HHS Public Access

Author manuscript

*Int J Biling Educ Biling*. Author manuscript; available in PMC 2018 April 16.

Published in final edited form as:

*Int J Biling Educ Biling*. 2014 ; 17(1): 90–110. doi:10.1080/13670050.2012.743960.

# The relation between language experience and receptive-expressive semantic gaps in bilingual children

**Todd A. Gibson**[*], **Elizabeth D. Peña**, and **Lisa M. Bedore**

Communication Sciences and Disorders, University of Texas at Austin, Austin, TX, USA

## Abstract

The purpose of the current study was to explore the influence of language experience on the presence of the receptive-expressive gap. Each of 778 Spanish-English bilingual children screened pre-kindergarten in Utah and Texas were assigned to one of five language experience groups, ranging from functionally monolingual to balanced bilingual. Children's scores from the language screener semantics subtest administered in both Spanish and English were standardized, and receptive and expressive semantic scores were compared. Children presented with a meaningful gap between receptive and expressive semantic knowledge in English but not Spanish. This gap increased as target-language exposure decreased. Results indicate that current language experience plays a dominant role in influencing the appearance and magnitude of the receptive-expressive gap.

### Keywords

childhood bilingualism; bilingualism; second language acquisition; sequential bilingualism

Differences between receptive and expressive language development are well documented. In the early stages of language development, children understand more words than they produce (Benedict 1979; Fenson et al. 1994; Nelson et al. 1978; however, see Gathercole 1988 for apparent exceptions to this rule). The pace of development for the modalities differs; children demonstrate an understanding of words at around nine months of age but first produce words at about 12 months (Benedict 1979). Receptive language appears to play an important role in the development of expressive language, but expressive language does not influence development of receptive language to the same degree (Bates et al. 1979; Benedict 1979; Goldin-Meadow, Seligman, and Gelman 1976). Thus, receptive and expressive language skills appear to be dissociated, at least to some extent (Bates, Dale, and Thal 1995; Bates et al. 1989; Chiappe, Chiappe, and Gottardo 2004; Chiat and Roy 2008; Watt, Wetherby, and Shumway 2006).

Despite these differences, receptive and expressive language skills are highly correlated (Mashburn et al. 2009), such that a relatively high score on a test of receptive language predicts a relatively high score on a test of expressive language. Recent studies, however, have illustrated that this prediction does not necessarily hold for bilingual children with

[*]Corresponding author. todd.gibson@austin.utexas.edu.

typical development (TD) (Gibson et al. 2012; Kan and Kohnert 2005; Miccio et al. 2003; Oller and Eilers 2002; Yan and Nicoladis 2009). Even after controlling for the inherent differences in difficulty between receptive and expressive language tasks, some aspects of expressive language performance lag behind what would be predicted on the basis of receptive language performance (Gibson et al. 2012). We use the term receptive-expressive gap to describe this unexpected discrepancy.

In the current study, we investigate the receptive-expressive gap in the performance of Spanish-English bilingual children on tests of semantic knowledge in both of their languages relative to their exposure to each language. The results inform our understanding of second language acquisition and aid in the interpretation of language test performance by bilingual children.

## Comparing receptive and expressive language performance

Most tests of receptive language require individuals to identify, typically via pointing to a picture or object, a word or words uttered by the tester, while tests of expressive language often require individuals to generate spoken words when presented with an image or object. The former putatively is recognized to be easier than the latter because success with expressive language tasks appears to require more fine-grained phonological representations than does success with receptive language tasks (Bates 1993). Furthermore, individuals might be successful in performing receptive language tasks by employing a guessing strategy (Leonard 2009) that is not likely to be successful during expressive language tasks (Clark and Hecht 1983). One technique that allows for a valid comparison between receptive and expressive language performance despite their differences in difficulty is measurement standardization. When raw scores are converted to standard scores, the inherent differences between the two modalities are accounted for mathematically (assuming similar norming samples). This conversion allows for comparison between receptive and expressive performance beyond those that are already documented. When comparing standard scores of receptive and expressive language, there is no expectation that there will be systematic differences between standardized test outcomes. Standard score transformations are based on Z-scores where the mean is typically 100 with a standard deviation (SD) of 15.

It should be noted that receptive-expressive comparisons occur at both the individual and group levels. At the group level, the average for both receptive and expressive standard scores for a test norming group would be 100 with an SD of 15, such that no discrepancy exists between receptive and expressive standard scores. At the individual level, any single participant in the test norming sample may have scores that differ from the average of the group to which they belonged, with some individual children scoring higher in the receptive or in the expressive domain. Several standardized English-only language assessments report receptive-expressive gaps ranging from 7 to 14 standard score points for roughly 12.5% of their monolingual samples, including Oral and Written Language Scales (Carrow-Wool-folk 1995), the Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4; Semel, Wiig, and Secord 2003), and the *Preschool Language Scales – Fifth Edition* (*PLS-5*; Zimmerman, Steiner, and Pond 2011). Similar receptive-expressive discrepancies have been identified in co-normed single-word vocabulary tests such as the Peabody Picture

Vocabulary Test – Fourth Edition (PPVT-4; Dunn and Dunn 2007) compared to the Expressive Vocabulary Test – Second Edition (EVT-2; Williams 2007) and the Receptive One Word Picture Vocabulary Test – Third Edition (ROWPVT-3; Brownell 2000b) compared to the Expressive One Word Picture Vocabulary Test – Revised (EOWPVT-R; Brownell 2000a). This suggests that some amount of receptive-expressive discrepancy at the individual level is to be anticipated, even when tests are co-normed.

Given individual variation among standardized receptive and expressive performance, it is important to establish a data-based criterion for a true expressive-receptive gap. Based on the above review of standardized language measures, we consider a receptive-expressive gap to be meaningful when receptive standard scores are higher than expressive standard scores by 10 or more standard score points at the individual level or if there is a statistically significant difference at the group level. We reason that because such a small percentage of norming samples show this level of discrepancy, anything at this level or beyond can be considered atypical.

## Receptive-expressive gap in bilinguals

The receptive-expressive gap has been identified in bilingual individuals with TD (see Gibson et al. 2012 for a review). In some cases, the first language (L1) receptive-expressive gap is larger in magnitude than the same individuals' gap in the second language (L2). This phenomenon can be identified in bilinguals across a variety of languages and age groups, including Spanish-English bilingual adults (Muñoz and Marquardt 2003), preschoolers (Miccio et al. 2003), and school-age children (Gibson et al. 2012; Oller and Eilers 2002), as well as Hmong-English bilingual school-age children (Kan and Kohnert 2005) and French-English bilingual school age children (Yan and Nicoladis 2009).

For bilingual individuals, there appears to be a small but robust receptive-expressive gap in the L2 and a larger robust gap in L1. A review of the literature yielded 11 articles in which a receptive-expressive gap in vocabulary was identifiable via standardized single-word test scores, $M = 100$, SD = 15, for both Spanish and English (see Table 1). The average receptive-expressive gap in English (in all cases the L2) for the 11 published papers was 3.45 standard score points. By comparison, the average receptive-expressive gap in L1 (Spanish) was 15.63 standard score points, four times greater than the gap in L2 across these same investigations. The magnitude of the receptive-expressive gap ranged from one (expressive greater than receptive) to eight standard score points in English and from 3 to 28 standard score points in Spanish.

There are at least two explanations offered in the literature for the receptive-expressive gap in bilingual speakers. The first explanation, related to bilingual speakers' L1, is that speakers *inhibit or suppress* the activation of their first language as they focus their efforts on acquiring an L2 (Gibson et al. 2012; Linck, Kroll, and Sunderman 2009). The second explanation, related to bilingual speakers' L2, is the *weaker links hypothesis* (Gollan et al. 2008), which suggests that bilingual speakers have less well-established links between words and concepts associated with divided input in their two languages.

