# Did I Say *Cherry*? Error Patterns on a Blocked Cyclic Naming Task for Bilingual Children with and without DLD

Stephanie McMillen[*] Zenzi M. Griffin[†] Elizabeth D. Peña[‡]
Lisa M. Bedore[§] Gary M. Oppenheim[¶‖]

To appear in *Journal of Speech, Language, and Hearing Research*
(accepted December, 2019)

**Abstract**

**Purpose:** Using a blocked cyclic picture naming task, we compared accuracy and error patterns across languages for Spanish-English bilingual children with and without developmental language disorder (DLD).

**Methods:** Pictured stimuli were manipulated for semantic similarity across two (Same and Mixed) category contexts. Children's productions were scored offline for accuracy, error frequency, and error type.

**Results:** TD children were more accurate and produced fewer errors than their peers with DLD; however, this was moderated by task language and semantic context. Children were generally more accurate when naming pictures in English compared to Spanish and in the Mixed-category context compared to the Same-category context. Analyses of error types further showed that children with less English language exposure specifically

[*]Department of Communication Sciences & Disorders, Syracuse University, 621 Skytop Rd, Suite 1200, Syracuse, NY 13244
[†]Department of Psychology University of Texas at Austin
[‡]School of Education University of California, Irvine
[§]Department of Communication Sciences and Disorders, Temple University
[¶]School of Psychology, Bangor University
[‖]Miles Dyslexia Centre, Bangor University

Conflict of interest: Drs. Peña and Bedore are authors of the *Bilingual English-Spanish Assessment* (BESA) and receive royalties from its sale.
Funding Statement: Funding for this work was provided by the *Cross-language outcomes of bilingual language impairment* (R01 DC010366-01A1) and *Language, Communication and the Brain* (T32DC000041-20) grants from the National Institute of Deafness and Other Communication Disorders (NIDCD) and through the *Language Evaluation, Assessment, and Development in Educational Research* (LEADER) project H325D140096 from the Office of Special Education, U.S. Department of Education.

produced more nonresponses in English than in Spanish. Children with DLD produced more of each error type than their TD peers, particularly in Spanish.

**Conclusions:** Regardless of language ability, bilingual children demonstrated greater difficulty with lexical retrieval for pictured items in the semantically-related context than in the unrelated context. However, bilingual children with DLD produced more errors of all types than is typical for children developing more than one language. Their greater error rates are not secondary to limited second language exposure, but instead reflect deficits inherent to the nature of language impairment. Results from this study are discussed using a framework of semantic constraint, where we propose that because bilingual children with DLD have impoverished semantic networks, and this knowledge insufficiently constrains activation for lexical selection, thereby increasing error production.

# Introduction

When asked to name a picture of a strawberry, why might a child say *cherry*? Error production during naming tasks provides a glimpse of lexical organization and processing, and demonstrates differences between children with and without Developmental Language Disorder (DLD). Error production, including the number and types of errors produced, on behavioral tasks allows us to examine the nature of naming difficulties in children with DLD. Further, identifying the sources of difficulties contributes to understanding the mechanisms implicated in DLD. This is valuable clinically because the locus of language impairment informs best-practices as to how we can effectively identify children with DLD—particularly those who are developing more than one language. Error production can be the result of semantic deficit inherent to language impairment, can occur in developing bilinguals secondary to gaps in knowledge from limited exposure to one language, or a phenomenon associated with typical development. Differences, however, occur with regard to the frequency and types of errors produced by each of these groups. As such, disentangling the effects of language impairment and language exposure on error production in bilinguals is clinically and theoretically valuable. A unique avenue for investigating lexical processing and lexical-semantic knowledge via error production is blocked cyclic naming, which requires children to name pictured items in the context of either semantically-related or unrelated items. In this paper, we are particularly interested in determining how language exposure influences patterns of error production across languages in Spanish-English bilingual children with and without DLD.

Patterns of accuracy and error production during naming inform our understanding of lexical processing in bilingual children developing two lexicons and how these processes break down in DLD. The first aim of this investigation is to evaluate the patterns of accuracy (i.e., a score of 1 or 0 determined by the final response), error frequency (i.e., the total number of errors produced), and the types of errors produced by bilingual children with and without DLD across languages. Based on the extant literature (e.g., Lahey & Edwards, 1996; McGregor, Newman, Reilly, & Capone, 2002), we expect that children with DLD will produce patterns of error types that reflect difficulty with lexical processing as well as deficient semantic development relative to their typically-developing (TD) peers, even after accounting for language experience. The second aim is to determine whether the semantic-relatedness manipulation in blocked cyclic naming affects errors differently for bilinguals with DLD compared to those without DLD. Clinically, error production patterns that indicate a deficit in lexical processing, rather than a difference in language knowledge based on language exposure, could support best practices for assessment of developing bilinguals with a range of English exposure.

**Lexical access and retrieval for bilinguals**

Bilingualism may have multiple consequences for word production, including increased opportunities for both inter- and intra-lexical errors. Many studies have shown that bilingual language production co-activates semantic and phonological representations in both languages (e.g., Colomé, 2001; Costa, Caramazza, & Sebastian-Galles, 2000; Kroll, Dijkstra, Janssen, & Schriefers, 2000; Gollan & Acenas, 2000; Hermans, Bongaerts, De Bot, & Schreuder, 1998). If spreading activation spans both languages, bilinguals may have more word representations competing for selection, and thus more opportunities for error. Additionally, efficiency of lexical access and retrieval for bilinguals depends on experiences with both languages, and the language of exposure determines the type of words they encounter and the opportunities for constructing meaning associated with each word. Thus, language exposure directly impacts the breadth and depth of lexical-semantic knowledge. Bilinguals produce individual words less often in each language than monolinguals do because of divided opportunities to use each language. As a result, words are functionally lower in frequency which may potentially increase latencies and error frequency (e.g., Frequency Lag Hypothesis; Gollan, Slattery, Goldenberg, Van Assche, Duyck, & Rayner, 2011). Poorer naming efficiency in developing bilinguals could therefore stem from several sources, including weaker links between lexical-semantic information, increased lexical competition, deficient language ability, and/or gaps in lexical knowledge.

