Title
A longitudinal investigation of the semantic receptive-expressive gap in Spanish-English bilingual children

Journal
International Journal of Bilingual Education and Bilingualism, 25(3)

ISSN
1367-0050

Authors
Gibson, Todd A
Peña, Elizabeth D
Bedore, Lisa M
et al.

Publication Date
2022-03-16

DOI
10.1080/13670050.2020.1721427

Peer reviewed

Published in final edited form as:
Int J Biling Educ Biling. 2022 ; 25(3): 819–833. doi:10.1080/13670050.2020.1721427.

A Longitudinal Investigation of the Semantic Receptive-Expressive Gap in Spanish-English Bilingual Children

Todd A. Gibson,
Louisiana State University

Elizabeth D. Peña,
University of California, Irvine

Lisa M. Bedore,
Temple University

Kevin S. McCarter
Louisiana State University

Abstract

Purpose: Although a *semantic receptive-expressive gap* appears to be a universal feature of early bilingualism, little is known about its development. We sought to determine if the magnitude of the discrepancy between receptive and expressive standard scores changed over time in bilingual children's two languages.

Method: In this longitudinal study, standardized receptive and expressive semantics tests of 106 Spanish-English bilingual children with TD were taken at kindergarten and first grade in both English and Spanish. We used a multivariate analysis approach to identify interactions and main effects.

Results: Although both receptive and expressive standard scores improved across the year in both languages, the magnitude of the gap was similar for both languages at both time points. However, there was greater improvement in English than in Spanish. Expressive scores at the end of the year were similar to receptive scores a year earlier.

Conclusions: The magnitude of this gap remains relatively constant at kindergarten and first grade in both English and Spanish, despite overall improvements in semantic performance in both languages. There is on average roughly a one year lag between receptive and expressive semantics skills. Clinicians should take caution in interpreting receptive-expressive semantic gaps.

Keywords

childhood bilingualism; second language learning; school-age children; semantic development

Unlike their monolingual peers, bilingual children often have better receptive than expressive performance on measures of semantic knowledge, even after controlling for

Address for correspondence: Todd A. Gibson, Louisiana State University, 63 Hatcher Hall, Baton Rouge, LA 70806, toddandrewgibson@gmail.com.

the differences in difficulty between the two modalities (Gibson, Oller, Jarmulowicz, & Ethington, 2012). We call this discrepancy the *semantic receptive-expressive gap*. This phenomenon is important to understand both theoretically and clinically. It is important theoretically because its underlying mechanisms are not well understood and therefore it is difficult to predict under which conditions the phenomenon will appear. It is important clinically because the existence of this phenomenon creates the conditions for the misdiagnosis of language impairment in the bilingual population. For example, when monolingual children's receptive standard scores significantly exceed expressive standard scores, children are frequently interpreted as having an expressive language disorder (Leonard, 2009). Therefore, this pattern of performance among bilingual children likely lends itself to a similar interpretation by speech-language pathologists (SLPs), even when it is not warranted. Furthermore, as explained below, different theoretical models result in the use of different clinical techniques with this population, especially during language assessment. The current study sought to contribute to our understanding of bilingual language acquisition through the lens of the semantic receptive-expressive gap by measuring its change over time and by identifying its predictors.

Receptive and expressive language processes differ but can be compared

Levelt's model of speech comprehension and production (Indefrey & Levelt, 2004; Levelt, Roelofs, & Meyer, 1999) can be used to illustrate how receptive and expressive language processes differ. For example, the processes involved in understanding speech begin with hearing a spoken word. Its phonetic features are identified, which activates and selects phonological representations, which activate and select the appropriate lemmas (abstract representations containing semantic and syntactic information but absent a phonological representation). Finally, lemmas activate concepts. In the other direction, producing speech begins with a concept that activates and selects its associated lemma. This is followed by activation and selection of the phonological readout of the segments that will be produced, with syllabification of the segments immediately after. Finally, articulatory programs associated with the target word are then activated and fed forward to the articulators to be executed.

Because of the inherent processing differences between receptive and expressive language, receptive language tasks, like pointing to pictures, are often considered easier than expressive language tasks, like naming pictures. However, after test standardization, these inherent differences in difficulty should diminish such that an individual's receptive and expressive standard scores should be similar. For example, in the Receptive One-Word Picture Vocabulary Test, Spanish-Bilingual Edition (ROW; Martin & Brownell, 2011), the average raw score for a child age 5;0 is 49. In the expressive counterpart to the ROW, the Expressive One-Word Picture Vocabulary Test, Spanish-Bilingual Edition (EOW; Martin & Brownell, 2011) the average raw score for a child age 5;0 is 39. A comparison of the raw scores would give the impression that receptive exceeds expressive performance. However, after converting the raw scores to standard scores, this discrepancy disappears, and the average for both tests is 100 standard score points.

Changes in the receptive-expressive gap over time

Although the semantic receptive-expressive gap has been identified in both languages of bilingual children (Gibson, et al., 2012; Gibson, Peña & Bedore, 2014a; see Gibson, Jarmulowicz & Oller, 2018, for role of task type), the trajectory of its change over time is not clear. Gibson, et al. (2012), proposed that the onset of the gap in L1 might occur immediately after L2 immersion since it was present for Spanish-English bilingual children within the first months of English-immersion kindergarten. However, the study was not longitudinal, and the gap may have been present before children began kindergarten. Using a cross-sectional design, Oller & Eilers (2002) observed that the gap in Spanish (L1) among Spanish-English bilinguals in English-immersion schools was 34 standard score points in kindergarten but 20 in the fifth grade (tests had a mean of 100 and standard deviation of 15). There was no significant gap in English. Similarly, in a longitudinal study of Spanish-English bilingual children at fourth and fifth grades, Lesaux, et al., (2010) reported a diminution of the gap in Spanish (L1) from 8 to 3 standard score points. Because this gap was not the focus of the study, this change was not tested statistically so it is not clear if it was meaningful.