The explanation for a suppression mechanism argues that the L1 is suppressed in an L2 learning context in order to free cognitive space to learn the L2 (Gibson et al. 2012; Linck, Kroll, and Sunderman 2009). Evidence for suppression is found in the Linck et al. study in which university students learning Spanish in an immersion context generated fewer items in English (their first and primary language) compared to those learning Spanish in the United States, where they spoke English most of the time. Alternatively, what appears to be suppression instead might be a relative deactivation of the L1 while L2 is highly activated in the L2 learning context. A receptive vocabulary task might be performed successfully in the presence of L1 suppression or deactivation, but an expressive vocabulary task might not and this could result in a receptive-expressive gap.

Another explanation, related to speakers' L2, depends on bilingual speakers' experience, i.e. practice, in either of their languages. The linguistic experience of bilingual speakers is divided between two languages, resulting in less practice in either language compared to their monolingual peers (Oller, Pearson, and Cobo-Lewis 2007). This divided experience might result in weaker representations of both morphosyntactic (Bedore, Cooperson, and Boerger 2011) and lexical targets within each language. Gollan et al. (2008) proposed the *weaker links hypothesis* in which bilingual speakers have difficulty with expressive language tasks because the link between semantic and phonological representations is weak due to limited practice within each language compared to monolingual peers. With increased practice, the quality of the links improves, resulting in better performance on expressive tasks (Gollan et al. 2005).

Support for the explanation that the receptive-expressive gap depends on language experience comes from studies of the repetition of nonwords like *woogalamic*. Some authors have argued that nonword repetition is primarily a measure of the quality of an individual's phonological representations (Bowey 1996, 1997; Metsala 1999; however, see Gathercole 2006, for an argument that nonword repetition is primarily a measure of phonological short-term memory). Several studies with bilingual speakers have demonstrated that as knowledge of a language improves, so does performance on nonword repetition tasks, at least until a ceiling is reached (Cheung 1996; Masoura and Gathercole 2005; Thorn and Gathercole 1999; Thorn, Gathercole, and Frankish 2002). If nonword repetition is a measure of the quality of phonological representations, then it is reasonable to conclude that, due to divided input, the quality of phonological representations for bilinguals will be weaker than those for monolingual peers. Weak phonological representations in the bilingual's L2 might be sufficient to perform receptive but not expressive language tasks in L2 (Bates 1993), resulting in a receptive-expressive gap. Children presumably will show a larger gap in the language in which they have the least experience, which is often the L2.

A similar process appears to apply to monolingual speakers since there is a positive correlation between an individual's language input and the size of that individual's vocabulary. Hart and Risley (1995) found that children of mothers who spoke more frequently to their children had larger receptive and expressive vocabularies than children of mothers who spoke less frequently to their children. Whether due to the quantity of mothers' speech, limited access to linguistic activities (Hoff and Tian 2005), or deafness (Mann and

Marshall forthcoming), limited language experience might result in persisting weakness in phonological representations.

## Role of language experience in the receptive-expressive gap

As suggested by the weaker links hypothesis, language exposure plays a significant role in bilinguals' performance on language measures. Both cumulative exposure and current exposure play a role in first and second language performance. For English language learners in the United States, longer exposure to the second language is related to a shift in dominance from L1, Spanish, to L2, English (Kohnert, Bates, and Hernandez 1999). Similarly, Hammer, Lawrence, and Miccio (2008) found that Spanish-English bilingual children's receptive standard scores in English improved after entry to the English-speaking school system, but Spanish standard scores plateaued or declined. Focusing on current use and exposure to English and Spanish, Peña et al. (2011) found that performance on both receptive and expressive semantic and syntactic measures were positively associated with input and output in the target language. In a study examining performance relative to both current exposure and first year of exposure to English, Bedore et al. (2012) found that both predicted current performance on language measures, but current exposure to the target language was a better predictor of language dominance than the age of first regular exposure to English in preschool-age Spanish-English bilingual children.

Children differ not only in the amount of language they experience but also in the quality of the language they experience (Unsworth 2008). Native speakers of a language are likely to provide language input that is highly target-like, while second language speakers might provide low target-like input (Cornips and Hulk 2008). As noted by Oller and Eilers (2002), access to native-speaking peers who can provide models that are highly target-like is essential to second language learning (Gamez and Levine forthcoming). If parents or peers provide low target-like input, the result may be degraded or underspecified phonological representations that might contribute to the receptive-expressive gap.

Divided language experience might result in cross-linguistic competition for lexical access, which could contribute to a receptive-expressive gap. Gollan et al. (2007) and Gollan and Acenas (2004) found that during picture-naming tasks, bilingual speakers report more instances in which they know the target word but cannot access it (called a tip of the tongue state) than monolingual speakers. Similar results were found by Ivanova and Costa (2008). Gollan et al. (2007) argued that the greater number of tip of the tongue states for bilinguals was due to competition between languages. Because there is only one possible correct answer in a picture-naming task, competition might block access to the correct answer and increase the likelihood of failure on an item. However, on a task that allows multiple possible correct answers, one of a variety of possible correct answers might be accessed, even in the face of competition. Thus, the effects of competition or suppression might not be as pronounced in a more ecological expressive task, i.e. one which accepts multiple correct answers.

In addition to the quantity and quality of language experience, variables such as age, maternal education, and sex might play a role in the receptive-expressive gap. It is well

established that older children perform better than younger children on measures of language knowledge, and children of more highly educated mothers are more likely to have better language skills than children of less educated mothers (Hoff-Ginsberg 1998; Magnuson et al. 2009). In addition, several studies report that girls have an advantage over boys in vocabulary learning (Galsworthy et al. 2000; Huttenlocher et al. 1991), and this seems to be the case across languages (Eriksson et al. 2011); therefore, sex differences might also influence the receptive-expressive gap.

## Research questions

As reviewed above, the receptive-expressive gap in the performance of bilingual children with TD might be explained by a suppression mechanism or by weak phonological representations, depending on whether the gap appears in L1 or L2, respectively. We reasoned, however, that both a suppression mechanism and weak phonological representations might be involved in the performance of bilingual children on tests of receptive and semantic knowledge. Furthermore, we posited that with a more precise measure of current language experience than the demographic proxies used by previous researchers (see Gibson et al. 2012), we might be able to identify incremental relationships between language experience and the receptive-expressive gap. Based on these considerations, we formulated two research questions:

1. What is the relationship between the receptive-expressive gap and language experience, based on amount of current experience in English and Spanish?
2. What variables best predict the presence of the receptive-expressive gap?

In the current study, we addressed these questions by providing three unique data points that were not available in earlier studies. First, we were able to ascertain hour-by-hour language experience information using a parent questionnaire. Second, we used a screening test that assessed semantic knowledge more broadly than single-word vocabulary. Third, receptive and expressive modalities were co-normed.

## Method

### Participants

Participants were drawn from the screening phase of data collection from a longitudinal study of diagnostic markers in bilingual children (see Bohman et al. 2010; Peña et al. 2011). Following approval from the review boards of two universities, children participating in this study were recruited from 12 schools with large Latino populations in three school districts in Northern Utah and Central Texas. English-speaking only, Spanish-speaking only, and Spanish-English bilingual speakers were recruited. Over 85% of the caregivers of invited participants completed consent forms. The targeted districts utilized a variety of educational models, including English-only classes, ESL classes, and transitional bilingual classes (Bohman et al. 2010). In order to test before formal education began, children were screened either in the spring of the pre-kindergarten year or late summer during kindergarten enrolment for the fall semester.

A total of 1192 children were screened. From this set, we excluded 151 children whose parents did not identify them as Hispanic or provide data for all variables of interest. We further excluded 67 children who were either identified with language impairment or suspected of having language impairment due to low performance in their stronger language. In addition, we excluded 207 children below the age of five years because a screener with only one receptive item was used for that age group. Eleven children had both low language scores and missing data points. This process reduced the sample size to 778. The average age was five years five months; 50% were girls.

## Measures

**Caregiver interviews—**Caregivers were interviewed (Gutiérrez-Clellen and Kreiter 2003) in their preferred language (Spanish or English) either in person or by telephone. Questions focused on the children's education history as well as the level of the mother's education, which was included because of its reported relationship with child language development (Dollaghan et al. 1999; Hoff and Tian 2005). Based on Hollingshead's (1975) education categories, mother's education was scored on a scale of one to seven (0 = no formal education; 1 = less than 7th grade education; 2 = 9th grade education; 3 = partial high school; 4 = high school graduate; 5 = specialized training or partial college; 6 = college degree; 7 = graduate degree). Mothers of children in the current study fell into Hollingshead education categories that ranged from 0 to 7, with a mode of 4, an average of 3.12, and a Standard Deviation (SD) of 1.61.

