ASHA American Speech-Language-Hearing Association

AJSLP

Research Article

Semantic Difficulty for Bilingual Children: Effects of Age, Language Exposure, and Language Ability

Stephanie McMillen,[a] Nahar Albudoor,[b] Elizabeth D. Peña,[c] and Lisa M. Bedore[d]

[a] Department of Communication Sciences & Disorders, Syracuse University, NY [b] Department of Human Services and Sciences, Gallaudet University, Washington, DC [c] School of Education, University of California, Irvine, CA [d] Department of Communication Sciences & Disorders, Temple University, Philadelphia, PA

ARTICLE INFO

Article History:
Received January 19, 2022
Revision received September 1, 2022
Accepted December 2, 2022

Editor-in-Chief: Erinn H. Finke
Editor: Anny Castilla-Earls

https://doi.org/10.1044/2022_AJSLP-22-00018

ABSTRACT

Purpose: Semantic tasks evaluate dimensions of children's lexical-semantic knowledge. However, the relative ease of semantic task completion depends on individual differences in developmental and language experience factors. The purpose of this study was to evaluate how language experience and language ability impact semantic task difficulty in English for school-age Spanish–English bilingual children with and without developmental language disorder (DLD).

Method: Participants included 232 Spanish–English bilingual children in second through fifth grade with (n = 35) and without (n = 197) DLD. Data included children's performance on the English Semantics subtest of the Bilingual English–Spanish Assessment—Middle Extension Field Test Version (BESA-ME), age of English acquisition, and percent English language exposure. Task difficulty, a measurement of the relative ease of task completion, was calculated for six semantic task types included on the BESA-ME. Multilevel regression modeling was conducted to estimate longitudinal growth trajectories for each semantic task type.

Results: Results showed that language ability and grade level drive semantic task difficulty for all task types, and children with DLD experienced greater difficulty on all task types compared to their typically developing peers. Longitudinally, semantic task difficulty decreased for all children, regardless of language ability, indicating that semantic task types became easier over time. While children made gains on all semantic tasks, the growth rate of task difficulty was not equal across task types, where some task types showed slower growth compared with others. English language exposure emerged as a significant predictor of semantic task difficulty while age of acquisition was not a significant factor.

Conclusions: This study clarifies developmental profiles of lexical-semantic performance in bilingual children with and without DLD and supports clinical decision-making regarding children's English language learning.

What does it mean to "know" a word? *Knowing* a word requires developing a representation by mapping both the form—including the phonological constituents (e.g., /k/ + /ʌ/ + /p/ = *cup*)—and the semantic characteristics to the lexical item. While establishing an initial lexical entry with corresponding form requires only a few exposures through quick incidental learning for typically developing (TD) children (Rice, 1990), building semantic depth requires multiple exposures and use of the word across a variety of contexts (Hills et al., 2010). Rich, interconnected networks of semantic knowledge underpin robust word representations, and these are continually developed over time to create nuanced word meanings within the lexicon (Bloom, 2002). Theoretically, when a child has a robust representation for a lexical item, the child fully

Correspondence to Stephanie McMillen: smcmille@syr.edu. ***Disclosure:*** *Elizabeth D. Peña and Lisa M. Bedore receive royalties from sales of the Bilingual–English-Spanish Assessment (BESA), which was one of the measures used to classify children with or without DLD. The other authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

"knows" the meaning, form, and use for that word. In assessment of semantics, a robust representation of a word is likely associated with ease of performance on items that include that specific word. In contrast, having a sparse representation for a word likely results in increased difficulty for a specific test item; sparse representations for multiple words would result in increased difficulty for several test items, decreasing the child's overall score on the semantic task. The robustness of word representations is impacted by several intersecting factors, including sociocultural variations, cognitive-linguistic demands of test items, and developmental factors inherent to the child, such as language ability, language exposure, and cognitive maturation. Variations in these factors influence the robustness of children's lexical-semantic knowledge and, in turn, the relative ease children experience when completing semantic tasks. From a measurement perspective, children's accuracy on lexical-semantic tasks reflects their relative difficulty level, where higher accuracy represents advanced knowledge in the targeted lexical-semantic domain. The purpose of this study was to evaluate how language experience and language ability affect task difficulty for a semantics subtest designed to account for cultural and linguistic differences in school-age Spanish–English bilingual children with and without developmental language disorder (DLD).

Semantic Task Types

Semantic tasks are designed to assess the breadth and depth of children's word knowledge, including the number of lexical items that are stored in long-term memory and how much semantic information is connected to each word. While breadth and depth are interrelated, semantic tasks are created to elicit different aspects of lexical-semantic knowledge. For example, naming items on a single word picture vocabulary test targets the breadth of lexical knowledge, while describing how three items are alike captures depth of semantic knowledge. For children with DLD, single word picture vocabulary tests are poor for identification purposes (Gray et al., 1999); however, children with DLD often have deficits in semantic depth due to sparse semantic networks (Sheng et al., 2012), causing poor performance on tasks reflecting this language domain (Jasso et al., 2020). For monolingual children with DLD, deficits on lexical-semantic tasks persist through high school (McGregor et al., 2013).

Performance on semantic tasks provides key evidence needed for differential diagnosis of DLD. Semantic tasks are included within a subtest on the Bilingual English–Spanish Assessment–Middle Extension Field Test Version (BESA-ME; Peña, Bedore, Gutiérrez-Clellen, et al., 2016), which is a standardized assessment under development for Spanish–English bilinguals ages 7–12 years. The BESA-ME Semantics subtest includes seven semantic task types that probe a wide range of lexical-semantic abilities: analogies, associations, categories, characteristics, definitions, functions, and similarities and differences. Each semantic task type was created to assess meaning-based knowledge within the language system. The analogies task measures children's ability to describe items by relationship; items may be related by functions, anonyms/synonyms, characteristic properties, part-whole, and/or category membership (e.g., "Green is to grass as yellow is to _____"). The associations task provides children with a word and asks them to state three additional words that correspond to the target word (e.g., "If I say bird, you say 1. ___, 2. ___, 3. ___"). The categories task evaluates a child's ability to classify items into groups based on shared characteristics (e.g., Show me all of the zoo animals"). The characteristics task requires children to describe perceptual features and functions of objects (e.g., "Tell me three things about a bicycle"). The definitions task evaluates children's ability to describe developmentally appropriate vocabulary terms. The functions task asks children to name an action that corresponds with a specified item (e.g., "What do you do with a computer?"). The similarities and differences task requires children to demonstrate how a set of pictured objects is either alike or different (e.g., "How are *pencil*, *crayon*, and *marker* alike?"). Together, patterns of strengths and weaknesses across these semantic task types ultimately support clinical decision-making regarding eligibility for diagnosis and intervention services. While language ability significantly affects test performance, clinicians must also account for other factors unique to each child that influence developmental changes in lexical-semantic knowledge over time in order to understand profiles of lexical-semantic knowledge bilingual children. In this study, the factors we focus on include language exposure, age of acquisition (AoA) for English, educational experience represented by grade, and the cognitive-linguistic processes implicated in each semantic task type.