Unbalanced language exposure results in asymmetrical semantic performance, where bilinguals have better performance on semantic tasks in one language compared with the other. The variance in naming that is accounted for by experience with a particular language exceeds that which is explained by total language exposure (as measured by age; Sheng, Bedore, Peña, & Fiestas, 2013). Indeed, Sheng (2014) found that, as young TD Mandarin-English bilingual children became increasingly more dominant in their second language, English, they demonstrated greater gains in lexical-semantic development compared to their first language, Mandarin (see also Oppenheim, Griffin, Peña, & Bedore, 2019, for similar results from Spanish-English bilinguals). Extending this to the development and organization of semantic knowledge, Sheng et al. (2013) used a repeated word association task to index this knowledge in the context of bilingual children's varying language exposure. Unsurprisingly, bilingual children produced a greater number of and more developmentally mature responses in their more dominant language. This highlights the importance of language exposure for the types of knowledge bilinguals can express productively on structured language-based tasks. With regard to error production, Sheng (2014) found that TD bilingual children produce different error types across their languages, where more semantic errors occurred in the more advanced language while more phonological and indeterminate (i.e., "I don't know") errors occurred in the weaker language. Currently, we do not know how language exposure is related to the frequency and types of errors bilinguals produce. An important next step is to examine how language exposure influences error production for bilingual children as the types and frequency of errors may be exacerbated by DLD despite differences in language exposure. This information would enable clinicians to effectively gauge language impairment status among a continuum of language exposure for young bilingual children.

Returning to our original problem, where a bilingual child said "cherry" for the pictured item *strawberry*, we assume that attempting to name a picture involves co-activating all words that share the target's features, and these become alternative candidates (e.g., *orange* as another fruit; *red* as a shared perceptual feature; e.g., Oppenheim, Dell, & Schwartz, 2010). Additionally, these items are also phonologically related in Spanish (i.e., *cereza* and *fresa* for cherry and strawberry, respectively), which can cause activation to spread to their English-language translations (see Oppenheim, Wu, & Thierry, 2018, for a review and model). Because *cherry* and *strawberry* share overlapping features at both the semantic and phonological levels, it may be unsurprising that a bilingual child would produce this error. For a TD bilingual child who has a densely-packed lexical-semantic network in both languages, lexical selection of the target would likely be successful due to strong lexical connectivity and interaction between processing levels within the lexicon, despite high lexical coactivation (see Dell & O'Seaghdha, 1992, as well as Dell, Schwartz, Martin, Saffran, & Gagnon, 1997, for examples of interactive models). However, for a bilingual child with limited exposure to one of his/her languages or who has DLD, the connections between semantic and lexical information may be insufficient for efficient lexical processing. Combining bilingualism and DLD extends previous literature on lexical processing, providing unique insight into how lexical access and retrieval is dependent upon the structure of lexical-semantic networks for children developing more than one language.

### Error production patterns for children with and without DLD

Picture naming tasks have been used to explore accuracy and error types for bilingual children with and without DLD. TD bilingual children (e.g., Gross, Buac, & Kaushanskaya, 2014) and their peers with DLD (Degani, Kreiser, & Novogrodsky, 2019) have poorer accuracy on a variety of single-word picture-naming naming tasks in comparison to TD monolingual children. Monolingual children with DLD produce more and a greater variety of error types than their TD peers, including more semantic (Lahey & Edwards, 1999; McGregor, 1997) and phonological errors (Lahey & Edwards, 1999; Sheng & McGregor, 2010). On word association tasks, monolingual children with DLD produce fewer semantic associations and more unrelated or unclassifiable (e.g., "I don't know") responses than TD monolinguals, demonstrating deficits in lexical-semantic organization and/or retrieval; these deficits are greater than expected given expressive vocabulary knowledge. For monolinguals with DLD, error production is therefore believed to result from impoverished representations in long-term memory (Sheng & McGregor, 2010). In general, children with DLD are hypothesized to have underspecified word representations with fewer, weaker lexical-semantic connections, resulting in an overall poorer lexical quality in comparison to their TD peers (e.g., Kan & Windsor, 2010). This impedes speed and accuracy of lexical access and retrieval (see storage hypothesis by Kail & Leonard, 1986), which can also be called lexical processing efficiency.

**Lexical access and retrieval during blocked cyclic naming**

In a popular method of studying lexical processing efficiency, blocked cyclic naming, participants name pictured objects that appear individually on a computer screen, as quickly and accurately as possible. Interactive activation models of speech production posit that when an individual sees a pictured item, the semantic (and/or conceptual) information associated with it in the lexicon is activated. In a rich semantic network, nodes are interconnected via links: the greater the number of nodes and links, the more robust the representation. When a node is activated, activation automatically travels outward from the semantic node via links to related nodes within the semantic network and outward to the lexical and phonological levels (Collins & Loftus, 1975; Dell & O'Seaghdha, 1992). For example, the representation of the word *lamp* is realized when the corresponding features (e.g., bright, yellow, furniture, etc.) within the semantic network become activated, and activation spreads to the lexical entry and the phonological constituents (i.e., /l/ /æ/ /m/ /p/). At the semantic level, a greater number of features will result in a more refined representation; in contrast, few semantic features with less interconnectivity among nodes limit how strongly the representation can be activated. Additionally, representations with overlapping semantic features (e.g., *mirror*: furniture) become activated at each level but to a lesser degree. Accurate lexical selection is supported by feedback from lexical to semantic and phonological to lexical levels, increasing the relative activation level for the target lexical item (e.g. Dell, 1986). The greater activation of targets reinforce the lexical system, allowing the lexical selection mechanism to quickly and accurately select the target, which should be the most highly activated lexical item, for production (Dell & O'Seaghdha, 1992; Poulisse & Bongaerts, 1994).