The questionnaire was additionally comprised of an hour-by-hour report of typical language exposure for weekdays and weekends. For any given hour, parents reported the activity that was likely to occur, the participants in the activity, the language spoken to/around the child (input), and the language the child produced (output). Parents could report the language of the activity as English, Spanish, or both. Children's scores in each language are highly correlated to the language in which their interlocutors speak and the language in which children speak (see Bedore et al. 2012; Bohman et al. 2010; Peña et al. 2011). Average input and output over a typical week in Spanish and English was calculated. Input plus output scores were then averaged to create a language experience composite score in each language.

The language experience composite scores were used to categorize children into five language groups consistent with previous studies of bilingual language experience (Peña et al. 2011) and to create a norming sample. Children whose experience in English was 20% or less and in Spanish 80% or greater were categorized as functionally monolingual in Spanish (FMS). Children whose English experience was 20–40% and whose Spanish experience was 60–80% were categorized as bilingual but dominant in Spanish (BDS). Children whose English experience and Spanish experience were 40–60% were categorized as balanced bilingual (BL). Children whose English experience was 60–80% and Spanish experience 20–40% were categorized as bilingual but dominant in English (BDE). Finally, children were categorized as functionally monolingual in English if their English experience was 80% or greater and Spanish experience 20% or less (FME). Classification by language experience group allowed us to examine the receptive-expressive gap in each language

holding relative experience in Spanish and English at the group level constant. The average English language experience for the participants was 50% with an SD of 32% (see Table 2 for demographic information by language experience group).

Most participants were exposed to some Spanish from birth, while other children were exposed to English either from birth or at a later date. Our language experience categorizing scheme did not differentiate children who were simultaneous bilinguals from those who were sequential bilinguals. However, in addition to the current language exposure information, the questionnaire asked for the age at which the child was first regularly exposed to English. The average age of first English exposure for the current participants was 2.12 years, SD = 2.08. The distribution of age of first English exposure is included in Table 2.

Children were categorized into age groups of either 60–65 months or 66–81 months for the purpose of norming. Means and standard deviations from each group were used in standardizing test scores, thus controlling for age. Only 37 (4.7%) of the children were older than 72 months.

**Bilingual English-Spanish oral screener—**The English and Spanish semantics and morphosyntax subtests of the *Bilingual English-Spanish Oral Screener* (BESOS, Peña, Bedore, et al. in preparation) were administered (see Peña et al. 2011). The BESOS consists of four subtests of morphosyntax and semantics in English and Spanish. We focused on the semantics subtests in the current analysis. The BESOS items were drawn from the comprehensive Bilingual English Spanish Assessment (BESA; Peña, Gutierrez-Clellen, et al. in preparation). Items selected for the Spanish and English subtests are equated on the basis of difficulty and therefore not direct translations. Specifically, items selected for the BESOS are those that are difficult for preschool-age children with risk for language impairment (Peña et al. 2011), and they are *sensitive* to language development across different levels of language exposure and use (Bohman et al. 2010).

The BESOS semantic subtests target semantic knowledge via various expressive and receptive tasks (see below) in English and Spanish. The English semantic subtest includes 11 items (four receptive, seven expressive), while the Spanish version contains 12 items (five receptive, seven expressive). Analyses demonstrate a significant correlation between BESOS and the corresponding BESA, specifically for Spanish semantics, $r(172) = 0.855$, $p < 0.001$ and English semantics, $r(185) = 0.887$, $p < 0.001$ (Summers et al. 2010).

The BESOS semantics subtests were not designed to test receptive and expressive semantics separately; therefore, we reviewed each item and designated it as either receptive or expressive based on task requirements. Several types of semantic tasks were used in this study. These have been described in Peña, Bedore, and Rappazzo (2003) as associations (analogies), characteristic properties (descriptions), categorization, functions, linguistic concepts, and similarities and differences. Receptive semantic tasks required children to point to a target item or items in response to a verbal prompt, while expressive semantics tasks required the child to produce language when answering a question. For example, in a receptive task targeting similarities and differences in English, children were asked to point

to the piñata that was different from the others. In an expressive task targeting similarities and differences, children were shown a picture and asked ‘What is different about these jeans?’ to which they were required to verbalize an answer. This question allowed for a variety of correct answers, such as ‘The sizes are different,’ ‘Some are bigger than the others,’ ‘Some are smaller than the others,’ ‘Some are longer than the others,’ etc. Inaccurate responses, i.e. pointing to the non-target picture or verbalizing an incorrect response, were scored as zero, and accurate responses were scored as one. Total score for each test was the number of items answered correctly.

We used the scores of the functional monolingual speakers as the comparison groups (those exposed 80% or more to the target language). Raw score averages were derived for both receptive and expressive tasks for each age. We standardized scores in the following way in order to allow for comparisons between receptive and expressive scores consistent with standardization procedures (Allen and Yen 1979). First, the means and standard deviations of the receptive and expressive raw scores from the comparison groups (FMS and FME) were calculated for each age. Second, these means and standard deviations were used to develop Z-scores for each age group. Third, the Z-scores were then transformed into standard scores with a mean of 100 and an SD of 15. This allowed us to control for age and for differences in task difficulty. After development of these norms, we transformed all children’s raw scores for expressive and receptive semantics in English and Spanish into standard scores based on age. These standard scores were used in our planned analyses. This transformation based on monolingual norms allows us to examine potential receptive-expressive differences related to level of bilingual exposure.

## Procedures

Tests were administered by bilingual speech-language pathologies (SLPs) or students enrolled in programs in communication sciences and disorders. Students were directly supervised by certified bilingual SLPs. Screening took place at schools during regular school hours. Attempts were made to reduce distractions during testing. If children failed to respond after 5 consecutive items, the subtest was discontinued. However, if children responded at all, even incorrectly, the subtest was continued. Responses were written down verbatim and coded as correct/incorrect. In the current analysis, given the purpose to examine receptive vs. expressive knowledge in a given language, only correct responses in the target language were credited. Thus, responses given in the non-target language were scored as zero. We provide the raw scores by modality and age group in Table 3.

## Statistical analysis

Descriptive statistics were calculated for test scores and demographic variables of interest. A repeated measures analysis of variance (ANOVA) was performed to test differences between language groups across test types and languages and to analyze the interactions between them. With the ANOVA results, we report effect sizes, which measure the magnitude of the association between the dependent variable and independent variables. We selected partial eta squared ( $\eta_p^2$) as the effect size measure since it only includes variance from the target variable and not variance produced by other independent variables (Pierce, Block, and Aguinis 2004). This is similar to eta squared for sample sizes over 50 (Stevens 2002). No

guidelines exist for interpretation of the partial eta squared, but because this statistic is based on the general linear model, guidelines are adopted from correlation analysis. Thus, effect sizes between 0.00 and 0.10 are considered negligible; those between 0.10 and 0.25 are considered small; those between 0.25 and 0.50 are considered moderate; those between 0.50 and 0.80 are considered large; and those between 0.80 and 1 are considered very large. This was followed by paired *t*-tests within language and within group to determine if the receptive-expressive gap was statistically significant. Finally, multiple regression analyses with age, year of first exposure, language experience, mother education, and sex as the predictor variables were performed to predict the appearance of the receptive-expressive gap in both Spanish and English.

## Results

Our first research question explored how language experience and language of testing were related to the receptive-expressive gap (standard scores for each language group are provided in Table 4 for Spanish and Table 5 for English). We wanted to determine if performance on semantic testing differed by modality (receptive and expressive) and if there were differences related to test language (Spanish or English) and level of language exposure (Language Groups FME to FMS). A repeated measures ANOVA with Modality and Test Language as the within-subject variables and Language Group as the between-subjects variable was conducted.