Language Experience and Semantic Performance

Language experience, including AoA and the amount of exposure to each language, drives development of lexical-semantic knowledge (Bedore et al., 2016). Differences in opportunities for word learning occur based on the sociocultural environment, including the language(s) used by the cultural community and environmental context (e.g., home, school). For example, children in U.S. public school systems are more likely to learn the words *desk* and *homework* in English because these are high frequency words in predominantly English-speaking environments, whereas home-based words, such as *attic* and *stove*,

are more likely to be learned in the first language. This effect of distributed lexical-semantic knowledge is well documented (Oller et al., 2007; Pearson et al., 1993) and demonstrates positive associations between quantity of language exposure and performance on semantic tasks (e.g., Bedore et al., 2012; Jasso et al., 2020). For example, Peña et al. demonstrated with a large sample of school-age Spanish–English bilinguals that greater Spanish language exposure was associated with higher scores on a standardized Spanish semantics subtest, whereas children with greater English exposure achieved higher scores in English (Peña, Bedore, Shivabasappa, & Niu, 2020). AoA, or the age at which children first began learning their second language, also influences lexical-semantic development. Bedore et al. (2016) found that first- and third-grade Spanish–English bilinguals who began learning English at a younger age (lower AoA) and those who had greater English language exposure demonstrated higher scores on an English semantics test.

Language exposure is dynamic, varying over time as a function of sociocultural influences. In the United States, it is common for English to be the predominant language used for education and clinical services, as the majority of teachers and clinicians are monolingual English speakers. This increased English input often results in a shift in dominance to English over time, but this shift is moderated by task type and mode of elicitation (Kohnert et al., 1999; Oppenheim et al., 2020). Changes in language exposure directly influence the types of language structures a child is exposed to, potentially posing a challenge during language evaluation, as children may have advanced lexical-semantic knowledge in one language while simultaneously having more limited knowledge in the other. This unbalanced lexical-semantic knowledge between languages potentially leads to greater difficulty with comprehension and expression in one language compared with the other, possibly resulting in difficulty successfully completing semantic tasks (see Poulisse, 1997, for a discussion of vocabulary learning in bilinguals).

Development of robust lexical-semantic knowledge requires sufficient language exposure, as more opportunities to hear and use a word across a variety of contexts allows children to build more connections among lexical items on the basis of their semantic characteristics. Having more numerous connections increases robustness, allowing a lexical item increased potential for activation and retrieval for production. Less robust lexical items—those that have fewer links between the lexical item and its corresponding semantic properties—are slower and less accurately produced by adults (Gollan et al., 2011) and children (McMillen et al., 2020). Children who have relatively limited exposure to a language may not have the opportunity to learn the words needed for the task (Hoff et al., 2012; Pearson et al., 1997; Thordardottir, 2011), inherently increasing the relative difficulty of semantic task items. As such, children must have sufficient exposure to words across a variety of contexts in each language to be able to support robust lexical-semantic development, potentially resulting in relatively increased ease for accessing and retrieving lexical-semantic information during testing and decreasing the overall semantic task difficulty. While the impact of language exposure on semantic task difficulty has not been previously explored, language exposure has been shown to impact the difficulty of expressive vocabulary for bilingual children with and without DLD (McMillen et al., 2022).

Cognitive-Linguistic Processes Implicated in Semantic Tasks

As children age, their cognitive-linguistic system matures as a function of engaging in new learning opportunities through educational and daily living experiences. This is reflected in children's vocabulary growth over time, as indexed by standardized tests and language samples (Hoff & Ribot, 2017; Rojas & Iglesias, 2013). In the classroom, school-age children's learning focuses on mathematics and language arts, developing critical thinking skills, and gaining information about the world around them (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010); learning information in this context inherently includes acquiring the language forms and content associated with these concepts. As children advance through elementary school, they encounter progressively more complex syntactic structures in educational curricula. Curran (2020) evaluated textbooks and teachers' scripts in a science curriculum for first-, third-, and fifth-grade students. She found that complex syntactic structures were included in both the textbooks and teachers' scripts, and the frequency of complex sentence use significantly increased from first to third grade. In fact, almost 30% of sentences within the textbooks contained complex syntax. Additionally, complex syntactic structures often include advanced vocabulary words, including academic vocabulary and words containing morphological affixes (Beck et al., 2013). The interdependency of developing complex language structures challenges children's cognitive-language systems to meet the demands of their educational environment.

As the language system continues to become more complex, the cognitive system must also mature to allow children to perform progressively more demanding tasks. Increases in cognitive ability provide the resources necessary for children to engage in language-based tasks, including more complex types of semantic tasks; as such, older children will have greater resources available to

complete complex language tasks compared with younger children. Standardized assessments are constructed to evaluate children's language ability across a range of time, encompassing developmental windows throughout childhood. As children age, they are able to rely on their increased language knowledge and more mature cognitive skills to complete semantic tasks, requiring fewer cognitive resources and increasing the ease children experience during task completion. In the context of language-based tasks, including semantic tasks, a child's unique profile of strengths and weaknesses across skills and abilities, as well as their relative cognitive maturation, may influence the overall difficulty of each task type.

The cognitive demand of semantic tasks differentially taxes the cognitive-linguistic system. For example, a task with reduced cognitive load, such as a functions task, often includes high frequency words corresponding to real objects that children would likely have experience with, and elicits perceptual and action features, which are learned early in childhood (see Bloom, 2002, for information on word learning). In the stimulus question, "What do scissors do?" children must recognize the target object and its corresponding action-based features, then subsequently describe the associated verbs ("scissors *cut*"). Because these lexical-semantic concepts are established at a young age, school-age children have developed a robust representation of the words and their meanings, allowing for greater ease of processing the requested information and producing an accurate response. In contrast, a relatively more cognitively demanding task, such as similarities and differences, requires abstract thinking and advanced language skill. In this type of task, children are asked to look at an array of two to four pictured items and describe how they are alike and/or different. For example, the question, "How are a pencil, a marker, and chalk alike?" would require children to respond that these items all *write*. This requires activation of each item, holding the lexical-semantic information in working memory, systematically evaluating features for each item, and determining which feature is shared across all three items. This is a more taxing task to the cognitive-linguistic system, resulting in increased difficulty for children with weaknesses in other factors, including those who have DLD, relatively immature cognitive-linguistic systems, and/or sparse lexical-semantic representations.

