naming Test (BNT), which was designed for use in English only, is more difficult in Spanish and so makes bilinguals look more English-dominant than they are when tested with Oral Proficiency Interviews (OPIs; Gollan et al., 2012, Figure 1 and Table 3). In that study, the OPI classified the same group of bilinguals as English-dominant 9.9% of the time, the MINT, 16.0% of the time and the BNT, 28.1% of the time. Thus, there was a 6.1% bias towards English in the MINT (or about 8.6% in older bilinguals), but a much larger 18.2% bias in the BNT. When looking at Figure 1 in the present study, especially at lower self-rated proficiency level, the higher scores in Spanish-dominant bilinguals seem much higher than that attained by English-dominant bilinguals with the same self-rating (e.g., at a rating of 2, Spanish-dominant bilinguals scored about 18% higher than English-dominant bilinguals, and at a rating of 3, they scored about 13% higher). This suggests English bias alone could not explain the systematic between-group difference in naming ability given the same self-rating levels. A more likely possibility is that Spanish-dominant bilinguals might have expected better naming scores for equivalent ratings because they may have had a different standard of comparison. Spanish-dominant bilinguals may have compared themselves to monolingual speakers of English (who could be characterized as hyper-proficient in the one language they know), whereas English-dominant bilinguals may have rated themselves compared to other bilinguals with non-immersed and relatively lower proficiency level in Spanish.

To note, self-rated proficiency might have emerged as a more powerful predictor in Experiment 1 (Figure 1) compared to Experiment 2 (Figure 3) because of higher variability in self-rated proficiency level in Experiment 1, and the exclusion of bilinguals with low education level. In Experiment 2, the self-rating scale for

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

proficiency may have been truncated by our matching process across age groups (see Experiment 2 - Participants section). All young bilinguals were undergraduates, meaning they had at least some college level of education (and education level was significantly correlated with self-rated proficiency in the nondominant language; see Tables B1-B2). When studying bilinguals who all rate themselves as relatively proficient in their nondominant language, self-ratings lose some of their predictive power, possibly making room for percent use to explain unique variance. This could suggest an avenue for improving the predictive power of self-rating scales if, for example, college educated bilinguals may be reluctant to assign themselves very low proficiency rating in any language – they could perhaps be encouraged to do so when appropriate with more detailed definitions of proficiency level or peer group comparison than used in the present study.

Percent current use

In Experiment 1, self-rated percent current use of the nondominant language did not explain unique variance in predicting nondominant language naming accuracy when examined jointly with self-rated proficiency and immersion in older bilinguals. This was true even for older bilinguals with no immersion experience (all English-dominant bilinguals). In Experiment 2, with a larger number of bilinguals with less varied self-rated proficiency levels, and collapsing together young and older bilinguals, percent use did emerge a significant predictor of naming accuracy, with a similar effect size as years of immersion (though still smaller than self-rated proficiency level) when jointly examined. Percent current use did not interact with language dominance or age group. Similarly, a reanalysis of a much larger data set (Tomoschuk et al., 2019) with participants drawn from the same population as Experiment 2 confirmed that both current use and percent use in childhood explained unique variance in predicting naming accuracy in the nondominant language when considered jointly with self-rated proficiency level in young Spanish–English bilinguals.

This suggests that percent current use of the nondominant language is a weaker predictor of picture naming accuracy in the nondominant language compared to self-rated proficiency but can explain unique variance when limited to bilinguals with relatively higher self-rated proficiency level. Importantly, this does not necessarily mean that greater use of the nondominant language fails to increase proficiency level. Instead, bilinguals might not be very good at rating how often they use each language. Another possibility is that some bilinguals with very low proficiency level might use their nondominant language often, but in similar ways repeatedly and without opportunity for improving proficiency. This could especially be true for bilinguals with lower education level (as was the case for some Spanish-dominant older bilinguals in Experiment 1) who remain at a low proficiency level despite being immersed in the nondominant language. Alternatively, percent use could be a weak predictor of naming accuracy because the context of language use, or the number of speakers with which bilinguals communicate in the nondominant language (Gollan et al., 2015), matter more than the sheer number of hours using the language (e.g., Gullifer & Titone, 2020; Gullifer et al., 2021).