In the blocked cyclic naming paradigm, pictures are divided into small sets composed of items from either the same semantic category (e.g., bus, car, and airplane are all members of the transportation category) or mixed semantic categories (e.g., airplane, chair, and apple are members of different categories: transportation, furniture, and fruit, respectively). In each block, participants repeatedly name the pictures from one set, in several randomly ordered cycles. The classic result is that individuals are slower and less accurate when trying to name pictures in the blocks that represent a single semantic category (i.e., Same-category context) than when trying to name them in mixed semantic category blocks (i.e., Mixed-category context). One potential explanation for this phenomenon assumes that as multiple category members with overlapping features simultaneously become more strongly activated, as a result of residual activation, they compete more for lexical selection (Damian, Vigliocco, & Levelt, 2001). However, this can also be explained as a result of more persistent changes to semantic-to-lexical connections that affect the candidates' re-activation (i.e. incremental learning, e.g., Oppenheim, Dell, & Schwartz, 2007, 2010; see also Howard, Nickels, Coltheart, & Cole-Virtue, 2006). Under both explanations, the target lexical representation becomes relatively less activated, compared to the alternatives, so the likelihood of selecting a non-target alternative increases. The intensity of semantic interference directly corresponds to the similarity between set members, where lexical items that are closer semantic neighbors will show greater effects—exhibited behaviorally by longer naming latencies and/or increased error production—than more distant members (Vigliocco, Vinson, Damian, & Levelt, 2002; see also Alario & Moscoso del Prado Martin, 2010). Thus, items presented in the Same-category context are named more slowly and are more prone to error (e.g., "cherry" for *strawberry*) than items in the Mixed-category context. When naming items during blocked cyclic naming, a child may therefore name a picture of a strawberry as "cherry" because 1) the items are semantically related, and 2) the target and alternative names have similar levels of activation than they typically would have in other every-day contexts. Errors of this nature may further be exacerbated by DLD, where impoverished semantic representations contribute to less accurate lexical retrieval. Importantly, understanding typical and atypical patterns of error production in the context of dual language development could provide a useful basis for clinicians to identify bilinguals who potentially have DLD.

### Monolingual children's performance on blocked cyclic naming tasks

Until now, error production on blocked cyclic naming tasks has not been explored in bilingual children with and without DLD. There is, however, evidence that even young TD monolinguals experience difficulty rapidly retrieving lexical items in the context of other semantically-related items. Two studies of blocked cyclic naming performance with monolingual English-speakers ages 5 to 7 years and 10 to 12 years old showed that children were slower and less accurate at naming items in a Same-category context than in a Mixed-category context (Boelens & La Heij, 2017; Snyder & Munakata, 2013). To date, only one study has evaluated the types of errors that young children produced in the context of semantic interference (Charest, 2017). Of the 1047 responses produced by 3-year-old monolingual English-speakers, 190 responses were errors. Of these errors, 54 were classified as naming errors, and included 44 contextual word intrusions (defined as the production of another item name from the same set of items). Overall, Charest proposed that her results showed continuity between the picture naming processes of children and adults, demonstrating that even young children have shared activation among semantically related items, which produces lexical co-activation during picture naming; thus, children's error production can be assumed to result from typical language production processes.

Only one study, to our knowledge, has evaluated performance on the blocked cyclic naming paradigm for children with DLD (i.e., Ladányi & Lukács, 2016). This study considered monolingual Hungarian-speaking children with and without DLD and focused on naming latencies instead of errors. Excluding trials with incorrect responses, these children with and without DLD showed similar naming latencies in their first naming cycles for pictured items in the Same semantic category context. The investigators suggest this pattern of performance may be secondary to decreased co-activation, or competition, for children with DLD due to weaker lexical-semantic connections, resulting in weaker reactivation of alternative names (Ladányi & Lukács, 2016). Lexical processing in the context of blocked cyclic naming has yet to be explored in bilingual children. This avenue of research may provide new insight into how bilinguals with DLD differ from their TD peers with regard to lexical processing and structure. Specifically, findings from this work may inform clinical decision-making pertaining to how bilingual children with DLD are identified. As bilinguals with DLD have sparse semantic networks (Sheng, Peña, Bedore, & Fiestas, 2012), we anticipate that they will produce quantitatively more (i.e., higher error frequency) and qualitatively different patterns of error types in comparison to their TD peers. This information lends itself to clinical practice, as clinicians can use this as an indicator of potential impairment, where error frequency and error types may point to breakdowns in lexical processing associated with DLD in contrast to patterns typical for bilinguals.

### Summary and research questions

The patterns of error production during a blocked cyclic naming task provide a way to evaluate how lexical processing and organization interact in bilingual children with typical and disordered language. In line with previous research (McGregor, Newman, Reilly, & Capone, 2002; Sheng, Peña, Bedore, & Fiestas, 2012), we assume that children with DLD have impoverished word representations due to sparser semantic knowledge and fewer lexical-semantic connections compared to their TD peers. The degree of knowledge represented in the lexicon makes words more or less vulnerable to retrieval failure and, thus, more prone to error production. Because of this deficit and additional potential effects that are characteristic of bilingualism (e.g., weaker links, limited language exposure, lexical competition), there is a greater likelihood for increased rates of error production resulting from less densely-packed lexical items and limited semantic depth to constrain the selection process. A rich semantic network comprising depth of semantic knowledge and interconnectivity is required to appropriately constrain activation and support efficient lexical processing. We asked the following questions to guide our investigation:

1. How does language exposure affect accuracy and error frequency during blocked cyclic naming for Spanish-English bilingual children with and without DLD?
    a. Is the pattern dependent upon the blocked cyclic naming context?
    b. Is the pattern dependent upon the test language?

Based on previous comparisons of monolingual children with and without DLD (e.g., Lahey & Edwards, 1999; Sheng & McGregor, 2010), we expect that TD bilinguals will ultimately be more accurate and produce fewer errors than their peers with DLD in each of their languages (Spanish and English). While accuracy was determined by scoring the child's final answer for each target, responses before the final answer could be classified as an error. For example, a child who often accurately self-corrects (e.g., "cat… no—dog" for a stimulus picture *dog*) would have high accuracy as well as high error frequency. Consistent with previous blocked-cyclic naming studies with monolingual children (Boelens & La Heij, 2017; Charest, 2017; Ladányi & Lukács, 2016; Snyder & Munakata, 2013), we expect that bilinguals with and without DLD will produce more errors in the Same-category context than the Mixed-category context, due to the relatively greater coactivation of alternative items. If children with DLD have impoverished semantic representations, dispersing activation to nontarget lexical entries (i.e., a lack of constraint), or weaker links between lexical-semantic information, we would expect this difficulty to increase their error production in the Same-category context secondary to deficient lexical retrieval.

2. After accounting for language exposure, do TD and DLD bilingual children produce different patterns of error types?
    a. Is the pattern of error types dependent upon the blocked cyclic naming context?
    b. Is the pattern of error types dependent upon the test language?

Due to the locus of impairment in bilingual children with DLD, we expect their error types to reflect difficulty accessing (i.e., no responses) and retrieving semantic (i.e., intrusions and semantic errors) knowledge; however, this may also be dependent upon the children's language exposure, ability status, and blocked cyclic naming condition.