Task Difficulty and Semantic Task Types

The level of difficulty for a test item can be calculated based on a child's accuracy, as their performance reflects the culmination of intersecting sociocultural and developmental factors. *Item difficulty* is a psychometric property that measures the ease of a test item. In classical test theory, item difficulty is estimated by calculating the proportion of individuals who responded correctly to an item, such that higher scores (closer to 1.0) indicate an easier item whereas lower scores (closer to 0.0) indicate a relatively more difficult item (Allen & Yen, 1979). A high proportion correct on an item signifies greater ease with success on that item, whereas a lower proportion correct on an item demonstrates greater difficulty. Items on tests can be grouped together based on the underlying information targeted within each item, specifically by task type. These groups of items within each task type measure a child's skill and knowledge in a specific area and can be used to determine patterns of strength or weakness.

Not all semantic task types have equal task difficulty. TD Spanish–English bilingual children show cross-linguistic differences in performance on semantic task types. In a study by Peña et al. (2003), the easiest tasks in Spanish for children were expressive and receptive functions, which require children to identify or describe objects by their function, and characteristic properties, which require children to identify and/or describe perceptual features (e.g., What shape is the box?). The most difficult tasks in Spanish included expressive linguistic concepts and expressive associations. In contrast, the easiest tasks in English were receptive similarities and differences, as well as expressive functions. The most difficult tasks in English included expressive characteristic properties, expressive linguistic concepts, and expressive associations. These findings demonstrate that the language of testing and the response mode (i.e., receptive or expressive) influence children's performance on semantic task types.

This Study

When learning a second language, children begin with lexical-semantic knowledge in their first language, and they can use this knowledge to scaffold language learning. However, the extent to which children can successfully accomplish this depends on their language ability, language experience, and the semantic task types, as tasks inherently tap different aspects of lexical-semantic knowledge, requiring children to differentially leverage cognitive-linguistic resources. In this study, we focus on evaluating changes in English semantic task difficulty over time because English is the language that bilinguals in the United States are in the process of acquiring and—in monolingual English public school systems—have few supports to do so. For children who are referred for an evaluation due to difficulty with learning language, there is a small proportion of clinicians available who can assess bilingual children in their home language (American Speech-Language-Hearing Association, 2022). Given this limitation, it is important to understand how children

perform on different semantic task types in English, as clarifying this aspect of developmental profiles will help clinicians make more accurate judgments about children's language learning.

This study uses the English Semantics subtest from the BESA-ME Field Test Version (Peña et al., 2018). While tasks tapping semantic depth are important for informing children's language profile, it has yet to be determined how the difficulty of these semantic tasks change over time and between language ability groups for children along a continuum of bilingual experiences. Given the dynamic nature of language exposure over time and its influence on lexical-semantic development, the difficulty of each semantic task type would also be expected to change with natural advances in cognitive maturation. The developmental trajectory in each language is directly proportional to the quantity of language exposure the child has to each language. This is particularly interesting to examine in English as children gain progressively more exposure to English across a wider variety of contexts as they age. Studying how children's performance across semantic tasks in English changes over time informs clinical decision making and the development of psychometrically valid assessment measures for culturally and linguistically diverse populations.

Research Questions

The research questions for this study are as follows:

1. What are the rates of growth in English task difficulty across semantic task types for children with DLD compared with their TD peers?
2. What are the predictive roles of grade, language exposure, language ability status, and English AoA on task difficulty across semantic task types for children with DLD compared with their TD peers?

First, we predicted that semantic task difficulty would decrease over time, as reflected in increasing proportion correct, for both language ability groups as a result of overall development and increased English language exposure (Bedore et al., 2016). However, in line with previous research comparing language ability groups (e.g., Jasso et al., 2020), we also predicted that children with DLD would experience greater difficulty on each of the semantic tasks compared with TD peers. Here, we extend this work by examining individual variations in semantic task difficulty through growth curve modeling. As DLD inherently inhibits lexical-semantic development, we expect that growth rates for more challenging semantic task types would show limited or no acceleration over time compared with faster growth rates for relatively easier semantic task types. In contrast, we expected that TD children would show continuous acceleration over time for all semantic task types.

Second, we hypothesized that factors important for language learning, including grade, language exposure, language ability status, and English AoA, would predict semantic task difficulty. While we believed clear differences across ability groups would emerge for these semantics task types (e.g., Peña et al., 2015), children should demonstrate gains across grades due to cognitive maturation and language growth inherent to development. However, children with lower English language exposure and higher AoA may experience greater difficulty on all semantic task types, regardless of language ability status (Bedore et al., 2016). Clinically, relatively low performance on lexical-semantic tasks would cause TD children to superficially appear similar to their peers with DLD, as there are some errors that both TD and children with DLD produce when they are experiencing increased demands on vocabulary tasks (McMillen et al., 2020). Because of this, it is important to understand how semantic task difficulty varies as a function of language ability, language exposure, AoA, and grade for bilingual children.

Method

Participants

Participants included 232 English–Spanish Latinx bilingual children between the ages of 7;2 (years;months) and 12;2 who attended public schools in the central Texas region of the United States. This sample was derived from a 4-year longitudinal parent study evaluating typical and atypical bilingual language development during elementary school (Peña, 2010), which included a total of 323 English–Spanish Latinx bilingual children. The parent study comprised three entry cohorts, with children entering the study at kindergarten, second grade, or fourth grade and remaining for up to four consecutive annual time points. Given the analyses adopted, inclusion for the current retrospective study required complete data on the following measures: (a) language ability status, (b) current English language exposure, (c) age of English acquisition, and (d) semantic task performance as indexed by the Semantics subtest of the BESA-ME (Peña, Bedore, Gutiérrez-Clellen, et al., 2016). We include the data at time points corresponding to second through fifth grade, as these were the grades during which the BESA-ME Semantics subtest was administered. This resulted in 503 observations from 232 distinct participants, with children contributing an average of 2.2 (range: 1–4) annual observations. Count and descriptive data for these participants are shown in Table 1, and children's semantic difficulty performance is shown in Figure 1. The resulting sample

Table 1. Participant counts and descriptives by grade.

Analysis grade	Entry grade			Obs	Descriptives					
	K	Second	Fourth		Age in years *M (SD)*	% female	% DLD	MED *M (SD)*	AoA *M (SD)*	CEE *M (SD)*
Second	54	112	—	166	7.9 (0.3)	51%	18%	2.8 (1.7)	2.6 (1.7)	46 (20)
Third	34	100	—	134	9.0 (0.4)	51%	21%	2.8 (1.7)	2.6 (1.7)	58 (21)
Fourth	—	49	66	115	9.9 (0.4)	47%	12%	2.9 (1.7)	2.4 (2.0)	61 (19)
Fifth	—	29	59	88	10.9 (0.5)	48%	10%	2.7 (1.6)	2.4 (2.0)	69 (18)
	Total distinct *N* = 232			Total Obs = 503	9.2 (1.2)	50%	16%	2.8 (1.7)	2.5 (1.9)	57 (22)

Note. Em dashes indicate that there are no data for the corresponding grade. Obs = observations; *M* = mean; *SD* = standard deviation; DLD = developmental language disorder; MED = maternal education per Hollingshead (1975), where 1 = ≤ seventh grade, 2 = eighth to ninth grade, 3 = tenth to eleventh grade, 4 = high school graduate, 5 = partial college, 6 = college education, and 7 = graduate degree; AoA = age of first English acquisition in years; CEE = current percent English exposure.

included 197 TD children and 35 children with DLD at study entry. Table 2 displays descriptive data by language ability group. The groups did not differ significantly ($p > .05$) on age, $t(54.90) = -1.78$, $p = .081$, mother's education level, $t(50.75) = -1.11$, $p = .274$, or biological sex, $\chi^2(1) = 3.54$, $p = .060$. TD and DLD children differed significantly on mean English language exposure, $t(49.03) = -2.21$, $p = .032$, and English AoA, $t(46.79) = 2.12$, $p = .039$, with TD children demonstrating 8% higher mean language exposure and 0.7 year earlier AoA on average. We controlled for these differences across all statistical analyses.