Overall, we consistently found that self-rated proficiency was a better predictor of naming accuracy than percent use of the nondominant language (but see Del Maschio et al., 2019[2]). Importantly, there was variability between subjects in percent current use in the full samples of Experiments 1 and 2 (see standard deviations for this variable in Table 1 and Table 3), thus the absence or relatively weak size of effects in the present study should not be attributed to lack of variation in the samples. Percent current use might be more difficult to accurately estimate compared to proficiency level when evaluating it relative to other bilinguals. One factor that might explain this is that we might not be paying close attention to how much we use each language as a bilingual on a daily/weekly/monthly basis.

Immersion across dominance and age groups

Overall, our results suggest that immersion experience consistently explains unique variance in predicting naming accuracy, but that it affected subgroups of bilinguals differently. Specifically, we had a robust interaction in Experiment 1 showing that in older bilinguals, each additional year of immersion had a greater effect when bilinguals had relatively shorter compared to longer periods of immersion. Length of immersion coincided with language dominance, such that bilinguals who reported shorter periods of immersion were English-dominant whereas those with many years of immersion were Spanish-dominant (and currently immersed in the nondominant language). Similarly, in Experiment 2 young bilinguals with relatively few years of immersion benefitted more than older bilinguals with many years of immersion, especially at the low end of the scale from 0-80 years of immersion (see Supplementary Materials, Appendix SD, Figure SA). However, when separated by language dominance, young bilinguals benefitted more than older bilinguals from each year of immersion between 0-20 years (see Figure 4).

Effects of immersion may be more readily observable when relatively few in number years, but curb after many years. Progress on naming accuracy could be getting slower after being immersed for decades because most of the frequently used words will be mastered by then, and the remaining fewer infrequent words, whether learned or not, consist of a smaller proportion of one's overall vocabulary, therefore impacting naming accuracy to a lesser extent. Practically, this means that even short periods of immersion should not be discounted as they can have a strong positive effect on proficiency (e.g., see the difference between being immersed 0 versus 5 years (Figures 2 and A). Immersion might be a more consistent predictor of nondominant language proficiency than percent current use because it is easier to count the number of years immersed than it is to estimate the proportion of time spent using each language. Another factor to consider is if percentage of immersion relative to one's lifetime might better predict naming scores (than raw number of years). However, this does not seem to be the case as, in the present study, older bilinguals had more years of immersion than young bilinguals but were matched in percent of lifetime immersed.

Onset of immersion in the developmental trajectory could be a more relevant indicator of proficiency: immersion during the sensitive period for language acquisition (e.g., Newport et al., 2001) may have a different impact on language proficiency measured in adulthood compared to immersion after this sensitive period, and these effects could vary depending on the length of immersion (note here we found young bilinguals benefitted more than older bilinguals). We did not have information about onset in our data, but it would be an interesting avenue for research, especially since a few years of immersion in younger bilinguals

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

strongly affected proficiency. In addition, context of language use while immersed (e.g., using language for work versus for everyday tasks like grocery shopping) can also influence level of proficiency attained over time. While the role of immersion is recognized to be a key factor in evaluating proficiency (Li et al., 2006; Marian et al., 2007), no previous study has examined the predictive role of self-reported years of immersion on objectively measured proficiency level in bilingual adults. The results we reported suggest this could be a fruitful avenue to pursue in future studies.

Conclusions

To summarize, even when predictive power seems high, comparisons across different types of bilinguals is problematic, and reveals extreme discrepancies in how different individuals interpret rating scales (e.g., eyeballing Figure 1, bilinguals with ratings as low as 1 sometimes named as many as 50% of pictures on the MINT in their nondominant language; conversely bilinguals with a rating of about 5 sometimes named only less than 40% of pictures and at other times named almost all pictures correctly). Bilinguals of different language dominance and age groups can interpret self-rated proficiency scales differently. Therefore, questionnaires need to be designed such that different types of bilinguals evaluate themselves with the same frame of reference: young vs. old, English- vs. Spanish-dominant, immersed vs. not immersed. More work is needed to pinpoint how to do that but, in the meantime, collapsing bilinguals across these variables should be avoided unless objective measures are used.