# Methods

### Participants

Data for the current study were drawn from Spanish-English bilingual children who participated in a larger study, *Cross-Language Outcomes of Typical and Atypical Development in Bilinguals* (Peña, Bedore, & Griffin, 2010), in Texas. Protocols for this study were approved by the Institutional Review Board at the University of Texas at Austin. Participants comprised a total of 238 bilinguals, including those who were diagnosed with DLD ($n = 36$) and those with typical development ($n = 202$). While 251 children were originally administered the blocked cyclic naming task, children were excluded based on outlier scores indicated by Mahalanobis distance ($n = 9$), or for missing impairment status and/or monolingual English status during the confirmatory phase indicated by their ability to respond in English only ($n = 4$). To reach an adequate sample size for our DLD group, we over-sampled children with DLD, exceeding the 7.4% prevalence reported by Tomblin et al. (1997). To avoid bias, children with DLD who participated in this study were not specifically recruited based on their deficits in semantic skills. We recruited and included children with DLD who represented the range of linguistic strengths and weaknesses—a heterogeneous profile consistent with the population of children with DLD (e.g., Leonard, 2014). Children who were classified as TD had no known language, hearing, or health concerns at the time of testing. A cross-sequential design was implemented such that children were recruited and tested in either kindergarten through 3$^{rd}$, 2$^{nd}$ through 5$^{th}$, or 4$^{th}$ through 5$^{th}$ grade. All children were raised in homes where Spanish was the primary language spoken, and they had varying levels of exposure to English, their second language. Across all children the mean age of first exposure to English was approximately 2.8 years, while they received exposure to Spanish from birth. The average combined language input and output was 57.3% in Spanish for the TD group and 56.1% in Spanish for the DLD group. See Table 1 for demographic information.

Table 1: Demographics for children with and without DLD.

| | | | | n | |
|---|---|---|---|---|---|
| | | TD | DLD | TD | DLD |
| Mean age in years (range) | | 7.7 (5.1-11.5) | 7.5 (5.4-10.5) | 202 | 36 |
| Sex (girls, boys) | | 109, 93 | 13, 23 | 202 | 36 |
| Mean percent Spanish input/output (SD) | | 57.3% (23.8%) | 56.1% (26.1%) | 202 | 36 |
| Mean age of first English exposure in years (SD) | | 2.7 (1.9) | 3.0 (2.2) | 200 | 34 |
| Mean Hollingshead score (SD) | | 2.67 (1.51) | 2.46 (1.34) | 173 | 28 |
| Mean ITALK Parent/Teacher rating (SD) | | 101.77 (10.27) | 86.89 (15.52) | 202 | 36 |
| Mean BESOS composite score (SD) | | 97.25 (13.27) | 72.56 (9.56) | 201 | 36 |
| Mean BESA/BESA-ME standard scores (SD) | | | | 198 | 36 |
| English | Semantics | 101.58 (21.50) | 74.26 (13.00) | | |
| | Morphosyntax | 86.87 (17.56) | 61.63 (9.88) | | |
| Spanish | Semantics | 99.41 (14.70) | 76.03 (14.53) | | |
| | Morphosyntax | 92.10 (17.19) | 66.53 (10.71) | | |
| Mean TNL score (SD) | | | | 202 | 36 |
| English | | 79.22 (16.88) | 64.66 (9.47) | | |
| Spanish | | 100.47 (13.78) | 80.00 (9.69) | | |

Note: All measures were collected during the child's first year of testing. BESOS, BESA, and BESA-ME scores were calculated using each child's best language scores for semantics and morphosyntax. The BESOS score was a composite score of semantics and morphosyntax.

Participants in this study were included in two phases of testing: a screening phase and a confirmatory phase. The screening phase was implemented to identify children who were suspected of having DLD. The confirmatory phase began one year later, and children in kindergarten through fifth grade could participate up to three times over three consecutive years. During the confirmatory phase, children were administered standardized and experimental measures in both Spanish and English to either rule in or rule out a diagnosis of DLD. Children between the ages of 4;0 and 6;11 were administered the *Bilingual English-Spanish Assessment* (BESA; Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2018), a standardized assessment measure, while children ages 7;0 to 11;6 were administered the *Bilingual English-Spanish Assessment-Middle Extension* (BESA-ME) field test version (Peña, Bedore, Gutiérrez-Clellen, Iglesias, & Goldstein, 2016), which is an assessment measure under development. Each year children participated in three or four testing sessions that were approximately 30 to 45 minutes that were held in relatively quiet locations at their schools. The order of the tests and the test languages was randomized across participants.

To determine ability status we followed the same procedures as Bedore et al. (2018). Children were qualified as having DLD if they met 4 of 5 criteria: 1) a rating of 4.2 or below out of 5 in both languages by the parent and/or teacher on the *Instrument to Assess Language Knowledge* (ITALK; Peña et al., 2018), 2) a score greater than one standard deviation below the mean in both languages on the morphosyntax subtest of the BESA (Peña et al., 2018) or the BESA-ME field test version (Peña et al., 2016), 3) a score greater than one standard deviation below the mean in both languages on the semantics subtest of the BESA (Peña et al., 2018) or the BESA-ME field test version (Peña et al., 2016), 4) a composite score on the *Bilingual English-Spanish Oral Screener* (BESOS; Peña, Bedore, Iglesias, Gutiérrez-Clellen, & Goldstein, 2008) that was greater than one standard deviation below the mean in both languages, and 5) a score on the *Test of Narrative Language* (TNL; Gillam & Pearson, 2004; TNL-Spanish; Gillam, Peña, Bedore, & Pearson, 2006) that fell greater than one standard deviation below the mean in both languages.

**Measures**

*Bilingual Input Output Survey* (BIOS; Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2018). Children's language exposure was calculated by obtaining parent and teacher reports of children's language use and language input for each hour of the day in their home and school environments.

*Blocked cyclic naming task.* Stimuli comprised 40 (20 for each language) color photographs of real objects. Audio recordings of each item name from a female voice were also used in a pre-experiment familiarization procedure. Pictured items were chosen based on high-familiarity and high convergence in naming for bilingual preschool children during a pilot testing phase. Based on these criteria, five cognates in the English version and four cognates in Spanish version of the blocked cyclic naming task were used. Different stimuli in each language were used to limit the effect of cross-linguistic transfer. Phonologically, initial consonants were varied across the Same- and Mixed-category conditions to avoid phonological priming. In each language, four items from four semantic categories were selected to form the Same- and Mixed-category item sets (see Appendix A). An additional unrelated set of four filler pictures was included for each language in order to evaluate repetition priming independent of semantic overlap among items.