However, the distribution of the gap might be related to the types of tasks used. The studies reported above used only comparisons of single-word picture naming and picture pointing tasks. When broader measures of semantic knowledge have been analyzed (Gibson et al., 2014a; Gibson, Peña & Bedore, 2014b), a semantic receptive-expressive gap in L2 English (but not L1 Spanish) has been observed and shown to be related to daily language experience. For example, Gibson et al. (2014a) administered the Bilingual English Spanish Oral Screener (BESOS; Peña, Bedore, Gutiérrez-Clellen, Iglesias, & Goldstein, 2014), bilingual Spanish-English screening instrument, to over 700 Spanish-English bilingual 5-year-olds. The screener tapped into a broad range of semantic knowledge by asking children to identify or produce categories, similarities and differences, characteristic properties, etc. (for review of task types, see Peña, Bedore & Rappazzo, 2003). Results revealed a meaningful receptive-expressive gap in English that had an inverse relationship with amount of English experience (i.e., children with the least English experience had the largest semantic receptive-expressive gap, and children with the most English experience had the smallest). There was no gap in Spanish. Similar results were found for older Spanish-English bilingual children (age 7 years) using similar tasks but with many more items (Gibson et al., 2014b). Together, these studies suggest that increased practice with a language leads to a diminution of the receptive-expressive gap. Therefore, we would anticipate a longitudinal decline in the magnitude of the gap.

The frequency-lag hypothesis

Because a semantic receptive-expressive gap is not a common feature in monolingual children's language performance, the development of this phenomenon presumably is a function of qualities related to the bilingual circumstance. Unlike monolingual speakers, the bilingual speaker's language experience is divided across two languages, a situation which has been termed the *distributed characteristic* (Oller et al., 2007). Because of the distributed characteristic, bilingual children have less practice in each of their languages compared to monolingual peers in their single language. Gollan and colleagues (Gollan, Fennema-Sotestine, Montoya, & Jernigan, 2007; Gollan, Montoya, Fennema-Notestine, & Morris, 2005) have argued that the monolingual/bilingual discrepancy in language frequency of use results in a word frequency effect (i.e., word tokens in each of the bilingual's languages occur less frequently than word tokens in the monolingual's language). Words that occur more frequently are easier to produce, and this results in a bilingual disadvantage in language production. The psycholinguistic operations of the bilingual word frequency effect have been specified by Gollan and colleagues in the *frequency-lag hypothesis* (also known as the *weaker links hypothesis*).

As reviewed by Gollan, Montoya, Cera, and Sandoval (2008), the frequency-lag hypothesis asserts that limited experience with words (as in the case of bilinguals compared to monolinguals) results in weak associations between semantic and phonological representations for those words. This impedes lexical access and might derail a bilingual speaker's ability to name known words (Gollan & Acenas, 2004). However, as experience increases, these links are enhanced, improving lexical access. The frequency-lag hypothesis has been used to explain bilingual disadvantages in word recall (Gollan & Acenas, 2004), naming accuracy (Kohnert, Hernandez, & Bates, 1998), and naming speeds (Ivanova & Costa, 2008).

An extension of the frequency-lag hypothesis has also been used to explain the semantic receptive-expressive gap. In both Gibson et al. (2014a) and Gibson et al. (2014b), mentioned above, we extended the frequency-lag hypothesis and proposed that it not only applied to the links between semantic and phonological representations but also to the representations themselves which become more highly specified with increased language use. Bates (1993) posited that successful performance on receptive language tasks could be achieved when phonological representations were imprecise; however, success on expressive language tasks required phonological representations that were highly specified. An example of this behavior is observed in 5-year old children learning novel words (Dollaghan, 1987) In this study, 5-year-old monolingual children were shown novel objects and asked to provide names for the objects. Later, when tested on their retention, children were decidedly successful when asked to identify those objects (receptive language task) but were mostly unsuccessful when asked to name the novel objects (expressive language task; note, however, that this task did not control for the inherent differences in the differences between the two modalities). Similar patterns have been found for bilingual children (Kan & Kohnert, 2010). The phonological representations of individuals learning a L2 become better specified with increased knowledge of that language (Summers, Bohman, Gillam, Peña, & Bedore, 2010). Those individuals with more phonological knowledge should be less likely to present with a semantic receptive-expressive gap than those individuals with less phonological knowledge.

In this model, underspecified semantic and/or phonological representations may be sufficient for success on less demanding receptive tasks (pointing to pictures) but not more demanding expressive tasks (producing words; Bates, 1993). Such a discrepancy could lead to a semantic receptive-expressive gap.

Suppression (or relative activation) mechanism

It is widely accepted that both of the bilingual speaker's languages are activated when speaking (Costa, 2005). Support for this assertion comes from a variety of methods. For example, Meuter and Allport (1999) asked bilingual speakers to name numbers in both their L1 and L2. Changes in the color of the background signaled to the speaker to change languages Response times were faster when the language of the current and preceding trials were the same. However, response times were slower when there was a language switch. This was interpreted to mean that participants were inhibiting the non-target language, which necessarily meant that both languages were being activated. Results from similar studies have argued that instead of inhibition, languages differed with respect to the magnitude of their activation, with greater activation for the target language (Costa & Santesteban, 2004; Levy, McVeigh, Marful, & Anderson, 2007). In either case, the evidence suggests that both of the bilingual's languages become activated while speaking. When language switch tasks have been performed during the use of event-related potentials, those brain waves associated with inhibition were significantly produced (Jackson et al. 2001; Misra, Guo, Bobb, & Kroll, 2012). And neuroimaging studies have found that brain areas associated with inhibition become activated during switch tasks (Abutalebi & Green, 2008; Guo, Liu, Misra, & Kroll, 2011). Taken together, these studies provide evidence that both of the bilingual's language are activated during communication tasks, and the non-target language becomes inhibited (or relatively deactivated) to allow access to the targeted language.

Several models have invoked a top-down suppression mechanism to account for this inhibition (Green, 1998; Hermans, Bongaerts, De Bot, & Schreuder; Lee & Williams, 2001). For example, in Green's (1998) Inhibitory Control model, a bilingual speaker's two languages share a single lexico-semantic field. Concepts activate lexical representations from both languages, which compete for selection. The item from the non-target language is suppressed after its activation. For example, for a Spanish-English bilingual speaker wishing to communicate the concept "cow" in English, the concept activates both the English word *cow* and its Spanish equivalent *vaca*. After both words have been activated, *vaca* is suppressed, which leaves *cow* for selection.

Although the Inhibitory Control model might be described as a reactive model, proactive inhibition models have also been proposed. Linck, Kroll, and Sunderman (2009) compared the verbal fluency skills of English-speaking college students studying Spanish. One group of students studied in the U.S. with only classroom experiences while others studied in Spain in a Spanish language immersion program. The two groups were carefully matched on their Spanish proficiency. Participants were given 30 seconds to name as many members of conceptual categories as possible. For example, they named as many fruits as they could. This task was performed in both English and Spanish. Both groups produced more responses in English, but there were differences by experience. Analyses revealed that the Spanish-immersion group's ability to perform this task in English was substantially lower than that of their peers studying in the U.S. Simultaneously, the immersion group was able to list more category members in Spanish than their peers studying in the U.S. This was interpreted to mean that when the L2 learner is immersed in an L2 context, the L1 becomes inhibited. This frees cognitive resources to focus on learning the majority language.