Practically, small changes might optimize the predictive power of self-report questions for predicting language proficiency. As a group, bilinguals may be less biased and more likely to use the full range of a rating scale if they are given explicit points of reference for comparison (e.g., instead of "like a native speaker" could mention age and years of education "like a highly educated native speaker, with 10+ years of immersion experience"). Finally, questionnaires could be improved by focusing on questions that are easier to answer in an objectively accurate manner: for example, counting years of immersion seems to be a powerful predictor, even (and in some cases especially) in bilinguals with relatively little immersion experience, and may be less prone to subjective bias than rating one's proficiency level or estimating percent current use of a language.

Supplementary Materials. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728924000038

Table SA.1
Correlations between variables in the dominant and nondominant language and MINT scores, for English-dominant participants of Experiment 1 (n = 36). Participant information is listed in Table 1.
Table SA.2
Correlations between variables in the dominant and nondominant language and MINT scores, for Spanish-dominant participants of Experiment 1 (n = 32). Participant information is listed in Table 1.
Table SB.1a.
Nondominant language naming accuracy predicted jointly from self-rated proficiency, years immersed and percent current use of the nondominant language.
Table SB.1b.
Nondominant language naming accuracy predicted jointly from self-rated proficiency and percent current use of the nondominant language, in bilinguals with no immersion experience.
Table SB.2.
Nondominant language naming accuracy predicted from self-rated proficiency, dominance group, and their interaction.
Table SB.3.
Nondominant language naming accuracy predicted from immersion, dominance group, and their interaction.
Table SB.4.
Nondominant language naming accuracy predicted from immersion, dominance group, and their interaction, in bilinguals with 12 or more years of education.
Table SB.5.
Nondominant language naming accuracy predicted from percent use, dominance group, and their interaction.
Table SC.1
Correlations between variables in the dominant and nondominant language and MINT scores, for older adults in Experiment 2, collapsed across dominance groups (n = 41). Participant information is listed in Table 2.
Table SC.2
Correlations between variables in the dominant and nondominant language and MINT scores, for younger adults in Experiment 2, collapsed across dominance groups (n = 82). Participant information is listed in Table 2.
Table SD.1.
Nondominant language naming accuracy predicted jointly from self-rated proficiency, years immersed and percent current use of the nondominant language.
Table SD.2.
Nondominant language naming accuracy predicted from self-rated proficiency, age group, and their interaction.
Table SD.3.
Nondominant language naming accuracy predicted from immersion, age group, and their interaction.
Figure SA.
Experiment 2 - Predicted proportion correct on MINT accuracy in the nondominant language by the number of years immersed in the nondominant language, across age groups but collapsing across language dominance (note that the apparent interaction is spurious; see Figure 4). Raw and predicted data are superimposed; grey ribbons to show standard error.
Table SD.3a.
Nondominant language naming accuracy predicted from immersion, age group, and their interaction, in English-dominant bilinguals only.
Table SD.3b.
Nondominant language naming accuracy predicted from immersion, age group, and their interaction, in Spanish-dominant bilinguals only.
Table SD.4.
Nondominant language naming accuracy predicted from percent use, age group, and their interaction.

Acknowledgements. This research was supported by grants from the National Institute on Deafness and Other Communication Disorders (DC011492), the National Institute on Aging (AG076415; P30 (AG062429), and the National Science Foundation BCS1923065.

We thank Mayra Murillo and Rosa Montoya for data coding, and Dalia Garcia for comments on an earlier version of this paper.

Data availability. The data, materials and script that support the findings of this study are openly available on OSF at https://osf.io/732p9.

Competing interests. The author(s) declare none.

Ethics statement. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Notes

1 The percentages were derived using a percentage increase calculation: 100 x ([Spanish score - English score/Spanish score). For example, for the Spanish-dominant bilingual group referenced whose average self-rating was 9.4/10 in Spanish and 6.9/10 in English, the calculation is: 100 x ([9.4-6.9]/6.9) = 36.23%. Percentages were rounded to the closest whole number.

2 Percent use might be a more reliable and important predictor in different types of bilinguals and with different measures. For example, in Del

Maschio et al. (2019), participants learned the L2 (which was also the nondominant language) later in childhood, were not immersed, and language usage was measured in terms of raw number of hours per day.