### Main effects for modality, test language, and language group

Results showed a main effect for Modality $F(1777) = 109.65$, $p<0.0001$, with a small effect size, $\eta_p^2 = 0.124$. Children's standard scores were higher on Spanish and English receptive tasks combined, $M = 91.84$, SD = 12.45, than expressive scores, $M = 86.32$, SD = 14.65.

There was also a main effect for Test Language, $F(1777) = 22.59$, $p<0.0001$, but its effect was negligible, $\eta_p^2 = 0.028$. The average standard score for Spanish testing was higher, $M = 90.88$, SD = 19.78, compared to the average for English, $M = 87.28$, SD = 18.07.

There was a main effect for Language Group, $F(4773) = 37.37$, $p<0.0001$, with a small effect size, $\eta_p^2 = 0.162$. The total Spanish standard score was averaged with the total English standard score to create a combined Spanish-English standard score. As expected, functionally monolingual children scored lower than bilingual children. Overall standard scores were lower for children from the FME group, $M = 82.10$, SD = 7.80 and FMS group, $M = 86.19$, SD = 10.85 than were scores of children from the BDE group, $M = 93.64$, SD = 13.09; BL group $M = 93.08$, SD = 10.79; and BDS group $M = 93.27$, SD = 12.61.

We were also interested in how test language and language group mediated the relationship between receptive and expressive testing performance. We found a 3-way Modality × Test Language × Language Group interaction, $F(4980) = 27.61$, $p< 0.0001$, $\eta_p^2 = 0.125$. Because we were primarily concerned with the receptive-expressive gap in each language and not

across languages, we decomposed the three-way interaction by looking at the two-way interactions of interest.

### Interactions between modality, test language, and language groups

First, there was a statistically significant interaction between Modality and Test Language, $F(1777) = 62.10$, $p<0.0001$, $\eta_p^2 = 0.074$. As a follow up, we calculated the average receptive-expressive gap in Spanish (the expressive standard score in Spanish subtracted from the receptive standard score in Spanish), $M = 1.97$, SD = 15.63, and in English, $M = 9.08$, SD = 21.01, and compared them using paired samples *t*-tests. Results showed a statistically significant difference in the magnitude of the receptive-expressive gaps, $t(777) = 7.875$, $p<0.0001$.

There was also an interaction between Modality and Language Group, $F(4773) = 7.35$, $p<0.0001$, $\eta_p^2 = 0.037$. To explore this interaction, we calculated the average receptive-expressive gap in Spanish and English separately. We then performed paired-samples *t*-tests to compare the Spanish gap with the English gap for each language exposure group. Because multiple comparisons increase the risk that statistically significant differences between means will be due to chance, we used a Bonferroni correction, which reduced the critical *p*-value to 0.01 (0.05 divided by five comparisons = 0.01). Results showed that the gap was larger in English than in Spanish for all groups. This difference was statistically significant for all but the BDE group.

Tukey post-hoc analyses of one-way ANOVA comparing levels of language exposure found that for both receptive and expressive Spanish tasks, there was no statistically significant difference between means for the FMS, BDS, and BL groups who scored significantly higher than the BDE and FME groups (see Table 4). The BDE group scored significantly higher than the FME group. In English receptive and expressive semantics, there was no statistically significant difference between the performance of the FME and BDE groups, whose scores were higher than those of the BL, BDS, and FMS groups (see Table 5). The BL and BDS groups' scores were not statistically significantly different from one another but were higher than the scores for the FMS group.

To understand better how individual variation contributed to the pattern of gaps, we determined the number of children in each language experience group who demonstrated a receptive-expressive gap. Seventy-eight percent of the children in the FMS group had a gap in English, but only 21% of the children in the FME group had a gap in Spanish (see Table 6 for the number of children with a receptive-expressive gap by language experience group).

### Variables that predict the receptive-expressive gap

We used multiple regression analyses to pursue our second research question, which sought to determine which variables most contributed to the receptive-expressive gap. Variables of interest included age, current average exposure to English, age at which the child was regularly exposed to English, mother's education, and sex. These variables were selected because they have been shown to be influential in bilingual language development (Gibson et al. 2012; Hakuta and D'Andrea 1992). Although current average exposure to English was

used to create categorical variables in the repeated measures ANOVA, this variable was treated as interval data in the regression analysis in order to increase the statistical power of the model.

To determine if the model could predict a receptive-expressive gap in English or Spanish, we regressed the set of independent variables on the receptive-expressive gap in Spanish and again on the receptive-expressive gap in English (see Tables 7 and 8). Results revealed that the five independent variables included in the model explained only four percent of the variance in the Spanish receptive-expressive gap $F(5772) = 7.69$, $p<0.001$, $R^2= 0.041$, with no variable uniquely contributing to variance explained in a meaningful way. Age explained 1.2% of variance; average English experience explained 2.2% of the variance, and age of first regular exposure to English explained 0.6% of the variance. On the other hand, the same model explained 11.3% of the variance in the English receptive-expressive gap, $F(5772) = 20.73$, $p<0.001$, adjusted $R^2 = 0.113$. The unique contribution to the explanation of the variance was made by average English exposure, $R^2\Delta = 0.087$, $p<0.001$; Beta = 0.263, $p<0.001$, and age $R^2\Delta = 0.030$, $p<0.001$; Beta = 0.136, $p<0.001$.

## Discussion

The current investigation was guided by two research questions. First, what is the relationship between the receptive-expressive gap and language experience? Second, what variables best predict the presence of the receptive-expressive gap?

For the children in this study, the receptive-expressive gap was considerably larger in English, 9.07 standard points, than in Spanish, 1.97 standard points. In both languages, the magnitude of the gap was related to the amount of English input that parents reported for their children. The receptive-expressive gap in English was greatest for children who were exposed to very little English (FMS), 19.45 standard points, and least for children who were exposed to a good deal of English (BDE), 4.81 standard points. A similar pattern was observed in Spanish where the receptive-expressive gap was greatest for children who were exposed to very little Spanish (FME), 6.64 standard points, and least for children who were exposed to a good deal of Spanish (BDS and BL), 1.47 standard points and 0.89 standard points, respectively.

Consistent with previous research (Bedore et al. 2012; Gibson et al. 2012), multiple regression analyses identified English language experience as the strongest predictor of performance in both English (a positive relationship) and Spanish (a negative relationship) semantics testing, uniquely explaining 33% and 47% of the variance, respectively. But when the same model was used to predict the receptive-expressive gap itself, language exposure had much less influence in both English and Spanish, explaining 11.3% and 4.1% of the variance, respectively.

In the current study, it appears that language exposure plays a dominant role in the appearance of the receptive-expressive gap compared to other variables related to language development and that gap is much more likely to appear in the L2 than the L1. These results

are in conflict with previous studies in which the receptive-expressive gap was larger in the L1 than in the L2 (Table 1).

### Possible explanation: Weaker links hypothesis

Results of the current study might be related to a logical extension of the weaker links hypothesis in which phonological representations become stronger with practice. Metsala and Walley (1998) proposed the lexical restructuring hypothesis, which claimed that phonological representations are stored first as large chunks of information at the phrasal and lexical levels, and with experience those representations are decomposed into segmental representations. At the segmental level, representations are presumably more fine-grained and precise than at the lexical level, making them easier to access and produce. The shift from whole-word representations to segmental representations is said to be piecemeal and tied to practice, i.e. those sounds that appear most frequently are the first to be decomposed and stored segmentally.

A similar process of lexical restructuring in the L2 might be taking place among the bilingual participants in the current study. The receptive-expressive gap was relatively small in Spanish compared to English, even for children with limited exposure to Spanish. It may have been the case that many of their phonological representations had undergone the lexical restructuring process in Spanish. When asked to produce a word in Spanish, children were able to do so because those phonological representations were well specified and stored at the segmental level. In English, an equivalent amount of lexical restructuring may not have taken place. The average age at which children were regularly exposed to English was 2.12 years, and the quality of phonological representations in English may not have been as refined as those in Spanish. This made access and production of words in English more difficult than in Spanish.