Language Ability Status

Language ability status was determined using multiple indicators to inform clinical diagnosis; this framework is consistent with the concept of utilizing a converging evidence approach to identify DLD in bilingual children (Castilla-Earls et al., 2020). Children were classified as having DLD during confirmatory testing of the parent study (Peña, 2010). Criteria for a diagnosis of DLD required four of the five following criteria: (a) a parent and/or teacher rating of ≤ 4.2 (max = 5) in Spanish and English on the Instrument to Assess Language Knowledge (ITALK; Peña et al., 2018), (b) a score lower than 1 *SD* below the normative mean in Spanish and English on the Morphosyntax subtest of the BESA-ME Field Test Version, (c) a score lower than 1 *SD* below the normative mean in Spanish and English on the Semantics subtest of the BESA-ME Field Test Version, (d) a composite score that was lower than 1 *SD* below the mean in Spanish and English on the Bilingual English–Spanish Oral Screener (BESOS; Peña et al., 2008), and (e) a score that fell lower than 1 *SD* below the normative mean in Spanish and English on the Test of Narrative Language (TNL: Gillam & Pearson, 2004; TNL-Spanish: Gillam et al., 2006). A better-language approach was used, such that, for a particular domain or language measure, only scores in the better language were considered. Children were classified as TD if their scores fell within the average or above average range on four or more of the diagnostic measures used in this study.

Figure 1. Observed mean semantic difficulty by task, grade, and language ability. DLD = developmental language disorder; TD = typically developing.

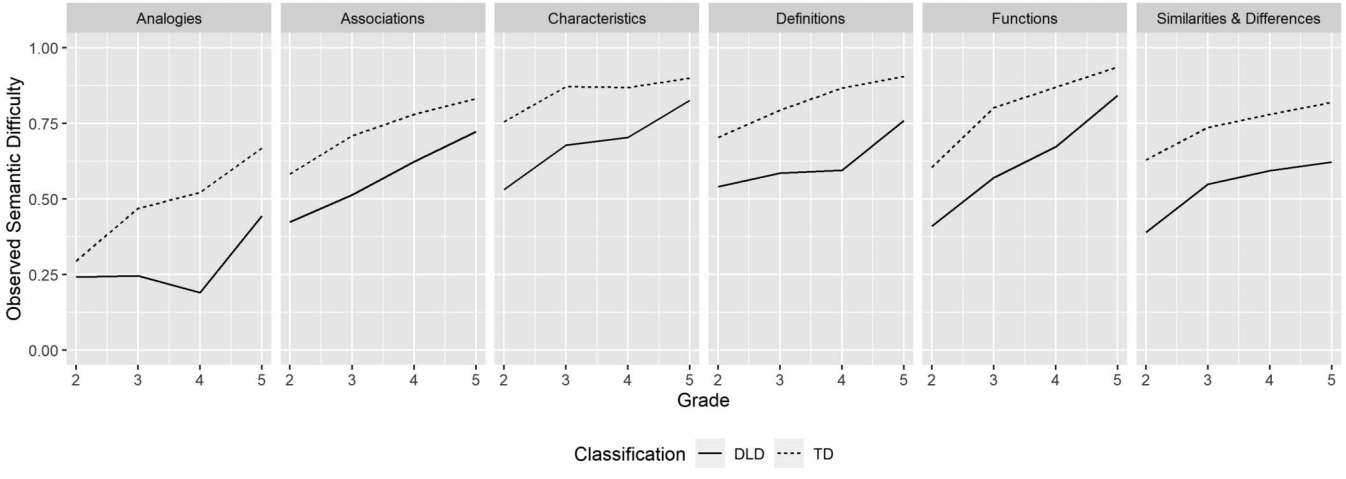


Table 2. Participant descriptives by language ability group (N = 232).

Variable	TD	DLD
n	197	35
Age in years*	8.5 (1.0)	8.2 (.8)
% female	52%	34%
MED	2.9 (1.7)	2.5 (1.5)
AoA	2.4 (1.8)	3.1 (1.9)
MEE	58 (19)	50 (19)
BESA/BESA-ME SS*		
English[a]	98 (14)	73 (14)
Spanish[b]	92 (17)	72 (14)
TNL*		
English[c]	81 (18)	64 (12)
Spanish[d]	100 (15)	80 (12)
ITALK*		
English	3.5 (1.1)	2.5 (.9)
Spanish	4.0 (1.2)	3.8 (1.1)

Note. Values outside of parentheses represent means, whereas values inside parentheses represent standard deviations. TD = typically developing; DLD = developmental language disorder; MED = maternal education per Hollingshead (1975), where 1 = ≤ seventh grade, 2 = eighth to ninth grade, 3 = tenth to eleventh grade, 4 = high school graduate, 5 = partial college, 6 = college education, and 7 = graduate degree; AoA = age of first English acquisition in years; MEE = mean percent English exposure; Bilingual English–Spanish Assessment—Middle Extension Field Test Version; SS = semantics standard score; TNL = Test of Narrative Language; ITALK = parent/teacher language report averages or (in cases where only one was reported) either parent or teacher.

*At study entry.
[a]6 missing. [b]25 missing. [c]6 missing. [d]23 missing.

Materials

Language Exposure

The Bilingual Input–Output Survey (BIOS; Peña et al., 2018) was used to quantify bilingual language exposure upon entry into the study and in every year afterward. Caregivers and teachers independently reported children's hour-by-hour input and output of English and Spanish on a typical weekday and weekend day. Raw hour totals were used to generate weighted averages, which was converted into a percentage of exposure relative to the other language.

BESA-ME Semantics Subtest

The BESA-ME Semantics subtest is composed of seven semantic task types: analogies, associations, categories, characteristics, definitions, functions, and similarities and differences. In this study, the categories task was excluded due to the limited number of items (i.e., two items) included on the BESA-ME. Each task type contains both receptive and expressive item types; for the purpose of this study, responses across language modalities (i.e., receptive and expressive) were not disaggregated due to factor structure observed for a younger group of children (BESA; Peña et al., 2018).