References

Anderson, J. A., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*, 250–263.

Ardila, A., Roselli, M., & Puente, A. E. (1994). *Neuro-psychological evaluation of the Spanish speaker*. New York: Plenum Press.

Artiola i Fortuny, L., Heaton, R. K., & Hermosillo, D. (1998). Neuropsychological comparisons of Spanish-speaking participants from the US–Mexico border region versus Spain. *Journal of the International Neuropsychological Society*, *4*(4), 363–379.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1), 1–48. doi:10.18637/jss.v067.i01.

Birdsong, D. (2014). Dominance and age in bilingualism. *Applied Linguistics*, *35*(4), 374–392.

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults?. *Journal of Memory and Language*, *30*(5), 542–579.

Carroll, S. E. (2017). Exposure and input in bilingual development. *Bilingualism: Language and Cognition*, *20*(1), 3–16.

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—Simulation and Computation*, *39*(4), 860–864.

De Cat, C. (2021). Socioeconomic status as a proxy for input quality in bilingual children? *Applied Psycholinguistics*, *42*(2), 301–324.

Del Maschio, N., Sulpizio, S., Toti, M., Caprioglio, C., Del Mauro, G., Fedeli, D., & Abutalebi, J. (2019). Second language use rather than second language knowledge relates to changes in white matter microstructure. *Journal of Cultural Cognitive Science*, *4*, 165–175.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Bloomington, MN: Pearson Assessments.

Garcia, D. L., & Gollan, T. H. (2022). The MINT Sprint: Exploring a fast administration procedure with an expanded Multilingual Naming Test. *Journal of the International Neuropsychological Society*, *28*(8), 845–861.

Gasquoine, P. G., & Gonzalez, C. D. (2012). Using monolingual neuropsychological test norms with bilingual Hispanic Americans: Application of an individual comparison standard. *Archives of Clinical Neuropsychology*, *27*(3), 268–276.

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*(2), 213–237.

Gollan, T. H., & Goldrick, M. (2016). Grammatical constraints on language switching: Language control is not just executive control. *Journal of Memory and Language*, *90*, 177–199.

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*(3), 787–814.

Gollan, T. H., Starr, J., & Ferreira, V. S. (2015). More than use it or lose it: The number-of-speakers effect on heritage language proficiency. *Psychonomic Bulletin & Review*, *22*, 147–155.

Gollan, T. H., Stasenko, A., & Salmon, D. P. (in press). Which language is more affected in bilinguals with Alzheimer's Disease? Diagnostic sensitivity of the Multilingual Naming Test. *Neuropsychology*.

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*(3), 594–615.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149.

Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, *23*(2), 283–294.

Gullifer, J. W., Kousaie, S., Gilbert, A. C., Grant, A., Giroud, N., Coulter, K., Klein, D., Baum, S., Phillips, N., & Titone, D. (2021). Bilingual language experience as a multidimensional spectrum: Associations with objective and subjective language proficiency. *Applied Psycholinguistics*, *42*(2), 245–278.

Hartig, F. (2022). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.4.5. https://CRAN.R-project.org/package=DHARMa

Hoff, E. (2018). Bilingual development in children of immigrant families. *Child Development Perspectives*, *12*(2), 80–86.

Hoff, E., Welsh, S., Place, S., Ribot, K., Grüter, T., & Paradis, J. (2014). Properties of dual language input that shape bilingual development and properties of environments that shape dual language input. In J. Paradis & T. Grüter (Eds.), *Input and experience in bilingual development*, *13*, 119–140. John Benjamins Publishing Co.

Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, *8*(3), 229–249.

Hurtado, N., Grüter, T., Marchman, V. A., & Fernald, A. (2014). Relative language exposure, processing efficiency and vocabulary in Spanish–English bilingual toddlers. *Bilingualism: Language and Cognition*, *17*(1), 189–202.

Kastenbaum, J., Bedore, L., Pena, E., Sheng, L., Mavis, I., Sebastian, R., Vallila-Rohter, S., & Kiran, S. (2018). The influence of language combination and proficiency on bilingual lexical access. *Bilingualism Language and Cognition*, 1–31.

Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2020). The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition*, *23*(5), 945–950.