Alternatively, instead of appealing to a suppression mechanism, it might be that when an L2 learner is in a majority L2 context, the L2 receives relatively higher levels of activation than the L1. Costa, Santesteban, and Ivanova (2006) proposed just such a relative activation approach. Extending the example above, for a Spanish-English bilingual speaker wishing to communicate the concept “cow” in English a relative activation model would assert that the concept “cow” might activate both *cow* and *vaca* but *cow* would receive greater levels of activation and thus be selected. Suppressed or relatively deactivated representations may be sufficient for success on the less demanding receptive task (pointing to pictures) but not the more demanding expressive task (producing words; Bates, 1993). Such a discrepancy could lead to a semantic receptive-expressive gap.

Compatibility of psycholinguistic mechanisms

It is not necessarily the case that the frequency-lag hypothesis and suppression/relative activation mechanism are mutually exclusive. The frequency-lag hypothesis emphasizes the quality of representations as they are related to frequency of use. However, the frequency of use appears also to be associated with relative activation. For example, in the Bilingual Interactive Activation model and its derivatives (Dijkstra & van Heuven, 1998; Dijkstra & van Heuven, 2002; Grainger, Midgley, & Holcomb, 2010), resting activation levels of phonetic representations lower as frequency of use increases. That is, as the use of a representation increases, the difficulty in accessing that representation decreases. Spreading activation models (Collins & Loftus, 1975) posit that activation of nearby representations spread to adjacent representations in the lexico-semantic space. Therefore, as a language’s use in general increases, the resting activation levels of its lexico-semantic constituents lower. Such a circumstance should predict a reduction of the magnitude of a receptive-expressive gap over time (i.e., with increased experience).

Clinical significance of the semantic receptive-expressive gap

Many standardized language measures provide distinct standard scores for receptive and expressive performance (e.g., the *Clinical Evaluation of Language Fundamentals, Fourth Edition - Spanish*, Wiig, Semel & Secord, 2006; the *Preschool Language Scales - Fifth Edition, Spanish*, Zimmerman, Steiner & Pond, 2012; the *Test of Early Language Development - Third Edition, Spanish*, Ramos, Ramos, Hresko, Reid & Hammill, 2006). Receptive and expressive standard scores are expected to be similar for children with typical language development. In fact, expressive language disorders are diagnosed when receptive standard scores are significantly higher than expressive (American Psychiatric Association, 2000, p. 58). The World Health Organization’s (2005) ICD-10 system has formalized this expectation by providing distinct codes for receptive and expressive language disorders. However, a receptive-expressive gap is associated with typical language development in bilingual children. This gap is larger for bilingual children with language impairment (Gibson et al., 2014b). Understanding the development of the gap should help clinicians accurately interpret language testing results.

Whether the gap is best explained by the frequency-lag hypothesis or a suppression mechanism has direct impact on clinical decisions made by SLPs. For example, if the receptive-expressive gap is the result of a suppression mechanism, then prior to a language assessment, SLPs should spend time performing tasks that diminish the magnitude of such suppression. This might be accomplished in a tier-like fashion by having children watch a movie or television program in the targeted language first, then engage in conversation/play in the targeted language, and then, after the targeted language is highly activated, the formal assessment would begin. However, if the receptive-expressive gap is the result of underspecified phonological representations due to a lack of practice with the language, then activities to prime the targeted language are less important. Instead, clinicians should focus on providing children with the sort of targeted practice that would enhance phonological representations.

Research question

Because it has both theoretical and clinical significance, it is important to understand the development of the semantic receptive-expressive gap. However, no longitudinal data exists to examine its change over time. Therefore, we asked the following question.

1. Is there a receptive-expressive gap in Spanish? If so, does the magnitude of the gap change over time? Could the change or lack of change be explained by language experience factors?

2. Is there a receptive-expressive gap in English? If so, does the magnitude of the gap change over time? Could the change or lack of change be explained by language experience factors?

If the receptive-expressive gap is the result of weak phonological representations or weak links between semantic and phonological representations (as per the frequency-lag hypothesis), then we would anticipate the presence of a receptive-expressive gap in English (L2) but not Spanish, or a comparatively diminished gap in Spanish compared to English. On the other hand, if the receptive-expressive gap is the result of a suppression mechanism (or relative activation), then we would anticipate a gap in Spanish (L1) but not English, or a comparatively diminished gap in English compared to Spanish. Furthermore, if indeed the mechanism resulting in a gap is suppression (or relative activation), then suppression should occur both for those individuals with much and those with little exposure to English. Therefore, if a model built on indicators of language experience fits well with the receptive-expressive gap, then we can reason that the mechanism underlying the receptive-expressive gap is related to language experience and not suppression.

Methods

Participants

Participants were drawn from a longitudinal investigation of diagnostic markers of language impairment in bilingual children (see Bohman, Bedore, Peña, Mendez-Perez, & Gillam, 2010; Peña, Gillam, Bedore, & Bohman, 2011). In the first phase of a three-phase study, 1192 bilingual children from school districts across the USA were screened for semantic knowledge. In the second phase, roughly six months later, 186 of those children were administered a wide range of language measures using an experimental version of the Bilingual English Spanish Assessment (BESA; Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2018), including measures of semantic, morphosyntactic and phonological knowledge. Due to attrition, 19 of the 186 children did not participate in the third phase a year later; the remaining 167 children were re-administered the same tests from phase two. Gillam, Peña, Bedore, Bohman, and Mendez-Perez (2013) reported on the 167 children's performance on additional language measures that are not included in the current study.

All participants had a minimum of 20% of their current linguistic experience in either Spanish or English, based on results of a parent questionnaire (Peña et al., 2018; see Table 1 for participant characteristics). Children scored within typical limits on the Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998), M = 100.27, SD = 11.65. They were all rated as typically developing in language abilities based on a 6-point scale from Records and Tomblin (1994). Participants passed a hearing screening, and there were no reported histories of behavioral, social, or emotional concerns. Only children with complete data sets at both timepoints were included, which reduced the pool of participants to 108 children. We excluded from analysis two children whose standard scores exceeded three standard deviations on the target variable, which we interpreted as outliers. The final number of participants was 106 (57 boys and 49 girls).