Procedure

Data were collected in two phases of the longitudinal parent study (Peña, 2010): a screening phase and a confirmatory phase. During the screening phase, trained Spanish–English bilingual examiners administered the BESOS to children individually in approximately 30-min sessions in quiet locations within their schools. The confirmatory phase began 1 year later and lasted up to four consecutive years. During this phase, children completed a battery of tests evaluating language ability and cognition administered by trained Spanish–English bilingual examiners. Confirmatory phase testing occurred over three to six sessions lasting up to 60 min in quiet locations within children's schools. Parents and teachers completed the ITALK and BIOS during in person or phone interviews with trained examiners. Language exposure, AoA, and language ability ratings were calculated from parent and teacher responses on these measures (see Bedore et al., 2018, and Peña, Bedore, Lugo-Neris, & Albudoor, 2020, for additional details).

In this study, we conducted a secondary analysis of longitudinal data to investigate growth rates in semantic task difficulty from second through fifth grade. Child responses were obtained from the English semantics subtest of the BESA-ME Field Test Version (Peña, Bedore, Gutiérrez-Clellen, et al., 2016), which was administered during the confirmatory testing phase in second through fifth grades. The outcome of interest for the present study was semantic task difficulty. Specifically, a difficulty index (i.e., proportion of correct answers) was calculated for each semantic task type, yielding grade-specific difficulty indices. Importantly, a higher difficulty index is the result of higher accuracy, indicating an easy semantic task; the opposite is also true, where a hard task would have a low difficulty index due to low accuracy on the semantic task.

Analysis Method

To estimate the effects of language ability, grade, AoA, and language exposure on semantic task difficulty, multilevel regression modeling was conducted. This method enables the estimation of a longitudinal trajectory using cross-sequential data by accounting for the clustering associated with repeated measurements (Hoffmann, 2016). One multilevel regression model was estimated, with children's semantic task difficulty serving as the outcome variable. The fixed effects were (a) Task, (b) Language Ability Status (levels: TD = 0, DLD = 1), (b) Grade, (c) time-varying Current English Language Exposure, and (d) time-invariant AoA. To aid interpretation of model estimates, grade was centered at second grade (the earliest grade in the sample), current language exposure was centered at .5 (representing an equal amount of

Table 3. Multilevel regression model predicting semantic difficulty.

Fixed effects		*F*	*p*
Task		116.75	**< .001**
Language ability status[a]		76.46	**< .001**
Grade		272.31	**< .001**
AoA		2.75	.099
CEE		8.45	**.004**
Task × Language Ability Status[a]		0.53	.754
Task × Grade		10.78	**< .001**
Task × AoA		1.89	.092
Task × CEE		4.68	**< .001**
Random Effects			
σ^2	.0340		
τ_{00}	.0177		
τ_{01}	.0007		
ρ_{01}	−.9900		
N	232		
Observations	503		

Note. Bolded *p* values indicate statistical significance.

[a]typically developing = 0, developmental language disorder = 1; AoA = age of first English acquisition (mean centered at 2.31 years); CEE = current percent English exposure (centered at .5).

English and Spanish exposure), and English AoA was centered at its grand mean (2.31 years). The random intercept, random slope, and the intercept by slope correlation were estimated to account for correlations between repeated measurements. For all models, restricted maximum likelihood estimation was used, and *p* values were approximated using Satterthwaite's method (Kuznetsova et al., 2017). These analyses were completed using the *mixed* function of the *afex* package, version 1.1–1 (Singmann et al., 2022), which employed the R statistical language, version 4.2.1 (R Core Team, 2022). Finally, to determine estimated effects of the higher order interactions, estimated marginal means and estimated mean trend analyses were conducted using the *emmeans* package, version 1.8.1–1 (Lenth, 2022) in R (R Core Team, 2022).

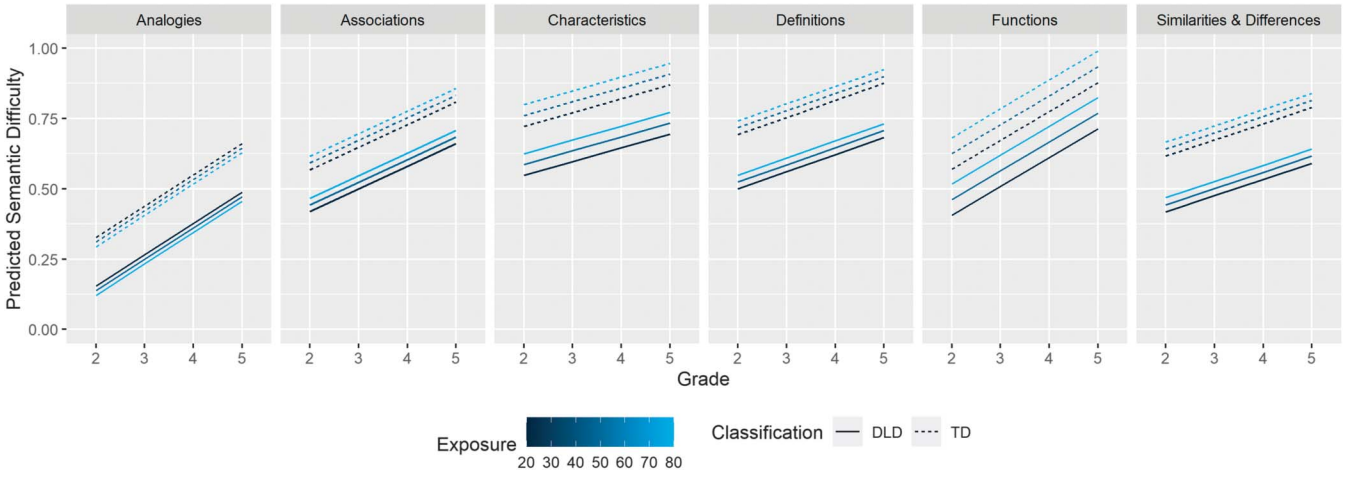
Results

Table 3 and Figure 2 present the results of the multilevel regression model predicting semantic task difficulty. There was a significant main effect of Task (F = 116.75, p < .001), indicating that there were substantial differences in semantic task difficulty depending on the semantic task type. Estimated marginal means on semantic task type demonstrated the following difficulty order, from least to most difficult: (a) characteristics (M = .742), (b) definitions (M = .697), (c) functions (M = .680), (d) associations (M = .620), (e) similarities and differences (M = .618), and (f) analogies (M = .363).

Regarding change over time for semantic task difficulty, the main effect of Grade was significant (F = 272.31, p < .001) and there was a significant Task × Grade interaction (F = 10.78, p < .001), indicating that there was change over time in semantic task difficulty and that the rate of growth substantially varied between semantic task types. Specifically, estimated mean trend analyses demonstrated the following estimated slopes per grade level from slowest to fastest growing: (a) characteristics (b = .047), (b) similarities and differences (b = .063), (c) definitions (b = .065), (d) associations (b = .084), (e) functions (b = .103), and (f) analogies (b = .117).