### Procedures

The blocked cyclic naming data used in this study were obtained from the confirmatory phase of testing, and this task was administered to children once each year they participated in the study. As such, children could have completed this task a maximum of three times. Prior to completing the task, children were familiarized with the test stimuli to ensure that they could accurately produce the names of the target items. They were shown a total of 20 pictures and asked to repeat the name of each object after it was spoken by a recorded voice prompt. They were then given instructions to name each of the items as quickly and accurately as possible, just as they had practiced. During testing, each trial began with a fixation point, displayed in the center of the screen for 500 milliseconds (ms). A blank screen was then presented for 250 ms, followed by a stimulus picture, presented in isolation in the center of the screen. The picture remained on the screen for 1200 ms after the voice key detected a response, or a maximum of 3500 ms after picture onset if the voice key had not yet detected a response. The next trial began after a 750 ms intermission. To monitor accurate voice key sensitivity, a red frame appeared around the stimulus picture when a response was detected. If the voice key did not detect a response, a message appeared instructing an error with the microphone placement. These automated protocols allowed the examiner to monitor and adjust as needed. Children were redirected back to the task if they became noticeably fatigued or distracted. During testing, each object was named a total of 12 times in each language. These items were pseudo-randomly organized into blocks, where each set of four items was cycled through six times, and blocks containing the filler items initiated and ended the session. This resulted in a total of 96 trials for each semantic context (Same- and Mixed-category) in each language, comprising 384 total trials each year of testing.

### Transcription and coding

All test stimuli were transcribed offline from recordings by trained undergraduate students who were blind to child language ability. After the data were transcribed, a coding scheme was created to classify children's responses by overall item accuracy, error frequency, and error type. While accuracy measured how many items were correct for each child (maximum = 96 for each language and context), children may self-correct after producing an error. This means that although the final response was correct, earning a score of 1 for the item, they initially produced a word in error. This initial response would receive a code to quantify the type of error; thus, while a child could receive a high accuracy score, (s)he could also produce errors (e.g., "gato...uh-caballo" for *caballo*). To determine if patterns of error frequency differed from accuracy, we calculated the total number of errors produced (i.e., error frequency) across languages and contexts for children with and without DLD.

Four trained coders—the first author and three trained undergraduate students—then coded all of the responses for error type. Responses that did not match the target exactly were classified as either correct or incorrect (i.e., accuracy) and were further categorized using the following codes: intrusions, phonological errors, semantic errors, or other errors. See Appendix B for a coding decision tree. Intrusions were defined as substituting the name of an item previously pictured (e.g., production of "dog" for the pictured item *pumpkin* after a dog was shown in a previous block). Phonological errors comprised rhymes (e.g., "hook" for *book*) and alliterations (e.g., "seal" for *snake*); nonwords were also accounted for within this category (e.g., "baispa" for *backpack*). Based on the coding system by Sheng, Bedore, Peña, and Fiestas (2013), the semantic errors included syntagmatic, or thematic relationships (e.g., "dream" for *pajamas*), and paradigmatic, or categorical relationships (e.g., "fruit" for *apple*). The other errors included in this coding scheme were code switching (e.g., "bear" for *oso*), self-correction (e.g., "fish—I mean egg" for the target *egg*), unclassifiable (e.g., "feet" for *apple*), no response (e.g., "um no sé"), and cut-off (e.g., "back—[cut-off]" for *backpack*). Responses that included code switching and self-corrections were subsequently classified as either successful or unsuccessful. For responses that were cut-off, bisyllabic words were counted as correct if up to two phonemes were omitted from the recording and monosyllabic words were counted correct if up to one phoneme was cut-off. If a child produced a minor articulation error (e.g., "benguin" for *penguin*, "owange" for *orange*) that was determined to be developmentally appropriate, then the item was scored as correct. Synonyms for target items (e.g., "bici" for *bicicleta*) and dialectal variations (e.g., "pescao" for *pescado*) were also counted as correct.

Because responses could earn more than one code, productions were not necessarily mutually-exclusive. All responses were coded to account for potential overlap. Descriptively, we found that 2.91% of the total number of items were double-coded (see Appendix C).

For the purposes of our analyses, the following codes were not used for this study: minor articulation error, cut-off, successful code-switch and/or successful self-correction, phonological nonword (for rhyme and alliteration), and phonological resemblance. In order to better evaluate potential breakdowns in the lexical system (i.e., semantic, lexical, or phonological), the error category codes were collapsed into composites: phonologic (rhymes and alliterations), semantic (syntagmatic and paradigmatic), and miscellaneous errors (self-corrections, code-mixing, unclassifiable: real and non-words). Intrusions and no responses comprised their own categories.

### Reliability

Two trained research assistants coded a random sample of 10% of the total responses. Point-by-point reliability was 100% for accuracy and 96.5% for the total number of errors. Reliability for each error type is as follows: 99.7% for intrusions, 97.7% for phonologic, 99.7% for semantic, 99.7% for no responses, and 98.4% for miscellaneous.

# Results

### How does language exposure affect accuracy and error frequency?

The first research question evaluated the effect of language exposure on accuracy and error frequency produced across blocked cyclic naming contexts (Same- and Mixed-category) and across languages (Spanish and English) for children with and without DLD. We explored this using two mixed model Analyses of Covariance (ANCOVAs). Task language and semantic context served as the within subjects variables, while ability group (TD or DLD) served as the between subjects factor. Language exposure was the covariate for both analyses. The dependent variable for the first mixed model ANCOVA was the children's accuracy in each context and language, and for the second mixed model ANCOVA it was the number of errors produced (i.e., error frequency) in each context and language. For accuracy, a total of 96 trials per language and context (i.e., 384 total trials) were administered at each time. Because children could be administered this task up to once per year over a three year period between kindergarten and fifth grade, dependent variables comprised the raw score of their performance averaged across time for the years they completed the task. Thus, the highest average accuracy a child could obtain was 96 for each language and context. For error frequency, the sum of the number of errors produced at each year were averaged across time. See Table 2 for descriptive data for children's accuracy and error frequency across test languages and contexts. The assumption of homogeneity of variance was violated, as indexed by Levene's test of equality of variances, for English measures of accuracy and error frequency across the Same-category ($p$ = .001) and Mixed-category (accuracy: $p < .001$, error frequency: $p$ = .013) contexts; corresponding Spanish test statistics were nonsignificant for both contexts. To correct for this, we applied a natural log transformation to both the English and the Spanish data. Results remained the same; thus, we report the original results.