On average children in the current study were slightly over 5 and a half years old at kindergarten. Mother's education was rated on a 7-point scale (Hollingshead, 1975), from no formal education (1) to a graduate degree (7). Mothers in the current study had an average Hollingshead score of 2.84 (1.71), signifying partial high school attendance, which is interpreted as reflecting low socioeconomic status. On average, children in this study learned Spanish first and were introduced to English at an early age. They had more day-to-day experience in Spanish compared to English, and performed better on Spanish compared to English tests.

Measures

Language history questionnaire—In addition to the demographic information above, caregivers provided information about participants' first English experience and current Spanish and English experience by responding to a parent questionnaire (Peña et al., 2018). This included an hour by hour report of language input and output. Research assistants fluent in Spanish spoke to caregivers by telephone to administer the questionnaire. The caregiver was asked to consider the child's language context for each hour of the day, from 7:00 a.m. until 8:00 p.m. This included a report of the activity taking place, the language being heard by the child (English, Spanish, or both), and the language being spoken by the child (English, Spanish, or both). For example, the child's context on an average weekday from 6:00 p.m. - 7:00 p.m., might be dinner, with the child hearing Spanish and the child producing both English and Spanish. The parent questionnaire was augmented by a teacher questionnaire reporting language input and output, and the two were combined to calculate a measure of current language experience (Bedore, Peña, Summers, Boerger, Resendiz, Greene, & Gillam, 2012). Caregivers also provided the age at which participants began to be regularly exposed to English, which we treated as a measure of cumulative experience with English.

Experimental version of BESA semantics subtest—Children were administered an experimental version of the BESA (Peña et al., 2018). This is a standardized language test that assesses Spanish and English language ability in U.S. Spanish-English bilingual children. The test includes subtests for morphosyntax and semantics in both languages, but only results from the latter were included in the current study. Children simultaneously heard brief stories about culturally relevant events (e.g., a birthday party) and saw associated illustrations. Six types of questions were included to measure semantic knowledge, and for each question type there was a receptive portion and an expressive portion. Children generated or identified categories, functions, similarities/differences, analogies, characteristic properties, and where/who questions. Accurate responses in either English or Spanish were credited as correct, independent of the target language (see Bedore, Peña, Garcia, & Cortez, 2005; Peña et al., 2003). Items across the two languages were not translation equivalents but were equivalent in both by type and level of difficulty.

BESA testers were fluent in the language of testing. Children were tested in a single language per session, and both the morphosyntax and semantics subtests for a single language were completed in a single session lasting about 30 minutes (roughly 15 minutes per subtest).

Standardization

We mathematically controlled for the inherent differences in difficulty between the BESA receptive and expressive semantics tasks through standardization. We drew our comparison group from functionally monolingual participants in English (n = 143) and Spanish (n = 44) from several representative communities across the USA, including Pennsylvania, Georgia, Texas, and California (independent from the participants in the current study). Functionally monolingual individuals were those who had exposure to both languages but had 20% or less exposure in the non-dominant language. Controlling for age, average raw scores were derived for receptive and expressive tasks. We followed Allen and Yen's (1979) standardization procedures, converting raw scores to Z-scores, and Z-scores to standard scores with a Mean of 100 and Standard Deviation of 15. We operationalized the semantic receptive-expressive gap as occurring when there was a statistically significant difference between children's receptive and expressive standard scores on the BESA semantics test with receptive scores exceeding expressive scores. For the participants in the current study, most children's receptive standard scores were greater than expressive (though not necessarily at a statistically significant level). However, at kindergarten, expressive was greater than receptive for 15 and 28 of these children in Spanish and English, respectively. At first grade, expressive exceeded receptive for 16 and 25 children in Spanish and English, respectively.

Standardized scores were based on conceptual scoring, which was the method used in the development of the BESA. This form of scoring accepts correct responses from either language. For example, Spanish-English bilingual children taking the Spanish version of the test might code-switch to English in order to provide an accurate answer. This is important to consider given that the theoretical models mentioned above make predictions based on one language compared to the other. However, as reported by Greene, Peña, & Bedore (2013), such code-switching was rare. We reviewed the language-specific raw scores (i.e., not accepting correct answers from the non-target language) and the conceptual raw scores. There were no differences in receptive scoring for the two scoring regimens. Conceptual scoring improved expressive scores in English by .5 of a raw score point and improved expressive scores in Spanish by on average 1.25 raw score points. These differences are modest and do not impact the substance of the analysis using conceptual scoring; therefore, standard scores based on conceptual scoring are reported here.

Analyses

The current study included three independent variables each at two levels: Modality (Receptive, Expressive), Language (Spanish, English), and Time (Kindergarten, First Grade). This study was multivariate in nature, with eight scores for each subject (four measures at two time points). Although response vectors from different subjects were independent, the eight scores for any individual participant were likely correlated. To allow for these potential correlations, we used a profile analysis using Hotelling's multivariate T-statistic that takes into account the main effects and interaction effects of the treatment structure and that also takes into account the fact that responses were obtained from each subject under all eight treatment combinations.

Additionally, we performed a path analysis to determine the influence that language exposure variables had on the receptive-expressive gap at the two time points in the two languages while controlling for the potential influence of maturation. Path analysis is a comprehensive approach that tests hypotheses about relations among variables. We focused on percent of English exposure at Kindergarten and at First Grade. And because maturation is always an important feature of language development, we also focused on children's age at the two time points.

Results

To test for correlations across languages which might influence the results of our multivariate analysis, we performed a likelihood ratio test, χ^2 (16) = 34.32, $p = .005$, that indicated at least one of the correlations across languages was statistically significant. Follow-up testing found a correlation of .21 between English and Spanish Receptive scores at Kindergarten, which was considered weak but statistically significant, $p = .03$, 95% CI [.01, .38]. Even though the correlations between dependent variables across languages were weak, subsequent analyses incorporated a completely unstructured covariance matrix to guard against over-constraining the covariance matrix, thereby accounting for all possible correlations between dependent variables.

To determine whether there were statistically significant changes in performance across the year and whether those changes varied based on the language of testing, we performed a profile analysis using Hotelling's multivariate T-statistic. There was a statistically significant main effect for Modality, $F(1, 105) = 210.872$, $p < .0001$, 95% CI [11.13, 14.65] and a statistically significant interaction between Language and Time, $F(1, 105) = 6.04$, $p = .02$. However, there was not a statistically significant three-way interaction, $F(1, 105) = 2.20$, $p = .14$. Additionally, there were not statistically significant interactions between Modality and Language, $F(1, 105) = .25$, $p = .62$, or between Modality and Time, $F(1, 105) = .32$, $p = .57$ (see Figure 1).