While these weak phonological representations may have made production difficult, especially in English, the same phonological representations may have been sufficient to recognize words in receptive semantic tasks. Because receptive and expressive semantics tasks require different levels of precision from phonological representations (Bates 1993), the receptive-expressive gap might be attributable to the quality of phonological representations that these children had, especially in English. With extended exposure to English, the receptive-expressive gap in English presumably will wane. This phenomenon has been observed in the receptive-expressive gap in L1 of Spanish-English bilingual children in Miami from both high and low SES environments and from English immersion and bilingual classrooms (Oller and Eilers 2002). For those children, the magnitude of the receptive-expressive gap reduced by the fifth grade, indicating that the gap decreases with increased second language exposure. The role of underspecified phonological representations can be illustrated by the example above, in which similarities and differences in English were targeted both receptively and expressively. When children were asked to point to the piñata that was different from the others, underspecified phonological representations of the instructions may have been sufficient to decipher the message and successfully point to the piñata that was different. On the other hand, when children were shown a picture and asked ‘What is different about these two invitations?,’ they could not

have accurately produced the answer without precise phonological representations.<p> The finding in previous studies (Table 1) that the receptive-expressive gap was larger in L1 than in L2, the opposite pattern as found in the current study, might be related to the picture naming tasks used in those studies. Picture-naming tasks are decontextualized and, as noted by Oller et al. (2010), might activate a multitude of lexical semantic representations compared to the limited set of representations activated in a picture-pointing task. Thus, even for monolingual speakers, there will be competition for selection among lexical items during a picture-naming task. This phenomenon is exacerbated for bilinguals who have translation equivalents competing for selection (Gollan et al. 2005). Because picture naming allows for only a single correct answer, lexical selection in this task might be thwarted in the face of cross-linguistic competition. The expressive measure in the current study, however, allowed for multiple correct answers and may have increased the likelihood of success in the face of cross-linguistic competition, thus shrinking the size of the receptive-expressive gap in Spanish compared to previous studies.

## A sociolinguistic role

At least some part of the receptive-expressive gap seems to be due to the social roles of the two languages. Children from the FME group, whose Spanish language experience accounted for less than 20% of their total language experience, scored 69 standard points on the receptive semantics test in Spanish and 62 standard points on the expressive. On the other hand, children from the FMS group, whose English language exposure accounted for less than 20% of their total language input scored 84 standard points on the receptive semantics test in English and 63 standard points on the expressive. The two groups' performances on expressive semantics testing might be explained by the weak phonological representations account presented above. However, this account does not explain why children from the FMS group performed so much better on receptive English semantics testing than children from the FME group performed on receptive Spanish semantics testing.

The difference in receptive semantics scores between the FMS and FME groups' non-dominant languages might be attributed to a sociolinguistic cost–benefit analysis with respect to language learning. Children presumably are aware that they need Spanish for communication at home and English outside of the home. Based on this awareness, they appear to focus attention on English to which they are exposed, resulting in relatively high receptive semantics scores given their limited English exposure. This recognition might also be a probabilistic one. For example, De Houwer (2007) found that children were more likely to become bilingual if both parents used the minority language at home than children who had at least one parent who spoke the majority language at home. Although children from the FME group are exposed to as much of the non-dominant language as are children from the FMS group, their receptive semantics scores in Spanish suggest that they do not focus attention on the Spanish to which they are exposed. This presumably is due to their recognition that they need English for communication both at home and outside of the home.

While it may indeed be the case that multilingualism is the rule and not the exception internationally (Grosjean 2010), these data appear to support Grosjean's (2008) complementarity principle in which individuals only learn a language to the extent it is

needed for participation in a culture. We suspect that as phonological representations in English improve with practice in English, the magnitude of the receptive-expressive gap will shrink in English, and as Spanish phonological representations degrade with a decline in Spanish use, the magnitude of the gap will widen in Spanish.

## Summary and Conclusion

The results of the current study affirm a receptive-expressive gap in the L2 but not the L1, except for the Spanish results of children who were functionally monolingual in English. This might support an extension of the weaker links hypothesis, which argues that bilingual phonological representations are weaker than those of monolingual peers because of the divided language exposure that bilinguals experience. A shift in the importance of English over Spanish, due to living in an English dominant community, may have contributed to the gap in Spanish test scores for children who were functionally monolingual in English. Because there was no substantial receptive-expressive gap in Spanish for all but those children who were functionally monolingual in English, a suppression of Spanish was not supported but cannot be ruled out. The effects of L1 suppression may be unique to tasks like picture naming that were not used in the current study. In addition, the best predictor of the receptive-expressive gap was children's current experience in English as opposed to the age at which children were regularly exposed to English.

Educators should understand the importance of language experience with respect to the normal course of bilingual language development. The current study found a robust association between the receptive-expressive gap and children's level of language experience. These findings suggest that variability in children's receptive-expressive scores, especially during early exposure to a second language, is to be expected.

## Acknowledgments

This work was supported by the grant R01DC007439 from the National Institute on Deafness and Other Communication Disorders (NIDCD). We thank the anonymous reviewers for their comments on this manuscript. We are grateful to the families that participated in the study. We would also like to thank Anita Méndez Pérez and Chad Bingham for their assistance with coordination of data collection, the interviewers for their assistance with collecting the data for this project, and the school districts for allowing us access to collect the data. This report does not necessarily reflect the views or policy of NIDCD.

## References

Allen, MJ., Yen, WM. Introduction to measurement theory. Monterey, CA: Brooks; 1979.

Barnett WS, Yarosz DJ, Thomas J, Jung K, Blanco D. Two-way and monolingual English immersion in preschool education: An experimental comparison. Early Childhood Research Quarterly. 2007; 22:277–93.

Bates E. Comprehension and production in early language development: Comments on Savage-Rumbaugh et al. Monographs of the Society for Research in Child Development. 1993; 58(3–4): 222–42. [PubMed: 8366873]

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., Volterra, V. The emergence of symbols: Cognition and communication in infancy. New York: Academic Press; 1979.

Bates, E., Dale, PS., Thal, D. Individual differences and their implications for theories of language development. In: Fletcher, P., MacWhinney, B., editors. Handbook of child language. Oxford: Basil Blackwell; 1995. p. 96-151.

Bates E, Thal D, Whitesell K, Fenson L, Oakes L. Integrating language and gesture in infancy. Developmental Psychology. 1989; 25:1004–19.

Bedore, LM., Cooperson, SJ., Boerger, KM. Morphosyntactic development. In: Goldstein, B., editor. Bilingual language development and disorders in Spanish-English speakers. Baltimore: Paul H. Brookes Publishing Company; 2011. p. 175-92.

Bedore LM, Peña ED, Summers CL, Boerger KM, Resendiz MD, Greene K, Bohman TM, Gillam RB. The measure matters: Language dominance profiles across measures in Spanish-English bilingual children. Bilingualism: Language and Cognition. 2012; 15(3):616–29.

Benedict H. Early lexical development: Comprehension and production. Journal of Child Language. 1979; 6:183–200. [PubMed: 468932]

Bohman TM, Bedore LM, Peña ED, Mendez-Perez A, Gillam RB. What you hear and what you say: Language performance in Spanish-English bilinguals. International Journal of Bilingual Education and Bilingualism. 2010; 13:325–44. [PubMed: 21731899]

Bowey JA. On the association between phonological memory and receptive vocabulary in five-year-olds. Journal of Experimental Child Psychology. 1996; 63:44–78. [PubMed: 8812028]

Bowey JA. What does nonword repetition measure? A reply to Gathercole and Baddeley. Journal of Experimental Child Psychology. 1997; 67:295–301. [PubMed: 9398396]

Brownell, R. Expressive one-word picture vocabulary test. 3. Novato, CA: Academic Therapy Publications; 2000a.

Brownell, R. Receptive one word picture vocabulary test. 3. Novato, CA: Academic Therapy Publications; 2000b.

Brownell, R. Expressive one-word picture vocabulary test: Spanish–English edition. Novato, CA: Academic Therapy Publications; 2001.

Carrow-Woolfolk, E. Oral and written language scales. Circle Pines, MN: American Guidance Service; 1995.