*Accuracy patterns.* As seen in Figure 1, patterns of accuracy demonstrate that TD children were more accurate than children with DLD. While children with DLD maintained consistently lower accuracy across languages and semantic blocking contexts, TD children obtained slightly higher accuracy for English compared with Spanish. TD children also achieved somewhat better accuracy during the Mixed-category context as compared with the Same-category context. Results from the mixed model ANCOVA showed a significant main effect of Ability ($F(1, 234) = 9.89$, $p = .002$, partial $\eta^2 = .04$), where TD children were more accurate than children with DLD ($p$ = .002). Additionally, significant main

Table 2: Means (standard deviations) for accuracy and total number of errors produced across task languages and semantic blocking conditions for children with and without DLD.

| | | | Group | |
|---|---|---|---|---|
| Factor | Condition | Language | TD | DLD |
| Accuracy | Mixed | English | 89.88 (5.93) | 85.85 (9.84) |
| | | Spanish | 88.99 (8.20) | 85.96 (7.05) |
| | Same | English | 89.58 (6.29) | 85.81 (9.86) |
| | | Spanish | 87.75 (8.60) | 85.00 (6.45) |
| Total Errors | Mixed | English | 2.23 (1.78) | 3.48 (2.56) |
| | | Spanish | 2.13 (2.12) | 3.37 (2.00) |
| | Same | English | 2.36 (1.86) | 3.51 (2.69) |
| | | Spanish | 2.64 (2.30) | 3.91 (2.04) |

effects for Task Language ($F(1, 234) = 19.42$, $p < .001$, partial $\eta^2 = .08$) and Semantic Context ($F(1, 234) = 6.15$, $p = .014$, partial $\eta^2 = .03$) were found. While pairwise comparisons revealed no significant differences across languages ($p = .307$), children had higher accuracy in English ($M = 87.73$, $SE = .59$)[1] than Spanish ($M = 86.98$, $SE = .71$) and were significantly more accurate during the Mixed-category context than the Same-category context ($p = .001$).

A significant interaction emerged between Task Language and Semantic Context ($F(1, 234) = 5.19$, $p = .024$, partial $\eta^2 = .02$). Bonferroni post-hoc pairwise comparisons were conducted to further evaluate differences among variables in the significant interaction between Task Language x Semantic Context. Children had higher accuracy in the Mixed-category context than the Same-category context ($p < .001$) in Spanish, but the difference between the two contexts did not approach significance in English ($p = .490$). These findings indicate that—regardless of ability—children demonstrated a greater semantic blocking effect for accuracy in Spanish, their first language, than in English, their second language.

A significant Task Language x Language Exposure ($F(1, 234) = 19.64$, $p < .001$, partial $\eta^2 = .08$) interaction was also found. To further evaluate this interaction, we conducted Pearson product-moment correlations and found that children with more Spanish exposure were less accurate in English ($r$ = -.15, $p = .018$, $n = 237$) and more accurate in Spanish ($r = .17$, $p = .011$, $n = 238$).

*Error frequency patterns.* Children with DLD produced more errors than their TD peers, with the greatest number of errors produced by the DLD group in the Same-category context in Spanish (see Figure 2). Indeed, results from the mixed model ANCOVA revealed significant main effects for Ability ($F(1, 234) = 17.74$, $p < .001$, partial $\eta^2 = .07$), Semantic Context ($F(1, 234) = 5.59$, $p = .019$, partial $\eta^2 = .02$), and Task Language ($F(1, 234) = 10.16$, $p = .002$,

[1]Mean and standard error are reported.

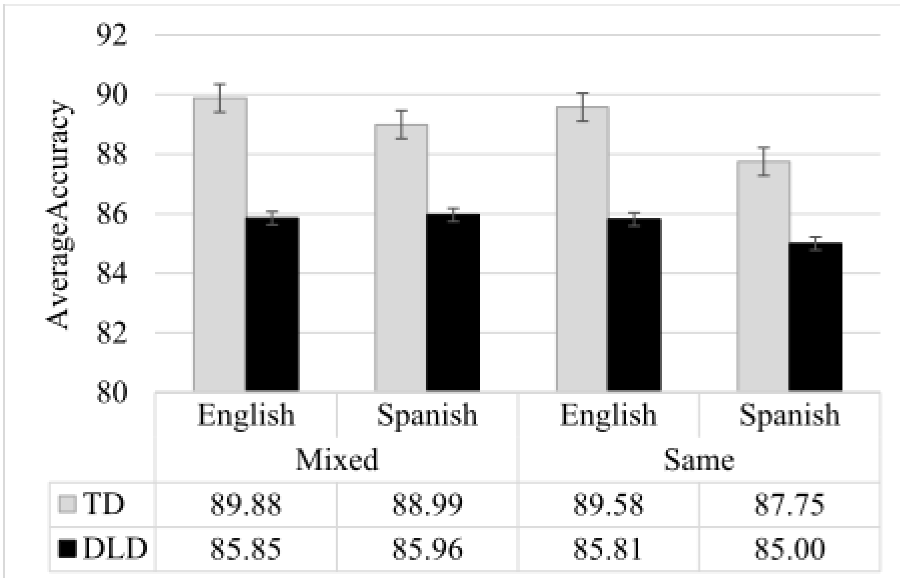


Figure 1: Average accuracy across blocked cyclic naming contexts by language across children with and without DLD.

partial $\eta^2$ = .04). Pairwise comparisons confirmed that TD children produced fewer errors than children with DLD ($p < .001$), and for both groups of children more errors were produced in the Same-category context as compared with the Mixed-category context ($p < .001$). Collapsing across groups of children and contexts, the pairwise comparison showed that English and Spanish error frequencies did not significantly differ ($p = .645$).

A significant interaction between Task Language and Semantic Context ($F$(1, 234) = 5.08, $p = .025$, partial $\eta^2$ = .02) emerged. Results from the Bonferroni pairwise post-hoc comparisons mirrored the results from the accuracy analyses, where children produced fewer errors during the Mixed-category context than the Same-category context in Spanish only ($p < .001$). No significant differences emerged for English ($p = .308$).

A significant Task Language x Language Exposure ($F$(1, 234) = 11.34, $p = .001$, partial $\eta^2$ = .05) interaction was found. We conducted Pearson product-moment correlation to further evaluate the relationship between variables. Results show that the more Spanish exposure children had, the fewer errors they produced in Spanish ($r$ = -.14, $p = .034$. $n$ = 238). The error frequency in English was not significantly related to percent Spanish exposure ($p = .089$).