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

Li, P., Zhang, F. A. N., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, *17*(3), 673–680.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*(12), 1507–1515.

Lorenzen, B., & Murray, L. L. (2008). Bilingual aphasia: A theoretical and clinical review. *American Journal of Speech-Language Pathology*, *17*, 299–317

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621.

Lynch, A., Klee, C. A., & Tedick, D. J. (2001). Social factors and language proficiency in postsecondary Spanish immersion: Issues and implications. *Hispania*, 510–524.

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*(4), 940–967.

Nanchen, G., Abutalebi, J., Assal, F., Manchon, M., Démonet, J. F., & Annoni, J. M. (2017). Second language performances in elderly bilinguals and individuals with dementia: The role of L2 immersion. *Journal of Neurolinguistics*, *43*, 49–58.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. *Language, brain and cognitive development: Essays in honor of Jacques Mehler*, 481–502.

Olson, D. J. (2023). A systematic review of proficiency assessment methods in bilingualism research. *International Journal of Bilingualism*, doi: 13670069231153720.

Paplikar, A., Alladi, S., Varghese, F. A., Mekala, S., Arshad, F., Sharma, M., Saroja, A.O., Divyaraj, G., Dutt, A., Ellajosyula, R., Ghosh, A., Iyer, G.K., Sunitha, J., Kandukuri, R., Kaul, S., Khan, A.B., Mathew, R., Menon, R.N., Nandi, R., Narayanan, J., Nehra, A., Padma, M.V., Pauranik, A., Ramakrishnan, S.,

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press

Sarath, L., Shah, U., Tripathi, M., Sylaja, P.N., Varma, R.P., Verma, M., & Vishwanath, Y. (2021). Bilingualism and its implications for neuropsychological evaluation. *Archives of Clinical Neuropsychology*, *36*(8), 1511–1522.

Rivera Mindt, M., Arentoft, A., Kubo Germano, K., D'Aquila, E., Scheiner, D., Pizzirusso, M., Sandoval, T. C., & Gollan, T. H. (2008). Neuropsychological, cognitive, and theoretical considerations for evaluation of bilingual individuals. *Neuropsychology Review*, *18*, 255–268.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, *15*(1), 1–20.

Rothman, J., Bayram, F., DeLuca, V., Alonso, J. G., Kubota, M., & Puig-Mayenco, E. (2023). Defining bilingualism as a continuum. In G. Luk, J. A. E. Anderson, & J. G. Grundy (Eds.). *Understanding Language and Cognition through Bilingualism: In honor of Ellen Bialystok* (pp. 38–67). John Benjamins Publishing Co.

Schmid, M. S., & Yılmaz, G. (2018). Predictors of language dominance: An integrated analysis of first language attrition and second language acquisition in late bilinguals. *Frontiers in psychology*, *9*, 1306.

Schrauf, R. W. (2009). English use among older bilingual immigrants in linguistically concentrated neighborhoods: Social proficiency and internal speech as intracultural variation. *Journal of Cross-Cultural Gerontology*, *24*, 157–179.

Sheng, L., Lu, Y., & Gollan, T. H. (2014). Assessing language dominance in Mandarin–English bilinguals: Convergence and divergence between subjective and objective measures. *Bilingualism: Language and Cognition*, *17*(2), 364–383.

Sinkeviciute, R., Brown, H., Brekelmans, G., & Wonnacott, E. (2019). The role of input variability and learner age in second language vocabulary learning. *Studies in Second Language Acquisition*, *41*(4), 795–820.

Stasenko, A., Kleinman, D., & Gollan, T. H. (2021). Older bilinguals reverse language dominance less than younger bilinguals: Evidence for the inhibitory deficit hypothesis. *Psychology and Aging*, *36*(7), 806.

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426–445.

Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition*, *22*(3), 516–536.

Treffers-Daller, J., & Silva-Corvalán, C. (Eds.). (2016). *Language dominance in bilinguals: Issues of measurement and operationalization*. Cambridge University Press.

Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan : Factors moderating language proficiency* (pp. 103–122). American Psychological Association.

Williams, K. T. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Service.

https://doi.org/10.1017/S1366728924000038 Published online by Cambridge University Press