Cheung H. Nonword span as a unique predictor of second-language vocabulary language. Developmental Psychology. 1996; 32:867–73.

Chiappe P, Chiappe DL, Gottardo A. Vocabulary, context, and speech perception among good and poor readers. Educational Psychology. 2004; 24:825–43.

Chiat S, Roy P. Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. Journal of Child Psychology and Psychiatry. 2008; 49(6):635–45. [PubMed: 18341547]

Clark EV, Hecht BF. Comprehension, production, and language acquisition. Annual Review of Psychology. 1983; 34:325–49.

Cornips L, Hulk A. Factors of success and failure in the acquisition of grammatical gender in Dutch. Second Language Research. 2008; 24(3):267–95.

De Houwer A. Parental language input patterns and children's bilingual use. Applied Psycholinguistics. 2007; 28:411–24.

Dollaghan CA, Campbell TF, Paradise JL, Feldman HM, Janosky JE, Pitcairn DN, Kurs-Lasky M. Maternal education and measures of early speech and language. Journal of Speech, Language, and Hearing Research. 1999; 42:1432–43.

Dunn, LM., Padilla, ER., Lugo, DE., Dunn, LM. Test de vocabulario en imagenes Peabody (TVIP) [Vocabulary test in Peabody images]. Circle Pines, MN: AGS Publishing; 1986.

Dunn, L., Dunn, L. Peabody picture vocabulary test – third edition (PPVT-III). Circle Pines, MN: AGS Publishing; 1997.

Dunn, LM., Dunn, DM. Peabody picture vocabulary test – fourth edition (PPVT-IV). Bloomington, MN: NCS Pearson; 2007.

Eriksson M, Marschik PB, Tulviste T, Almgren M, Pérez Pereira M, Wehberg S, Gallego C. Differences between girls and boys in emerging language skills: Evidence from 10 language communities. British Journal of Developmental Psychology. 2011; 30(2):326–43. [PubMed: 22550951]

Fenson L, Dale PS, Reznick JS, Bates E. Variability in early communicative development. Monographs of the Society for Research in Child Development. 1994; 59(5):1–185.

Galsworthy MJ, Dionne G, Dale PS, Plomin R. Sex differences in early verbal and non-verbal cognitive development. Developmental Science. 2000; 3(2):206–15.

Gamez P, Levine S. Oral language skills of Spanish-speaking English language learners: The impact of high-quality native language exposure. Applied Psycholinguistics. Forthcoming.

Gathercole S. Nonword repetition and word learning: The nature of the relationship. Applied Psycholinguistics. 2006; 27(4):513–43.

Gathercole VC. Some myths you may have heard about first language acquisition. TESOL. 1988; 22:407–35.

Gibson TA, Oller DK, Jarmulowicz L, Ethington CA. The receptive-expressive gap in the vocabulary of young second-language learners: Robustness and possible mechanisms. Bilingualism: Language and Cognition. 2012; 15:102–16.

Goldin-Meadow S, Seligman MEP, Gelman R. Language in the two-year old. Cognition. 1976; 4:189–202.

Gollan TH, Acenas LAR. What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and Tagalong-English bilinguals. Journal of Experimental Psychology: Learning, Memory, and Cogniton. 2004; 30(1):246.

Gollan TH, Fennema-Notestine C, Montoya RI, Jernigan TL. The bilingual effect on Boston Naming Test performance. Journal of International Neuropsychological Society. 2007; 13(2):197–208.

Gollan TH, Montoya RI, Cera C, Sandoval TC. More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. Journal of Memory and Language. 2008; 58:787–814. [PubMed: 19343088]

Gollan TH, Montoya RI, Fennema-Notestine C, Morris SK. Bilingualism affects picture naming but not picture classification. Memory and Cognition. 2005; 33(7):1220–34. [PubMed: 16532855]

Grosjean, F. Studying bilinguals. New York: Oxford University Press USA; 2008.

Grosjean, F. Bilingual: Life and reality. Cambridge, MA: Harvard University Press; 2010.

Gutiérrez–Clellen VF, Kreiter J. Understanding child bilingual acquisition using parent and teacher reports. Applied Psycholinguistics. 2003; 24:267–88.

Hakuta K, D'Andrea D. Some properties of bilingual maintenance and loss in Mexican background high-school-students. Applied Linguistics. 1992; 13(1):72–99.

Hammer CS, Lawrence FR, Miccio AW. Exposure to English before and after entry into Head Start 1: Bilingual children's receptive language growth in Spanish and English. International Journal of Bilingual Education and Bilingualism. 2008; 11(1):30–56. [PubMed: 26430382]

Hart, B., Risley, T. Meaningful differences in the everyday experience of young American children. Baltimore, MD: Paul H Brookes Publishing; 1995.

Hoff E, Tian C. Socioeconomic status and cultural influences on language. Journal of Communication Disorders. 2005; 38:271–8. [PubMed: 15862810]

Hoff-Ginsberg E. The relation of birth order and socioeconomic status to children's language experience and language development. Applied Psycholinguistics. 1998; 19:603–30.

Hollingshead, AB. Four factor index of social status. New Haven, CT: Department of Sociology, Yale University; 1975.

Huttenlocher J, Haight W, Bryk A, Seltzer M, Lyons T. Early vocabulary growth: Relation to language input and gender. Developmental Psychology. 1991; 27(2):236.

Ivanova I, Costa A. Does bilingualism hamper lexical access in speech production? Acta Psychologica. 2008; 127(2):277–88. [PubMed: 17662226]

Kan PF, Kohnert K. Preschoolers learning Hmong and English: Lexical-semantic skills in L1 and L2. Journal of Speech, Language, and Hearing Research. 2005; 48:372–83.

Kohnert KJ, Bates E, Hernandez AE. Balancing bilinguals: Lexical-semantic production and cognitive processing in children learning Spanish and English. Journal of Speech, Language, and Hearing Research. 1999; 42:1400–13.

Leonard LB. Is expressive language disorder an accurate diagnostic category? American Journal of Speech-Language Pathology. 2009; 18:115–23. [PubMed: 19029534]

Lesaux NK, Crosson AC, Kieffer MJ, Pierce M. Uneven profiles: Language minority learners' word reading, vocabulary, and reading comprehension skills. Journal of Applied Developmental Psychology. 2010; 31:475–83. [PubMed: 21243117]

Linck JA, Kroll JF, Sunderman G. Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second language learning. Psychological Science. 2009; 20:1507–15. [PubMed: 19906121]

Magnuson KA, Sexton HR, Davis-Kean PE, Huston AC. Increases in maternal education and young children's language skills. Merrill-Palmer Quarterly. 2009; 55(3):319–50.

Mann W, Marshall C. Investigating deaf children's vocabulary knowledge in British sign language. Language Learning. Forthcoming.

Mashburn AJ, Justice LM, Downer JT, Pianta RC. Peer effects on children's language achievement during pre-kindergarten. Child Development. 2009; 80(3):686–702. [PubMed: 19489897]

Masoura E, Gathercole S. Phonological short-term memory skills and new word learning in young Greek children. Memory. 2005; 13:422–9. [PubMed: 15948628]

Metsala, J., Walley, A. Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In: Metsala, J., Ehri, L., editors. Word recognition in beginning literacy. Hillsdale, NJ: Erlbaum; 1998. p. 89-120.

Metsala JL. Young children's phonological awareness and nonword repetition as a function of vocabulary development. Journal of Educational Psychology. 1999; 91:3–19.

Miccio, AW., Tabors, PO., Páez, MM., Hammer, CS., Wagstaff, DA. Vocabulary development of Spanish-speaking head start children of Puerto Rican descent. Paper presented at the 4th International Symposium on Bilingualism; April 30–May 3; Tempe, AZ. 2003.

Muñoz M, Marquardt T. Picture naming and identification in bilingual speakers of Spanish and English with and without aphasia. Aphasiology. 2003; 17:1115–32.

Nelson K, Rescorla L, Gruendel J, Benedict H. Early lexicons: What do they mean? Child Development. 1978; 49:960–8.