For children's language ability status, the main effect of language ability was significant (F = 76.46, p < .001) with children with DLD performing lower (M = .533) than their TD peers (M = .707) on average. This indicates that children with DLD experienced greater

Figure 2. Model predicted semantic difficulty by task, grade, language ability classification, and current English exposure. DLD = developmental language disorder; TD = typically developing.



difficulty on each semantic task type compared with their TD peers. However, the Task × Language Ability interaction was not significant, indicating that language ability status did not differentially influence semantic task difficulty on the various semantic task types.

Finally, while children's English AoA did not influence semantic task difficulty ($F = 2.75$, $p = .099$), there was a significant main effect of current English Language Exposure ($F = 8.45$, $p = .004$) and a significant Task × Language Exposure interaction ($F = 4.68$, $p < .001$). Estimated mean trend analyses demonstrated that for each unit increase in children's current English language exposure, semantic task difficulty on the following tasks were estimated to change by the following slope values, from least to most change: (a) associations ($b = .072$), (b) definitions ($b = .087$), (c) analogies ($b = -.097$), (d) similarities and differences ($b = .099$), (e) characteristics ($b = .131$), and (f) functions ($b = .202$).

Discussion

The purpose of this study was to evaluate developmental and exposure-related changes in English semantic task difficulty for school-aged bilingual children with and without DLD. We found that while English language exposure was important for building lexical-semantic representations, ability drives difficulty level on semantic tasks. Additionally, while all children made gains on each semantic task type over time, the growth rate of semantic task difficulty was not equal across all task types.

Language Ability and Semantic Task Difficulty

Longitudinally, semantic task difficulty decreased for each group, as indicated by an increase in the proportion correct, meaning the tasks became easier for all children over time. Descriptively, the patterns of hardest and easiest task types were also similar across language ability groups, where analogies emerged as the most difficult task type and characteristics was the easiest task type across all grades. These patterns of relative difficulty level may be due to the cognitive-linguistic demands inherent to task types, where accurately responding to items targeting characteristics relies on language-based knowledge of perceptual features and uses fewer cognitive resources for successful task completion. In contrast, the analogies task requires not only linguistic knowledge but also places a significant demand on the executive function system, including working memory, for successful completion (see Goswami, 2013, for a discussion). This increased pressure on the cognitive-linguistic system makes this type of task difficult for all children. For example, the analogy "*mouse* is to *small* as *elephant* is to _____" requires the child to understand the nouns *elephant* and *mouse* and the basic concepts of *small* and *big/large*, as well as extrapolate the intended comparison of these concepts for production. We expect that this task would be difficult for all young children, regardless of the number of languages known. In fact, tailored item analyses show that analogies and definitions items were poor discriminators of DLD in a larger set of these same children (Jasso et al., 2020). This study sheds light on these findings by demonstrating that poor classification accuracy in the Jasso et al. study was due to the similar high difficulty levels across the groups for the analogies task.

Differences by ability group did emerge for children with DLD, where they experienced greater difficulty overall on all semantic task types in comparison to their TD peers—even after accounting for language exposure and AoA. This finding is consistent with previous literature demonstrating that monolingual and bilingual children with DLD have impoverished semantic knowledge in comparison to their TD peers (McGregor et al., 2002; Sheng et al., 2012). While monolingual children with DLD may have breadth of lexical knowledge (i.e., number of vocabulary words) that is commensurate or in the same range as their TD peers, they often have slightly lower scores on single-word vocabulary tasks than their TD peers (Gray et al., 1999); these limitations in their vocabulary knowledge persist through childhood and adolescence (McGregor et al., 2013). The BESA-ME Field Test Version (Peña, Bedore, Gutiérrez-Clellen, et al., 2016) was used in our study because it goes beyond lexical breadth, tapping into depth of semantic knowledge. Differences between groups of TD and DLD children are evident when evaluating semantic depth, where children with DLD experience greater difficulty relative to their TD peers for each semantic task type.

The Effects of Grade, Language Exposure, and AoA on Semantic Task Difficulty

In our study, while grade may represent the natural progression of cognitive and linguistic development inherent as children age, it also reflects the knowledge children acquire in school; thus, while the relative importance of age verses school experience cannot be teased apart, it is likely the combination of these two factors that contributes to the importance of this variable in this study. That is, as children age, they have more experiences with words in a variety of contexts, while school has the effect of creating relatively similar experiences for children across educational settings, as the same concepts and material are taught in accordance with required educational competencies. The natural progression represented by grade illustrates the gradual development of robust semantic knowledge within the linguistic system. However, growth rate differs by task

type, where characteristics demonstrated the slowest growth and analogies demonstrated the fastest growth.

These patterns have implications for development, where children make significant and rapid gains in language acquisition with regard to the cognitive process of comparing and describing the relationship between two concepts (i.e., analogies task). This increase is coupled with maturation of the cognitive system in school-age children, allowing them to engage in more advanced language-based tasks that inherently require abstract reasoning. In comparison, slower growth is demonstrated for describing the perceptual features of objects (i.e., characteristics task); this skill requires fewer cognitive resources, allowing children to gain the cognitive-linguistic skills necessary for success on this task type earlier in development.

While language exposure was a significant predictor of semantic task difficulty for all semantic task types, AoA did not emerge as a significant factor. While previous research found that both language experience factors—language exposure and AoA—were significantly related to children's standard scores on a semantics test (Bedore et al., 2016), our results with an older group of children demonstrate that children's current language exposure is more important for predicting the relative difficulty level of semantic task types than AoA. However, this effect of language exposure was not equal across all semantic task types, demonstrating the most significant impact on associations, definitions, and analogies. While language exposure effectively increases children's opportunities to construct the language knowledge needed to successfully complete semantic tasks, it is likely that other factors also play an important role in influencing the relative ease of children's performance. Semantic task difficulty may be impacted by individual differences in phonological working memory capacity, the lexical-semantic characteristics of the words used in these tasks (McMillen et al., 2022), or a combination of these factors. Given the reciprocal relationship between working memory and language (e.g., Archibald, 2018), children who have better working memory may experience decreased difficulty on each semantic task type. Additional research is needed on the role of cognition for successful completion of semantic tasks.

In addition to internal factors unique to each child's profile, the response mode for each item within the semantic task types could alter the relative difficulty. For example, in this study, the mode for the items used in the functions task included five expressive and two receptive items. Peña et al. (2003) found that for younger TD bilingual children with a range of English language exposure, receptive items were relatively more difficult than expressive items on the functions task in English. For this task type, it may be that including more receptive items would result in an interaction between language ability and language exposure for children with greater Spanish exposure and those with DLD, as we would expect that children who fall into one or both of these groups to experience greater difficulty than their peers who have more Spanish exposure or who are TD. Future investigations are required to tease apart these nuances.