The effects of Language and Time were evaluated within the levels of the other. It was not meaningful to evaluate their main effects alone because they significantly interacted. There was a statistically significant difference between overall Spanish and English scores at kindergarten, $F(1, 105) = 9.0$, $p = .003$, 95% CI [2.30, 11.29], but not at first grade, $F(1, 105) = 1.93$, $p = .17$, 95% CI [−1.23, 7.02]. There were statistically significant increases in scores between Kindergarten and First Grade for both Spanish, $F(1, 105) = 49.92$, $p < 0001$, 95% CI [6.35, 11.30] and English, $F(1, 105) = 94.14$, $p < .0001$, 95% CI [10.13, 15.33].

There was an overlap in the confidence intervals for Time in both Spanish and English, which suggested no statistically significant difference in the change across time between the two languages. However, the existence of an interaction between Language and Time precluded this. We analyzed the difference in the time effect for the two languages. The point estimate of the difference between Time for the English and Spanish was 3.90. In other words, the increase in mean scores across the year was 3.90 units higher for English than Spanish, 95% CI [0.75, 7.05].

In order to examine the influence of language exposure variables on the receptive-expressive gap at two time points in two languages, we performed a path analysis. By convention, indices of fit confirm if a model acceptably fits the data. These indices included the Goodness of Fit Index (GFI), the Adjusted Goodness of Fit Index (AGFI), the Bentler Comparative Fit Index (BCFI), and the Root Mean Square Error of Approximation (RMSEA). In the current study, our model was found to fit well with the data: GFI = .99; AGFI = .99; BCFI = .98; RMSEA = .05. Figure 2 details the model with its results. Although the overall model fit the observed outcomes, only two of the variables were uniquely significant in explaining variance. Age at Kindergarten predicted Spanish receptive-expressive gap at Kindergarten, $t = -2.26$, $p = .02$, $r^2 = .09$, and English exposure at Kindergarten predicted English receptive-expressive gap at Kindergarten, $t = 2.90$, $p = .003$, $r^2 = .04$.

As a post-hoc analysis, we performed a Chi Square to examine the distribution of children with and without semantic receptive-expressive gaps, here operationalized as receptive standard scores 15 points or greater than expressive standard scores (see Table 2 for distribution). We created four distribution groups within each language. Children in the first group did not have a gap at either kindergarten or first grade. In the second group, children presented with a gap at both kindergarten and first grade. Children in the third group had a gap only at kindergarten, and children in the fourth group had a gap only at first grade. Because the multivariate analysis was based on group statistics it potentially masked individual change, thus the results of our multivariate analysis suggested that the same children who had a semantic receptive-expressive gap in kindergarten also had one in first grade. However, Chi squared follow up analyses showed that in English, the distribution across the four categories was unequal, with more children in the *never* category than anticipated, $X^2(3, 106) = 12.26$, p = .006. On the other hand, the receptive-expressive relationship was generally equally distributed across the four categories in Spanish, $X^2(3, 106) = 5.85$, p = .11. That is, the observed number of children who fell into each of the four categories was not statistically significantly different than the expected number of children in each category.

Discussion

We sought to identify changes in Spanish-English bilingual children's receptive-expressive semantic gaps in their two languages across a year. Children in the current study performed better in Spanish than in English but improved in both languages across the year, with more improvement in English than Spanish. Despite the improvement, the magnitude of the gaps in both English and Spanish did not change with statistical significance across the year. We proposed that support for the frequency-lag hypothesis would result from a gap in children's English but not Spanish, or at least to a significantly lesser degree in Spanish than in English. We also proposed that support for a suppression mechanism would result from a gap in Spanish but not English, or at least to a significantly lesser degree in English than in Spanish. Since there was a similar gap in both languages, this outcome does not fit neatly with either of our predictions.

We take the absence of such patterns as support for the assertion that the frequency-lag hypothesis and the suppression mechanism/relative activation models are not mutually exclusive, as indicated in the literature review above. Frequency effects impact the quality of linguistic representations (*a la* the frequency-lag hypothesis) as well as those representations' levels of activation (*a la* the suppression mechanism/relative activation models). Therefore, the two phenomena might be operating simultaneously in the current study. We reason that it is unlikely that English would be suppressed by an English learner in an English majority context. But it is likely that because they are L2 learners, they would have sparse representations of words and so not be able to produce items, which is consistent with the frequency-lag hypothesis. Therefore, although the suppression mechanism/relative activation might contribute to the gap in L1, we posit that it is unlikely to contribute to the gap in L2. Unfortunately, the design of the current study prevents us from falsifying either of these hypotheses. Future studies using an experimental design rather than the current correlational design should be able to tease apart these two explanations.

Participants were on average 2;3 when they began to be regularly exposed to English, meaning that by the time they were in kindergarten they had had roughly three and a half years of English language exposure. Roughly 45% of their current daily language exposure was in English. Therefore, the children in the bilingual context likely had weaker Spanish language representations than did functionally monolingual Spanish-speaking peers. The relatively weak representations should contribute to a receptive-expressive gap.

A comparison of the expressive scores at first grade to the receptive scores a year earlier at kindergarten shows that there was a lag between the time these children could match their functionally monolingual same-age peers on the receptive task and the time they could match them on the expressive task. Because receptive performance continued to improve, expressive performance, although itself improving, maintained its distance relative to receptive performance. Consistent with both the frequency-lag hypothesis and the suppression mechanism/relative activation, expressive performance should match receptive performance once receptive performance has remained stable over time. For children in the current study, average receptive standard scores for both English and Spanish had inched above average by first grade, thus one might anticipate a flattening out of these children's receptive semantic performance in the near future, allowing expressive performance to 'catch up'. Future studies should explore this possibility longitudinally.

We performed a path analysis to examine the role of language experience on the receptive-expressive gap. The results reveal that the observed outcomes fit well with our hypothesized model in which age and language experience exert a significant influence on the receptive-expressive gap. Two variables uniquely explain variance in the gap. Age at kindergarten explains 9% of variance in the Spanish receptive-expressive gap at kindergarten. This negative relationship means that older children have smaller gaps in Spanish than do their younger peers. However, Age does not also explain variance in the English gap. In addition to maturation, Age captures children's lifelong experience with Spanish, and older children had more experience with Spanish. However, age ceases to be a uniquely significant factor by the first grade. This suggests that although age is important for the gap in the first language during the earlier stages of academic life in the U.S., its importance diminishes after a year's experience in the school system. It is likely, therefore, that factors related to academics (e.g., early literacy skills, metalinguistic awareness, syntagmatic and paradigmatic connections between concepts, etc.) begin to exert more influence on the gap than does Age. For English, however, children were first regularly exposed to English (L2) at on average 2.25 years of age. Therefore, Age in relation to English likely reflects maturation more than language experience.