Oller, DK., Eilers, RE. Language and literacy in bilingual children. Clevedon: Multilingual Matters; 2002.

Oller, DK., Jarmulowicz, L., Pearson, BZ., Cobo-Lewis, AB. Rapid spoken language shift in early second-language learning. In: Durgunouglu, A., Goldenberg, C., editors. Dual language learners: Their development and assessment in oral and written language. Guilford, NY: The Guilford Press; 2010. p. 94-120.

Oller DK, Pearson BZ, Cobo-Lewis AB. Profile effects in early bilingual language and literacy. Applied Psycholinguistics. 2007; 28:191–230. [PubMed: 22639477]

Peña E, Bedore L, Gutiérrez-Clellen V, Iglesias A, Goldstein B. Bilingual English Spanish oral screener (BESOS). In preparation.

Peña E, Bedore LM, Rappazzo C. Comparison of Spanish, English, and bilingual children's performance across semantic tasks. Language, Speech, and Hearing Services in Schools. 2003; 34:5–16.

Peña E, Gutiérrez-Clellen V, Iglesias A, Goldstein B, Bedore L. Bilingual English Spanish assessment (BESA). In preparation.

Peña ED, Gillam RB, Bedore LM, Bohman TM. Risk for poor performance on a language screening measure for bilingual preschoolers and kindergarteners. American Journal of Speech-Language Pathology. 2011; 20:302–14. [PubMed: 21821821]

Pierce CA, Block RA, Aguinis H. Cautionary note on reporting eta-squared values from multifactor ANOVA designs. Educational and Psychological Measurement. 2004; 64:916–24.

Semel, E., Wiig, EH., Secord, WA. Clinical evaluation of language fundamentals – fourth edition (CELF-4). San Antonio, TX: Psychological Corporation; 2003.

Stevens, J. Applied multivariate statistics for the social sciences. Hillsdale, NJ: Lawrence Erlbaum; 2002.

Summers C, Bohman TM, Gillam RB, Peña ED, Bedore LM. Bilingual performance on nonword repetition in Spanish and English. International Journal of Language and Communication Disorders. 2010; 45:480–93. [PubMed: 20565327]

Thorn A, Gathercole S. Language-specific knowledge and short-term memory in bilingual and non-bilingual children. The Quarterly Journal of Experimental Psychology Section A. 1999; 52(2): 303–24.

Thorn A, Gathercole S, Frankish C. Language familiarity effects in short-term memory: The role of output delay and long-term knowledge. The Quarterly Journal of Experimental Psychology Section A. 2002; 55(4):1363–83.

Uchikoshi Y, Maniates H. How does Bilingual instruction enhance English achievement? A mixed-methods study of Cantonese-speaking and Spanish-speaking bilingual classrooms. Bilingual Research Journal: The Journal of the National Association for Bilingual Education. 2010; 33(3): 364–85.

Unsworth S. Age and input in the acquisition of grammatical gender in Dutch. Second Language Research. 2008; 24(3):365–95.

Watt N, Wetherby A, Shumway S. Prelinguistic predictors of language outcome at 3 years of age. Journal of Speech, Language, and Hearing Research. 2006; 49:1224–37.

Williams, KT. Expressive vocabulary test second edition (EVT-2). Circle Pines, MN: Pearson Assessments; 2007.

Windsor J, Kohnert K. The search for common ground. Part I: Lexical performance by linguistically diverse learners. Journal of Speech Language and Hearing Research. 2004; 47:877–90.

Woodcock, R. Woodcock language proficiency battery – revised (WLPB-R). Itasca, IL: Riverside Publishing; 1991.

Woodcock, R., Johnson, M. Woodcock-Johnson psycho-educational battery-revised (WJPEB-R). Allen, TX: DLM Teaching Resources; 1989.

Woodcock, R., Muñoz-Sandoval, A. Woodcock language proficiency battery – revised: Spanish form (WLPB–RS). Itasca, IL: Riverside Publishing; 1995.

Woodcock, RW., Sandoval, AFM. Batería Woodcock-Muñoz: Pruebas de habilidad cognitiva-revisada [Woodcock-Muñoz Battery: Tests of cognitive ability – revised]. Itasca, IL: Riverside Publishing; 1996.

Yan S, Nicoladis E. Finding le mot juste: Differences between bilingual and monolingual children's lexical access in comprehension and production. Bilingualism: Language and Cognition. 2009; 12:323–35.

Zimmerman, IL., Steiner, VG., Pond, RE. Preschool language scales – fifth edition (PLS-5). San Antonio, TX: The Psychological Corporation; 2011.

**Table 1**

Literature review of mean (standard deviation) for standard test scores in English and Spanish.

| | | Spanish | | | | English | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Study** | **Population** | **Num** | **Rec** | **Exp** | **Gap** | **Num** | **Rec** | **Exp** | **Gap** |
| Barnett et al. (2007) | Preschool English immersion | 27 | 83 (12) | 73 (12) | 10 | 52 | 78 (17) | 79 (18) | −1 |
| Barnett et al. (2007; Continued) | Preschool (two-way class) | 47 | 84 (13) | 75 (15) | 9 | 79 | 78 (16) | 80 (18) | 2 |
| Gibson et al. (2012) | Kindergarten | 124 | 88 (17) | 67 (21) | 21 | 124 | 61 (18) | 54 (21) | 7 |
| Lesaux et al. (2010) | Grade 4 | 87 | 94 (16) | 86 (22) | 8 | 87 | 82 (14) | 78 (16) | 4 |
| Lesaux et al. (2010; Continued) | Grade 5 | 87 | 92 (16) | 89 (20) | 3 | 87 | 85 (15) | 82 (15) | 3 |
| Miccio et al. (2003) | Preschool | 38 (rec), 41 (exp) | 75 (15) | 56 (14) | 19 | 38 (rec), 41 (exp) | 77 (15) | 72 (20) | 5 |
| Oller and Eilers (2002) | Kindergarten | 332 | 90 | 62 | 28 | 332 | 71 | 70 | 1 |
| Oller and Eilers (2002; Continued) | Grade 2 | 306 | 91 | 63 | 28 | 306 | 89 | 80 | 0 |
| Oller and Eilers (2002; Continued) | Grade 5 | 314 | 90 | 71 | 19 | 314 | 91 | 87 | 4 |
| Uchikoshi and Maniates (2010) | Grade 2 | 34 | 92 (22) | 78 (30) | 14 | 34 | 89 (20) | 81 (25) | 8 |
| Windsor and Kohnert (2004) | Average age 9 years, 9 months | 22 | 101 (14) | 88 (20) | 13 | 22 | 101 (16) | 96 (16) | 5 |

Note: All receptive vocabulary tests in Spanish used the TVIP (Dunn et al. 1986). Barnett et al. used the picture vocabulary subtest of the *Batería Woodcock-Muñoz: Pruebas de habilidad cognitive-Revisada* (Woodcock and Sandoval 1996) for expressive vocabulary in Spanish. Gibson et al. (2012), Lesaux et al. (2010), Miccio et al. (2003), Oller and Eilers (2002), and Uchikoshi and Maniates (2010) used the picture vocabulary subtest of the *Woodcock Language Proficiency Battery = Revised: Spanish Form* (Woodcock and Muñoz-Sandoval 1995) for Spanish expressive vocabulary. Windsor and Kohnert used the *Expressive One-Word Picture Vocabulary Test: Spanish-English edition* (Brownell 2001) for Spanish expressive vocabulary. All of the reported studies used the *Peabody Picture Vocabulary Test = Third Edition* (Dunn and Dunn 1997) to measure receptive vocabulary in English. Barnett et al. used the picture vocabulary subtest of the *Woodcock-Johnson Psycho-Educational Battery = Revised* (Woodcock and Johnson 1989) to measure expressive vocabulary in English. Windsor and Kohnert used the *Expressive One-Word Picture Vocabulary Test* (Brownell 2000a) to measure English expressive vocabulary. The remaining studies used the *Woodcock Language Proficiency Battery = Revised* (Woodcock 1991) to assess English expressive vocabulary. No standard deviations were reported for Oller and Eilers (2002).