Limitations and Future Directions

For the current investigation, we evaluated semantic task difficulty in English using a test that is under development for older elementary school children. Receptive and expressive items were combined into semantic task types based on the area of lexical-semantic knowledge tapped in the composite items. This approach differs from previous investigations that differentiate receptive and expressive items and account for differences in semantic performance on Spanish tasks (Peña et al., 2003). Possible future directions for this work include evaluating developmental patterns of semantic task difficulty across response modes (receptive vs. expressive) and between Spanish and English for younger bilingual children. Evaluating developmental trends in semantic task difficulty across languages would clarify how the language profile changes over time for bilingual children with and without DLD.

It is also noteworthy that language input and language output were combined to create a composite language exposure measure. While language input is important for lexical-semantic development in younger bilingual children (Bohman et al., 2010), there is evidence that language use, or output, strengthens linguistic representations, increasing expressive vocabulary knowledge for young children (Ribot et al., 2018). As opportunities to expressively practice language builds lexical-semantic knowledge, we chose not to separate language input from output in this sample of older children. Future investigations should account for the differential influence of language input verses output on children's performance.

Within this longitudinal study, an average of 2.2 observations per child were included in the statistical model. Children had the opportunity to participate in this study for up to four consecutive years, depending upon the entry year; however, older children were only eligible to participate for 1 or 2 years due to aging out of the study population. Additionally, other data points for younger children were lost due to attrition over time. By using growth curve modeling, we were able to include data from all children, even when multiple observations for each child were not available within the data set. While the number of longitudinal observations may be relatively limited, this statistical approach allows for accurate predictions of the nonlinear change in difficulty for each semantic task type over time.

While we know that aspects of cognition impact language development (e.g., Gray et al., 2019; Wood et al.,

2021), we did not directly assess the relative contributions of verbal and nonverbal cognitive processes to semantic task difficulty in this study. Future investigations should evaluate the cognitive and linguistic demands of semantic task types through a developmental lens to better understand the mechanisms underpinning performance on semantic tasks.

Clinical Implications

Our study provided empirical evidence that difficulty on semantic tasks is driven by language ability and age, where increased semantic task difficulty is experienced by children with impaired language ability and/or younger children. Additionally, growth rate in semantic task difficulty varies by the task type, reflecting children's development of language knowledge and cognitive maturation. While the items used in each semantic task were created to be sensitive to differences in cultural knowledge for school-aged Spanish–English bilinguals in the United States, some tasks are more affected by individual differences in language exposure than others. Because language exposure has a significant effect on children's standardized test performance, clinicians should adjust to their assessment plans to account for individual differences in second language input and output. To accomplish this, clinicians should collaborate with caregivers and teachers to obtain objective measurements of language exposure for each child. These data, in combination with practical assessment considerations (i.e., access to interpreters, availability of high quality culturally and linguistically sensitive tests), will guide the evaluation process (see Castilla-Earls et al., 2020, for an evaluation framework). Additionally, it is critical to contextualize children's assessment results given their current exposure to each language within the home and school environments. We recommend that clinicians consider using a best language approach when evaluating results and assigning diagnoses (see Peña, Bedore, & Kester, 2016). Semantic tests for bilingual populations should incorporate a variety of items tapping a range of semantic skills to best capture language disorder among the range of language experience.

Data Availability Statement

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request (Peña, 2010).

Acknowledgments

Funding for this work was provided by the Cross-Language Outcomes of Bilingual Language Impairment Grant R01 DC010366-01A1 (PI: Peña) from the National Institute of Deafness and Other Communication Disorders and Language Evaluation and Development in Educational Research training grant from the U.S. Department of Education. We thank the children and their parents who participated in this study, as well as the research assistants in the Human Abilities in Bilingual Language Acquisition Lab for their support on this project. We also want to express gratitude to Javier Jasso for his contributions to the initial design of this study.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Brooks/Cole.

American Speech-Language-Hearing Association. (2022). *2021 demographic profile of ASHA members providing multilingual services*. https://www.asha.org

Archibald, L. M. D. (2018). The reciprocal influences of working memory and linguistic knowledge on language performance: Considerations for the assessment of children with developmental language disorder. *Language, Speech, and Hearing Services in Schools, 49*(3), 424–433. https://doi.org/10.1044/2018_LSHSS-17-0094

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Bedore, L. M., Peña, E. D., Anaya, J. B., Nieto, R., Lugo-Neris, M. J., & Baron, A. (2018). Understanding disorder within variation: Production of English grammatical forms by English language learners. *Language, Speech, and Hearing Services in Schools, 49*(2), 277–291. https://doi.org/10.1044/2017_LSHSS-17-0027

Bedore, L. M., Peña, E. D., Griffin, Z. M., & Hixon, J. G. (2016). Effects of age of English exposure, current input/output, and grade on bilingual language performance. *Journal of Child Language, 43*(3), 687–706. https://doi.org/10.1017/S0305000915000811

Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., Bohman, T. M., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism: Language and Cognition, 15*(3), 616–629. https://doi.org/10.1017/S1366728912000090

Bloom, P. (2002). *How children learn the meanings of words*. MIT Press.

Bohman, T. M., Bedore, L. M., Peña, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: Language performance in Spanish–English bilinguals. *International Journal of Bilingual Education and Bilingualism, 13*(3), 325–344. https://doi.org/10.1080/13670050903342019

Castilla-Earls, A., Bedore, L., Rojas, R., Fabiano-Smith, L., Pruitt-Lord, S., Restrepo, M. A., & Peña, E. (2020). Beyond scores: Using converging evidence to determine speech and language services eligibility for dual language learners. *American Journal of Speech-Language Pathology, 29*(3), 1116–1132. https://doi.org/10.1044/2020_AJSLP-19-00179

Curran, M. (2020). Complex sentences in an elementary science curriculum: A research note. *Language, Speech, and Hearing Services in Schools, 51*(2), 329–335. https://doi.org/10.1044/2019_LSHSS-19-00064

Gillam, R. B., & Pearson, N. A. (2004). *TNL: Test of narrative language*. Pro-Ed.

Gillam, R. B., Peña, E. D., Bedore, L. M., & Pearson, N. (2006). *Test of Narrative Language–Spanish Adaptation Experimental Version (TNL-S)*. Pro-Ed.