Children's English language experience at kindergarten uniquely explains 4% of the variance in the English gap at kindergarten. The positive relationship indicates that the receptive-expressive gap in English grows with increases in English exposure. This is reasonable since those children who are most exposed to English likely will expand their receptive vocabularies fastest. Therefore the discrepancy between what they understand and what they can produce will be greater than for children with smaller receptive vocabularies. However, a similar relationship is not found at first grade. We posit that the importance of English language experience is diminished as its influence is offloaded onto those capacities and skills picked up in the first year of academic experience (e.g., early literacy). Although English language experience explains some variance in English, it does not uniquely explain variance in Spanish. This likely is due to the fact that, on average, these children had well-established Spanish language representations and this was coupled with slightly more day-to-day experience with Spanish than with English. Therefore, these children's Spanish abilities did not attrite in the way some children's first language does when immersed in a second language.

To more deeply understand the distribution of scores across time, we performed a post-hoc analysis using Chi Square. Some children's receptive and expressive scores converged across the school year, while others' diverged. Some children never had a gap and some always had a gap. Convergence indicates that a year's worth of practice likely enhances the quality of English representations, thus closing the gap. Divergence likely is the result of some children beginning kindergarten with very little target language knowledge and so have very low scores on both receptive and expressive measures; only after a year of language learning does receptive performance advance sufficiently for the gap to be identified. Children who maintain the gap across the year are representative of the larger group. However, there are more children than anticipated who do not experience a receptive-expressive gap in English at either kindergarten or first grade. This is likely due to the fact that this category captures both those children whose scores are substantially low (the target language has not sufficiently developed for the gap to be identified) and those whose scores are substantially high (expressive language already has 'caught up' with receptive language in kindergarten).

We cannot rule out some possible influence of the norming sample in the current results. In the studies reported in the literature review, children either presented with a receptive-expressive gap in Spanish but not English (when comparing single picture naming to picture pointing tasks) or in English but not Spanish (when using broader measures of semantic ability like the one used in the current study). However, in the current study, children presented with receptive-expressive gaps of similar magnitude in both their Spanish and English. It is possible that this result was influenced to at least some degree by the way in which the norming sample was established. In previous work using broad measures of semantic ability, the norming samples were created from the performance of children participating in the study. For example, the means and standard deviations of the performance of a subset of the participants in Gibson et al. (2014b) served as the norming group. In the current study, the norming sample was based on an independent group of children on whom the BESA itself was based. Therefore, the current study likely reflects the reality of these children's semantic abilities more precisely than in previous studies.

Clinical implications

Clinicians should not be surprised when standard scores for receptive exceed expressive semantics abilities, even when Spanish-English bilingual children are typically developing. Additionally, they should expect that the magnitude of this discrepancy will be similar in Spanish and English. If children from this study are representative, overall semantics scores for typically developing bilingual children should be near the mean of the performance of their functionally monolingual peers, despite the presence of a receptive-expressive gap. Consistent with Gibson et al. (2014b), clinicians should not assume an expressive semantic disorder due to a large receptive-expressive gap if the overall semantic standard scores are within typical limits.

It is not clear whether this phenomenon is the result of underspecified phonological representations due to a lack of practice with the language or due to a suppression mechanism/low activation levels. If both are operant, then clinicians might construct their interventions anticipating the effects of both. To address weak representations due to a frequency effect (the frequency-lag hypothesis), clinicians should focus on providing more targeted practice by increasing time to task. This might be accomplished by increasing the length of time of each session, adding sessions, or combining one-to-one sessions with group sessions. In countries like the U.S., this might be difficult to accomplish in the face of insurance coverage. In those cases, clinicians could extend children's time to task by providing training and home programs that parents could carry out. To address a suppression mechanism/low activation levels, clinicians could perform tasks prior to intervention that would increase the targeted language's activation levels. For example, prior to a session, children could watch a movie in the targeted language, play games in that language, look at a book in that language, converse with a peer in that language, etc. If this sort of pre-session task is not undertaken, precious moment of the clinical session will be devoted to weakening the effects of a suppression mechanism/low activation levels.

Summary and Conclusion

A semantic receptive-expressive gap appears to be a feature of typical bilingual language development. At the group level, the magnitude of this gap remains relatively constant at kindergarten and first grade in both English and Spanish, despite overall improvements in semantic performance in both languages. Both Age and language experience exert an influence on the magnitude of the receptive-expressive gap at kindergarten, with younger children having a larger gap in Spanish, and children with greater English language experience having a larger gap in English. However, by first grade these influences diminish. It appears that there is on average roughly a one year lag between receptive and expressive semantics skills. However, there are subgroups of children whose receptive-expressive gaps converge or diverge across the year, while others never experience a gap, and yet others whose gap persists at both time points. Therefore, clinicians should take caution in interpreting receptive-expressive semantic gaps.

Acknowledgements

This work was supported by the grant R01DC007439 from the National Institute on Deafness and Other Communication Disorders (NIDCD). 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

Abutalebi J, & Green D (2007). Bilingual language production: The neurocognition of language representation and control. Journal of Neurolinguistics, 20, 242–275.

Allen MJ & Yen WM (1979). Introduction to measurement theory. Monterrey, CA: Brooks/Cole.

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders IV Text Revision. Washington, DC: American Psychiatric Association.

Barnett WS, Yarosz DJ, Thomas J, Jung K & Blanco D (2007). Two-way and monolingual English immersion in preschool education: An experimental comparison. Early Childhood Research Quarterly, 22, 277–293.

Bates E (1993). Comprehension and production in early language development: Comments on Savage-Rumbaugh et al. Monographs of the Society for Research in Child Development, 58, 222–242. [PubMed: 8366873]

Bedore LM, Peña ED, Garcia M & Cortez C (2005). Conceptual versus monolingual scoring: When does it make a difference? Language, Speech, and Hearing Services in Schools, 36, 188–200.

Author Manuscript

Author Manuscript

Author Manuscript

Bedore LM, Peña ED, Summers CL, Boerger KM, Resendiz MD, Greene K, … & Gillam RB (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. Bilingualism: Language and Cognition, 15, 616–629.

Bohman TM, Bedore LM, Peña ED, Mendez-Perez A & Gillam RB (2010). What you hear and what you say: Language performance in Spanish–English bilinguals. International Journal of Bilingual Education and Bilingualism, 13, 325–344. [PubMed: 21731899]

Bracken BA & McCallum RS (1998). Universal Nonverbal Intelligence Test. Austin, TX: Riverside Publishing Company.