**Do TD and DLD bilingual children produce different patterns of error types?**

The second research question concerned the types of errors produced by children

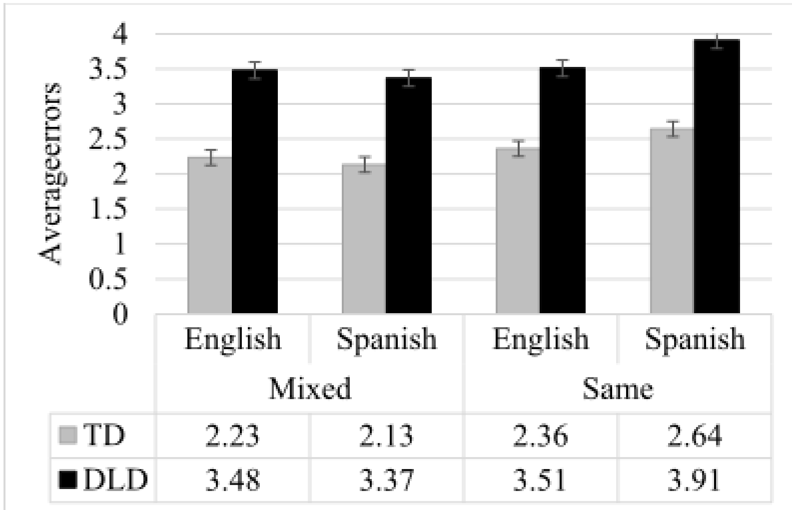


Figure 2: Average number of errors produced across blocked cyclic naming contexts by language for children with and without DLD.

with and without DLD across blocked cyclic naming contexts and languages. Descriptively, as seen in Figure 3, children with DLD produced more of each error type—with the exception of English semantic errors in the Mixed-category context—across languages and contexts in comparison to their TD peers. The average number of intrusion and semantic errors were similar across the two groups of children in English, and these error types were produced the least often across both languages and contexts. Phonological errors were produced the most frequently across both contexts in English and the Mixed-category context in Spanish, while no responses were the most common in the Same-category context in Spanish.

A series of mixed model ANCOVAs were conducted[2] to evaluate differences in the frequencies of the five types of errors (intrusions, phonological, semantic, no response, and miscellaneous) produced across ability groups (TD or DLD), task languages (Spanish and English), and semantic contexts (Same- and Mixed-category). Language exposure served as the covariate. As indexed by Levene's test of equality of variance, the assumption of homogeneity of variance was violated. All data were transformed using a natural log with the constants .01, 1, and 10. We report the best results obtained using the function ln(x + 1). For the function ln(x + 1), Levene's statistic was significant for the intrusions ($p < .001$) and semantic ($p < .001$) error types for the Same-category context in English, the intrusions ($p = .009$) and semantic ($p < .001$) error types in

[2]An omnibus mixed model ANCOVA was also conducted where the five error types were included as an additional within-subjects variable. Results for homogeneity of variance matched the results found for the series of ANCOVAs reported.

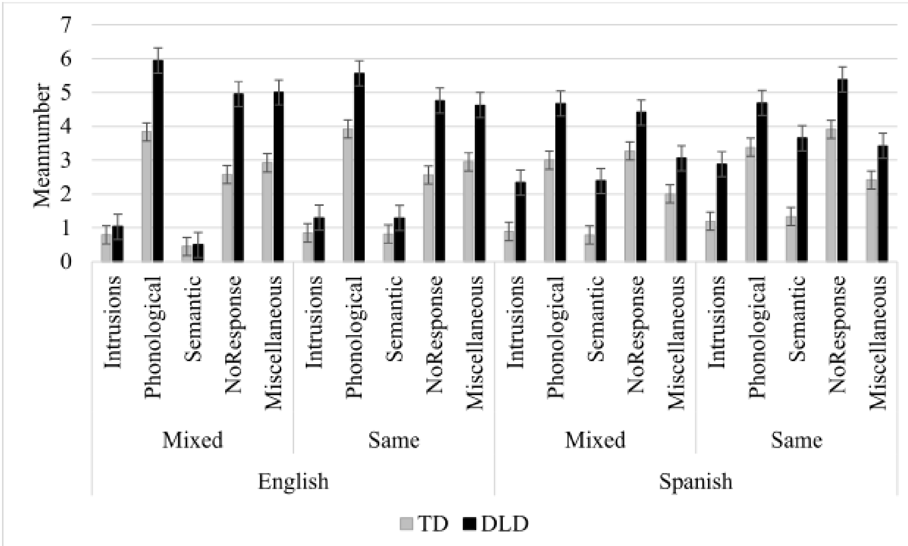


Figure 3: Average number of each error type produced across languages and blocked cyclic naming contexts by bilingual children with and without DLD.

the Mixed-category context in Spanish, and the intrusions ($p$ = .040) and no response ($p$ = .029) error types in the Same-category context in Spanish.

For each error type, the children with DLD produced significantly more total errors than their TD peers. See Table 3 for the results for these main effects. A significant main effect of Task Language also emerged for the intrusions ($F$(1, 215) = 9.00, $p$ = .003, partial $\eta^2$ = .04), semantic ($F$(1, 215) = 22.13, $p$ < .001, partial $\eta^2$ = .09), and no response ($F$(1, 215) = 20.61, $p$ < .001, partial $\eta^2$ = .09) error types, where children consistently produced fewer of each error type in English than in Spanish (intrusions and semantic: $p$ < .001, no response: $p$ = .028). A significant Task Language x Ability interaction emerged for the intrusions ($F$(1, 215) = 20.57, $p$ < .001, partial $\eta^2$ = .09) and semantic ($F$(1, 215) = 30.70, $p$ < .001, partial $\eta^2$ = .13) error types. Bonferroni post hoc comparisons showed that for both error types, both the TD and DLD groups produced significantly more errors in Spanish than in English (intrusions: TD $p$ = .047, DLD $p$ < .001; semantic: $p$ < .001). Additionally, for Spanish, the DLD group produced significantly more errors than their TD peers ($p$ < .001); productions of each error type in English did not significantly differ between groups (intrusions: $p$ = .104; semantic: $p$ = .115).