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology, 140*(2), 186–209. https://doi.org/10.1037/a0022256

Goswami, U. (2013). *Analogical reasoning in children*. Psychology Press. https://doi.org/10.4324/9781315804729

Gray, S., Fox, A. B., Green, S., Alt, M., Hogan, T. P., Petscher, Y., & Cowan, N. (2019). Working memory profiles of children with dyslexia, developmental language disorder, or both. *Journal of Speech, Language, and Hearing Research, 62*(6), 1839–1858. https://doi.org/10.1044/2019_JSLHR-L-18-0148

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools, 30*(2), 196–206. https://doi.org/10.1044/0161-1461.3002.196

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language, 63*(3), 259–273. https://doi.org/10.1016/j.jml.2010.06.002

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language, 39*(1), 1–27. https://doi.org/10.1017/S0305000910000759

Hoff, E., & Ribot, K. M. (2017). Language growth in English monolingual and Spanish-English bilingual children from 2.5 to 5 years. *The Journal of Pediatrics, 190,* 241–245.e1. https://doi.org/10.1016/j.jpeds.2017.06.071

Hoffmann, J. P. (2016). *Regression models for categorical, count, and related variables: An applied approach*. University of California Press. https://doi.org/10.1525/9780520965492

Hollingshead, A. (1975). *Four factor index of social status*. Yale University Department of Psychology.

Jasso, J., McMillen, S., Anaya, J. B., Bedore, L. M., & Peña, E. D. (2020). The utility of an English semantics measure for identifying developmental language disorder in Spanish-English bilinguals. *American Journal of Speech-Language Pathology, 29*(2), 776–788. https://doi.org/10.1044/2020_AJSLP-19-00202

Kohnert, K. J., Bates, E., & Hernandez, A. E. (1999). Balancing bilinguals. *Journal of Speech, Language, and Hearing Research, 42*(6), 1400–1413. https://doi.org/10.1044/jslhr.4206.1400

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lenth, R. (2022). *emmeans: Estimated marginal means, aka least-squares means*. https://CRAN.R-project.org/package=emmeans

McGregor, K. K., Newman, R. M., Reilly, R. M., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 45*(5), 998–1014. https://doi.org/10.1044/1092-4388(2002/081)

McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders, 48*(3), 307–319. https://doi.org/10.1111/1460-6984.12008

McMillen, S., Anaya, J. B., Peña, E. D., Bedore, L. M., & Barquin, E. (2022). That's hard! Item difficulty and word characteristics for bilinguals with and without developmental language disorder. *International Journal of Bilingual Education and Bilingualism, 25*(5), 1838–1856. https://doi.org/10.1080/13670050.2020.1832039

McMillen, S., Griffin, Z. M., Peña, E. D., Bedore, L. M., & Oppenheim, G. M. (2020). "Did I say cherry?" error patterns on a blocked cyclic naming task for bilingual children with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research, 63*(4), 1148–1164. https://doi.org/10.1044/2019_JSLHR-19-00041

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards*.

Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics, 28*(2), 191–230. https://doi.org/10.1017/S0142716407070117

Oppenheim, G. M., Griffin, Z., Peña, E. D., & Bedore, L. M. (2020). Longitudinal evidence for simultaneous bilingual language development with shifting language dominance, and how to explain it. *Language Learning, 70*(S2), 20–44. https://doi.org/10.1111/lang.12398

Pearson, B. Z., Fernández, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics, 18*(1), 41–58. https://doi.org/10.1017/S0142716400009863

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*(1), 93–120. https://doi.org/10.1111/j.1467-1770.1993.tb00174.x

Peña, E., Bedore, L. M., & Rappazzo, C. (2003). Comparison of Spanish, English, and bilingual children's performance across semantic tasks. *Language, Speech, and Hearing Services in Schools, 34*(1), 5–16. https://doi.org/10.1044/0161-1461(2003/001)

Peña, E. D. (2010). *Cross-language outcomes of bilingual experience in typical and atypical development* (project no. R01DC010366) [Grant]. National Institute of Deafness and other Communication Disorders. https://projectreporter.nih.gov/project_info_description.cfm?aid=8662743

Peña, E. D., Bedore, L. M., Gutiérrez-Clellen, V. F., Iglesias, A., & Goldstein, B. A. (2016). *Bilingual English Spanish Assessment–Middle Extension Field Test Version (BESA-ME)*.

Peña, E. D., Bedore, L. M., Iglesias, A., Gutiérrez-Clellen, V. F., & Goldstein, B. A. (2008). *Bilingual English Spanish Oral Screener–Experimental Version (BESOS)* [Unpublished instrument].

Peña, E. D., Bedore, L. M., & Kester, E. S. (2015). Discriminant accuracy of a semantics measure with Latino English-speaking, Spanish-speaking, and English–Spanish bilingual children. *Journal of Communication Disorders, 53,* 30–41. https://doi.org/10.1016/j.jcomdis.2014.11.001

Peña, E. D., Bedore, L. M., & Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders, 51*(2), 192–202. https://doi.org/10.1111/1460-6984.12199

Peña, E. D., Bedore, L. M., Lugo-Neris, M. J., & Albudoor, N. (2020). Identifying developmental language disorder in school age bilinguals: Semantics, grammar, and narratives. *Language Assessment Quarterly, 17*(5), 541–558. https://doi.org/10.1080/15434303.2020.1827258

Peña, E. D., Bedore, L. M., Shivabasappa, P., & Niu, L. (2020). Effects of divided input on bilingual children with language impairment. *International Journal of Bilingualism, 24*(1), 62–78. https://doi.org/10.1177/1367006918768367

Peña, E. D., Gutiérrez-Clellen, V., Iglesias, A., Goldstein, B., & Bedore, L. M. (2018). *BESA: Bilingual English–Spanish Assessment Manual*. Brookes.

Poulisse, N. (1997). Language production in bilinguals. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 201–224). Erlbaum.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org

Ribot, K. M., Hoff, E., & Burridge, A. (2018). Language use contributes to expressive language growth: Evidence from bilingual children. *Child Development, 89*(3), 929–940. https://doi.org/10.1111/cdev.12770

Rice, M. L. (1990). Preschoolers' QUIL: Quick incidental learning of words. In G. Conti-Ramsden & C. E. Snow (Eds.), *Children's language* (Vol. 7, pp. 171–195). Psychology Press.

Rojas, R., & Iglesias, A. (2013). The language growth of Spanish-speaking English language learners. *Child Development, 84*(2), 630–646. https://doi.org/10.1111/j.1467-8624.2012.01871.x

Sheng, L., Peña, E. D., Bedore, L. M., & Fiestas, C. E. (2012). Semantic deficits in Spanish–English bilingual children with language impairment. *Journal of Speech, Language, and Hearing Research, 55*(1), 1–15. https://doi.org/10.1044/1092-4388(2011/10-0254)

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. (2022). *afex: Analysis of factorial experiments*. https://CRAN.R-project.org/package=afex

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism, 15*(4), 426–445. https://doi.org/10.1177/1367006911403202

Wood, T., Pratt, A. S., Durant, K., McMillen, S., Peña, E. D., & Bedore, L. M. (2021). Contribution of nonverbal cognitive skills on bilingual children's grammatical performance: Influence of exposure, task type, and language of assessment. *Language, 6*(1), 36. https://doi.org/10.3390/languages6010036