Collins Allan M.; Loftus Elizabeth F. (1975). “A spreading-activation theory of semantic processing”. Psychological Review. 82 (6): 407–428

Costa A (2005). Speech production in bilinguals. In Bhatia T & Ritchie W (eds.), The handbook of bilingualism, pp. 201–223. Malden, MA & Oxford: Blackwell.

Costa A, & Santesteban M (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. Journal of Memory and Language, 50, 491–511.

Costa A, Santesteban M, & Ivanova I (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,1057–1074.

Dollaghan CA (1987). Fast mapping in normal and language-impaired children. Journal of speech and hearing disorders, 52(3), 218–222.

Gibson TA, Jarmulowicz L & Oller DK (2018). Difficulties using standardized tests to identify the receptive expressive gap in bilingual children's vocabularies. Bilingualism: Language and Cognition, 21, 328–339.

Gibson TA, Oller DK, Jarmulowicz L & Ethington CA (2012). The receptive–expressive gap in the vocabulary of young second-language learners: Robustness and possible mechanisms. Bilingualism: Language and Cognition, 15, 102–116.

Gibson TA, Peña ED & Bedore LM (2014a). The relation between language experience and receptive-expressive semantic gaps in bilingual children. International Journal of Bilingual Education and Bilingualism, 17, 90–110. [PubMed: 29670456]

Gibson TA, Peña ED & Bedore LM (2014b). The receptive–expressive gap in bilingual children with and without primary language impairment. American Journal of Speech-Language Pathology, 23, 655–667. [PubMed: 25029625]

Gillam RB, Peña ED, Bedore LM, Bohman TM, Mendez-Perez A (2013). Identification of specific language impairment in bilingual children, part 1: Assessment in English. Journal of Speech, Language, and Hearing Research, 56, 1813–1823.

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

Gollan TH, Fennema-Notestine C, Montoya RI & Jernigan TL (2007). The bilingual effect on Boston Naming Test performance. Journal of the International Neuropsychological Society, 13, 197–208. [PubMed: 17286875]

Gollan TH, Montoya RI, Fennema-Notestine C & Morris SK (2005). Bilingualism affects picture naming but not picture classification. Memory & cognition, 33, 1220–1234. [PubMed: 16532855]

Green DW (1998). Mental control of the bilingual lexicosemantic system. Bilingualism: Language and Cognition,1, 67–81.

Hermans D, Bongaerts T, De Bot K, & Schreuder R (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? Bilingualism: Language and Cognition, 1 (3), 213–229.

Indefrey P, & Levelt WJ (2004). The spatial and temporal signatures of word production components. Cognition, 92(1-2), 101–144. [PubMed: 15037128]

Ivanova I & Costa A (2008). Does bilingualism hamper lexical access in speech production? Acta Psychologica, 127, 277–288. [PubMed: 17662226]

Jackson GM, Swainson R, Cunnington R, & Jackson SR, 2001. ERP correlates of executive control during repeated language switching. Bilingualism: Language and Cognition, 4,169–178.

Kohnert KJ, Hernandez AE & Bates E (1998). Bilingual performance on the Boston Naming Test: preliminary norms in Spanish and English. Brain and language, 65, 422–440. [PubMed: 9843612]

Leonard L (2009). Is expressive language disorder an accurate diagnostic category? American Journal of Speech-Language Pathology, 18, 115–123. [PubMed: 19029534]

Lesaux NK, Crosson AC, Kieffer MJ & Pierce M (2010). Uneven profiles: Language minority learners' word reading, vocabulary, and reading comprehension skills. Journal of Applied Developmental Psychology, 31, 475–483. [PubMed: 21243117]

Levelt WJ, Roelofs A, & Meyer AS (1999). A theory of lexical access in speech production. Behavioral and brain sciences, 22(1), 1–38.

Levy BJ, McVeigh ND, Marful A, & Anderson MC (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition. Psychological Science, 18, 29–34. [PubMed: 17362374]

Linck J, Kroll J, & 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, 1507–1515. [PubMed: 19906121]

Martin NA, & Brownell R (2011). Expressive One-word Picture Vocabulary Test - 4: Spanish Bilingual Edition. Novato, CA: Academic Therapy Publications.

Martin NA, & Brownell R (2011) Receptive One-word Picture Vocabulary Test - 4: Spanish Bilingual Edition. Novato, CA: Academic Therapy Publications.

Meuter RFI, & Allport A (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. Journal of Memory and Language, 40, 25–40.

Miccio A, Tabors PO, Páez MM, Hammer CS, & Wagstaff D (2003, April). Vocabulary development in Spanish-speaking Head Start children of Puerto Rican descent. In 4th International Symposium on Bilingualism, Tempe, AZ.

Misra M, Guo T, Bobb S, and Kroll J (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition o =f the native languages. Journal of Memory and Language, 67, 224–237.

Oller DK & Eilers RE (2002). Language and literacy in bilingual children. Clevedon, UK: Multilingual Matters.

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

Peña ED, Bedore LM, Iglesias A, Gutiérrez-Clellen VF, & Goldstein BA (2014). Bilingual English Spanish Oral Screener–Experimental Version (BESOS). Unpublished instrument.

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

Peña ED, Gillam RB, Bedore LM & Bohman TM (2011). Risk for poor performance on a language screening measure for bilingual preschoolers and kindergarteners. American Journal of Speech-Language Pathology, 20, 302–314. [PubMed: 21821821]

Peña ED, Gutierrez-Clellen V, Iglesias A, Goldstein B & Bedore LM (2018). BESA: Bilingual English-Spanish Assessment Manual. Baltimore, MD: Brookes.

Ramos M, Ramos J, Hresko WP, Reid DK & Hammill DD (2006). Test of Early Language Development (TELD): Spanish. Austin, TX: PRO-ED.

Records NL & Tomblin JB (1994). Clinical decision making: Describing the decision rules of practicing speech-language pathologists. Journal of Speech, Language, and Hearing Research, 37, 144–156.

Summers C, Bohman TM, Gillam RB, Peña ED, & Bedore LM (2010). Bilingual performance on nonword repetition in Spanish and English. International Journal of Language & Communication Disorders, 45(4), 480–493. [PubMed: 20565327]

Uchikoshi Y, & Maniates H (2010). How does bilingual instruction enhance English achievement? A mixed-methods study of Cantonese-speaking and Spanish-speaking bilingual classrooms. Bilingual Research Journal, 33(3), 364–385.

Wiig EH, Semel E & Secord WA (2006). Clinical Evaluation of Language Fundamentals, Fourth Edition – Spanish (CELF-4). San Antonio, TX: Pearson.