Num, number; Rec, receptive vocabulary; Exp, expressive vocabulary; Gap, the difference resulting from expressive standard score subtracted from receptive standard score.

**Table 2**

Mean demographic information (standard deviation) by language experience group.

| Language group | Age in months | % English experience | Age of first English experience | Mother's education | % Female |
|---|---|---|---|---|---|
| FMS (*n* = 180) | 64.39 (3.14) | 9 (7) | 4.12 (1.07) | 2.54 (1.46) | 47 |
| BDS (*n* = 120) | 65.30 (3.74) | 31 (6) | 3.10 (1.76) | 2.71 (1.58) | 43 |
| BL (*n* = 211) | 65.65 (3.58) | 49 (5) | 2.05 (1.99) | 2.75 (1.65) | 46 |
| BDE (*n* = 90) | 66.67 (4.33) | 69 (6) | 0.98 (1.60) | 3.13 (1.55) | 51 |
| FME (*n* = 177) | 65.88 (4.13) | 96 (6) | 0.11 (0.64) | 4.36 (1.11) | 58 |

Note: Mother's education based on Hollingshead's (1975) categorization of education; a score of three represents completion of high school. Percentage of English experience reflects the weekly average of both input and output in English.

FMS, functionally monolingual in Spanish; BDS, bilingual but dominant in Spanish; BL, bilingual; BDE, bilingual but dominant in English; FME, functionally monolingual in English.

**Table 3**

Mean (standard deviation) of receptive and expressive Spanish and English semantics raw scores by language experience group and by age group.

| | | Language of testing | | | |
|---|---|---|---|---|---|
| | | Spanish | | English | |
| Language group | Age group | Rec | Exp | Rec | Exp |
| FMS | Older (*n* = 65) | 2.25 (.87) | 5.17 (1.96) | 0.95 (1.04) | 3.28 (2.12) |
| | Younger (*n* = 115) | 2.23 (0.87) | 5.50 (1.83) | 1.18 (1.25) | 1.46 (1.70) |
| BDS | Older (*n* = 53) | 2.64 (0.62) | 6.11 (1.96) | 1.92 (1.37) | 3.28 (2.12) |
| | Younger (*n* = 67) | 1.24 (.82) | 5.16 (2.28) | 1.40 (1.21) | 2.54 (2.05) |
| BL | Older (*n* = 102) | 2.21 (0.82) | 5.36 (2.10) | 1.90 (1.09) | 3.56 (1.81) |
| | Younger (*n* = 109) | 1.45 (1.11) | 3.50 (2.44) | 2.33 (.95) | 4.10 (1.49) |
| BDE | Older (*n* = 48) | 1.92 (1.15) | 4.85 (2.84) | 2.35 (1.30) | 4.33 (1.88) |
| | Younger (*n* = 42) | 1.45 (1.11) | 3.50 (2.44) | 2.33 (.95) | 4.10 (1.49) |
| FME | Older (*n* = 86) | 0.41 (0.74) | 0.44 (1.19) | 2.50 (0.92) | 4.756 (1.28) |
| | Younger (*n* = 91) | 0.30 (0.57) | 0.40 (1.08) | 2.33 (1.25) | 4.76 (1.35) |

FMS, functionally monolingual in Spanish; BDS, bilingual but dominant in Spanish; BL, bilingual; BDE, bilingual but dominant in English; FME, functionally monolingual in English; Rec, receptive semantic testing; Exp, expressive semantic testing. Younger age group = ages 60–65 months. Older age group = 66 months and older.

## Table 4

Mean (standard deviation) of receptive and expressive Spanish semantics standard scores by language exposure group.

| Language group | Rec | Exp | Gap | Effect size |
|---|---|---|---|---|
| FMS ($n$ = 180) | 100.00[a] (15.00) | 100.00[d] (15.00) | 0.00 | 0 |
| BDS ($n$ = 120) | 103.15[a] (13.14) | 101.68[d] (17.81) | 1.47 | 0.09 |
| BL ($n$ = 211) | 99.31[a] (14.22) | 98.42[d] (17.24) | 0.89 | 0.06 |
| BDE ($n$ = 90) | 90.79[b] (19.76) | 98.42[d] (17.24) | −0.28 | −0.01 |
| FME ($n$ = 177) | 67.52[c] (11.35) | 60.88[f] (9.43) | 6.64* | 0.64 |
| Total ($N$ = 778) | 91.86 (19.79) | 89.89 (22.64) | 1.97* | 0.09 |

Note: Means that share superscripts do not differ from each other with statistical significance (Tukey HSD, $p<0.05$).

FMS, functional monolingual Spanish; BDS, bilingual Spanish dominant; BL, bilingual; BDE, bilingual dominant in English; FME, functional monolingual English; Rec, receptive semantic testing; Exp, expressive semantic testing. Gap, the difference between receptive minus expressive means, which were compared by paired $t$-tests.

*$p<0.001$.

## Table 5

Mean (standard deviation) receptive and expressive English semantics standard scores by language experience group.

| Language group | Rec | Exp | Gap | Effect size |
|---|---|---|---|---|
| FMS ($n$ = 180) | 82.06[a] (16.68) | 62.61[d] (18.76) | 19.45 ** | 1.10 |
| BDS ($n$ = 120) | 89.62[b] (18.36) | 78.65[e] (23.85) | 10.97 ** | 0.52 |
| BL ($n$ = 211) | 91.58[b] (16.13) | 83.01[e] (20.83) | 8.57 ** | 0.46 |
| BDE ($n$ = 90) | 98.75[c] (17.31) | 93.94[f] (19.59) | 4.81 * | 0.07 |
| FME ($n$ = 177) | 100.00[c] (15.00) | 100.00[f] (15.00) | 0.00 | 0.00 |
| Total ($N$ = 778) | 91.82 (17.74) | 82.75 (23.64) | 9.07 ** | 0.43 |

Note: In the Rec and Exp columns, means that share superscripts do not differ from each other with statistical significance (Tukey HSD, $p<0.05$).

FMS, functional monolingual Spanish; BDS, bilingual Spanish dominant; BL, bilingual; BDE, bilingual dominant in English; FME, functional monolingual English; Rec, receptive semantic testing; Exp, expressive semantic testing. Gap, the difference between receptive and expressive means, which was compared by paired $t$-tests.

* $p<0.05$;

** $p<0.001$.

**Table 6**

Number of participants with a receptive-expressive in each language experience group.

| Language experience | Number of children with gap in Spanish | Number of children with gap in English |
|---|---|---|
| FMS ($n$ = 180) | 38 | 140 |
| BDS ($n$ = 120) | 32 | 71 |
| BL ($n$ = 211) | 63 | 102 |
| BDE ($n$ = 90) | 19 | 37 |
| FME ($n$ = 177) | 37 | 41 |

Note: A gap occurs when receptive standard scores exceed expressive standard scores by 10 points or more.

## Table 7

Summary of regression analyses for variables predicting gap in Spanish.

| Variable | $R^2\Delta$ | Beta |
|---|---|---|
| Age (in months) | 0.012 * | −0.113 ** |
| Average% English exposure | 0.022 ** | 0.043 |
| Age of regular English Exposure (in years) | 0.006 * | −0.109 * |
| Mother's education (Hollingshead scale) | 0.002 * | 0.050 |
| Sex | 0.005 | −0.069 |

Note: For sex variable, female was coded as 0. $F(5,777) = 7.685$, $p<0.001$; Adjusted $R^2 = 0.041$

* $p<0.05$;

** $p<0.001$.

## Table 8

Summary of regression analyses for variables predicting gap in English.

| Variable | $R^2\Delta$ | Beta |
|---|---|---|
| Age (in months) | 0.030 * | −0.136 * |
| Average% English exposure | 0.087 * | −0.263 * |
| Age of regular English | | |
| Exposure (in years) | 0.000 | 0.028 |
| Mother's education (Hollingshead scale) | 0.001 | −0.028 |
| Sex | 0.001 | −0.032 |

Note: For sex variable, female was coded as 0. $F(5,777) = 20.726$, $p<0.001$; Adjusted $R^2 = 0.113$.

* $p<0.001$.