Windsor J, & Kohnert K (2004). The search for common ground: Part I. Lexical performance by linguistically diverse learners. Journal of Speech, Language, and Hearing Research, 47(4), 877–890.

World Health Organization. (2005). International statistical classification of diseases and related health problems, 10th revision. Geneva.

Zimmerman IL, Steiner VG & Pond RE (2012). PLS-5: Preschool Language Scales – Fifth Edition, Spanish. San Antonio, TX: Pearson.

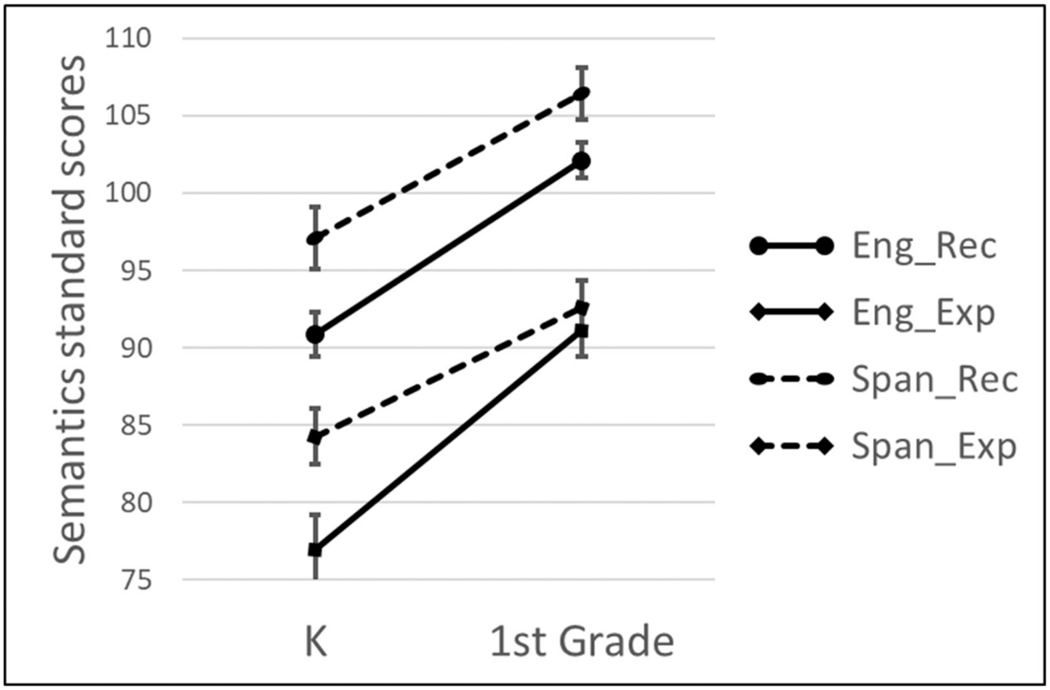

Figure 1.

Receptive and expressive semantics scores for English and Spanish in kindergarten and first grade. Bars indicate standard error. Eng_Rec = English semantic receptive score; Eng_Exp = English semantic expressive score; Span_Rec = Spanish semantic receptive score; Span_Exp = Spanish semantic receptive score.

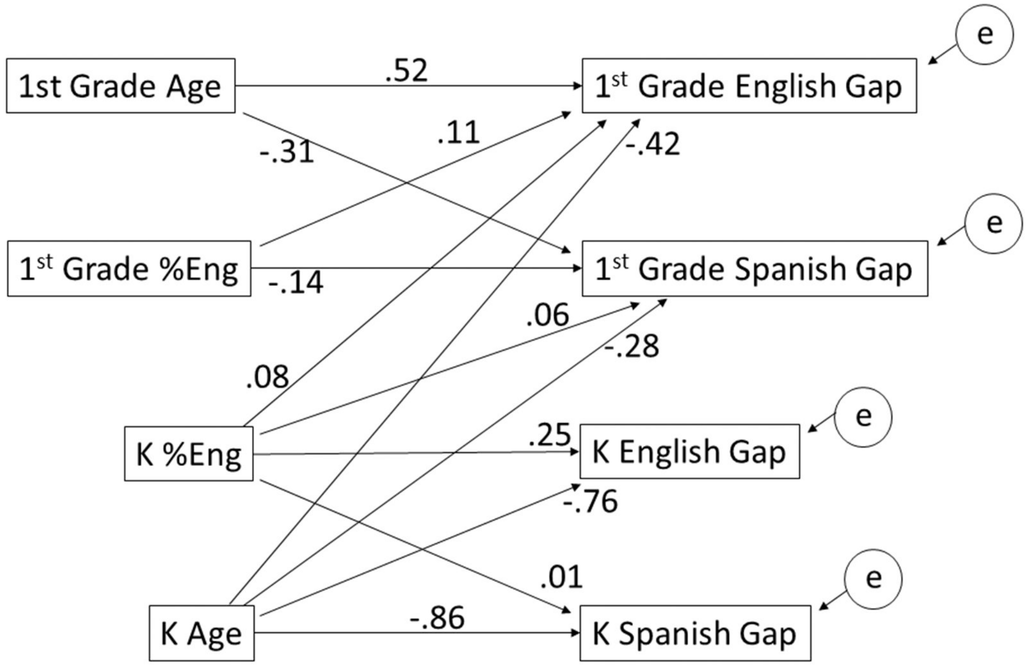

Figure 2.
Path analysis conducted to assess age and language exposure as predictors of the receptive expressive gap in Spanish and English.

Table 1.

Participant characteristics at kindergarten (N = 106)

	Mean (SD)
Age (months) at kindergarten	69.68 (4.20)
Age (months) at first grade	82.68 (4.64)
Mother education (Hollingshead)	1.84 (1.71)
Age English exposure (years)	2.25 (1.78)
% English daily exposure (at kindergarten)	44.13 (21.04)
% English daily exposure (at first grade)	47.76 (20.74)

Note. Age based on child's age at first day of that year's testing. Mother’s education is based on an eight point scale developed by Hollingshead (1975) where 0 = no formal education and 7 = graduate degree. A score of 1.84 indicates an eighth grade education. Age English exposure indicates the age at which children were first regularly exposed to English.

Table 2.

Distribution of the semantic receptive-expressive gap among participants

	Never	Always	K-only	1st Grade only
Spanish	32	16	30	28
English	42	21	23	20

Note. *Never* indicates no semantic receptive-expressive gap was present at either K (Kindergarten) or 1st grade. *Always* indicates the presence of a semantic receptive-expressive gap at both K and 1st grade.