Children also produced differences in the types of errors across contexts. A significant main effect for Semantic Context ($F$(1, 215) = 12.64, $p$ < .001, partial $\eta^2$ = .06) emerged for the semantic errors. Bonferroni post hoc comparisons showed that children produced significantly fewer semantic errors in the Mixed-

Table 3: Main effects for each error type across ability groups.

| | TD Mean (*SE*) | DLD Mean (*SE*) | *df* (error) | *F* | *p* | partial $\eta^2$ |
|---|---|---|---|---|---|---|
| Intrusions | .53 (.03) | .85 (.06) | 1 (215) | 22.82 | $< .001$ | .10 |
| Phonological | 1.15 (.04) | 1.57 (.10) | 1 (215) | 13.30 | $< .001$ | .06 |
| Semantic | .48 (.02) | .83 (.06) | 1 (215) | 30.90 | $< .001$ | .13 |
| No response | 1.04 (.05) | 1.49 (.12) | 1 (215) | 11.68 | .001 | .05 |
| Miscellaneous | 1.04 (.04) | 1.43 (.09) | 1 (215) | 16.95 | $< .001$ | .07 |

category context as compared with the Same-category context ($p < .001$). For the no response error type, a significant Task Language x Semantic Context ($F(1, 215) = 10.99$, $p = .001$, partial $\eta^2 = .05$) interaction emerged, where no responses were significantly more frequent in Spanish than in English for the Same-category context only ($p = .003$); the difference was not significant in the Mixed-category context ($p = .394$).

Language exposure was found to impact the occurrence of the no response error type, where significant Language Exposure x Task Language ($F(1, 215) = 14.71$, $p < .001$, partial $\eta^2 = .06$) and Language Exposure x Task Language x Semantic Context ($F(1, 215) = 4.72$, $p = .031$, partial $\eta^2 = .02$) interactions emerged. Pearson product-moment correlation coefficients were conducted to further examine the relationship between these factors. Language exposure was significantly related to both the Mixed-category ($r = .16$, $p = .014$) and the Same-category ($r = .22$, $p = .001$) contexts in English only; thus, no significant correlations were found for either context in Spanish.

# Discussion

The purpose of this study was to investigate differences in error production patterns on a blocked cyclic naming task for Spanish-English bilingual children with and without DLD who have a range of second language exposure. Error production provides a window into how bilingual children organize and process language across multiple lexicons. For children with DLD, deficits result in the production of errors that are not readily identified in a standard picture naming task (e.g., receptive vocabulary test); errors are typically accounted for using a broad accuracy measure, which is the common scoring method utilized in assessments of language ability. An additional consideration for bilingual children is the quantity and quality of exposure to each of their languages, as this influences the depth and breadth of their lexical-semantic knowledge. In this study, we found that children with DLD produced significantly more of each of five error types in comparison to their TD peers, and in our sample this difference was robust against variations in the amount of language exposure. Importantly, however, these error patterns were influenced by the context of test stimuli (Mixed- or Same-category blocked cyclic naming context) and the language of testing (Spanish or English).

**Accuracy and error frequency**

TD children in our study were significantly more accurate and produced fewer errors than those with DLD. However, the effects of test languages and blocked cyclic naming contexts did not significantly differ between the groups of children, suggesting that poorer naming was due to deficient lexical processing that is inherent to DLD. Considering our semantic constraint hypothesis, we suggest that fewer lexical-semantic connections and weaker semantic representations might be associated with insufficient activation and/or retrieval of the target name, resulting in lower accuracy and increased error production. For instance, substituting "orange" for the pictured item *pumpkin* demonstrates that the semantic competitor was more strongly activated at the lexical level, leading to inaccurate lexical retrieval. While a child with DLD may know both words to a degree (i.e., indicating lexical breadth), (s)he may have few connections supporting and distinguishing them (i.e., limited semantic depth). As such, this child with DLD may know that a *pumpkin* is round, orange, grows, one can pick it in the fall, and it is edible. However, many of these semantic features overlap with those of an *orange*, including the color of both items. This results in strong activation for both of these lexical items. In contrast, TD children have additional features stored for each item (e.g., for *pumpkin*: has a stem, one can carve it, make soup out of it, it has seeds and ribbed skin, it grows on a vine, it is a type of gourd, etc.). The stronger and/or more numerous semantic features would guide activation to the lexical item for *pumpkin*, allowing this item to be more highly activated relative to its lexical competitor, *orange*. This hypothesis of semantic constraint, in concordance with theories of interactive models of lexical access (e.g., Dell & O'Seaghdha, 1992) and proposals of lexical retrieval for monolingual English-speaking children with DLD (e.g., McGregor, Newman, Reilly, & Capone, 2002), could account for the overall higher accuracy and reduced error production for TD bilingual children relative to their peers with DLD.

While our findings provide preliminary evidence in support of the hypothesis of semantic constraint, more research is needed to confirm this account. Alternatively, weaker links (e.g., Gollan, Montoya, Cera, & Sandoval, 2008) would result in difficulty accessing and retrieving lexical information for production due to limited availability for spreading activation. Another possible explanation may emphasize the role of memory in encoding semantic information. As monolingual children with DLD are reported to have deficient working memory, it may be that limitations in working memory would restrict the linguistic details encoded—including extraction of semantic features from environmental input. As such, fewer features are stored in semantic networks within long term memory; reciprocally, impoverished language knowledge places a greater demand on working memory for retaining unfamiliar information (Archibald, 2018). Each of these accounts could potentially explain why children with DLD have poorer accuracy and increased error frequency on picture naming tasks in comparison to their

TD peers. While the latter explanation concerning deficient working memory complements our hypothesis of semantic constraint, our current investigation does not disambiguate between these accounts. Future research should test these accounts to determine the differential effects secondary to DLD or divided language exposure.

Language exposure is key for bilingual language development (Bedore, Peña, Griffin, & Hixon, 2016). In the context of picture naming during the blocked cyclic naming task, when the children with and without DLD were combined, language exposure was found to be related to both accuracy and error frequency in Spanish. Children, regardless of language ability, who had greater Spanish exposure were more accurate and produced fewer errors when naming pictured items in Spanish. Language exposure was also positively correlated with the number of no response error types children produced in each semantic context in English, indicating that the greater the child's relative Spanish exposure (and therefore the less English exposure), the more difficulty the child experienced with lexical access in English. The language exposure metric used in this study was the average percentage of input and output to Spanish throughout each child's typical week. Given that greater exposure to the language would allow for children to build more robust lexical representations, these findings were unsurprising. However, an unexpected finding was that language exposure was not correlated with error frequency in English, as we had anticipated a greater number of errors produced in the child's second language.

One possibility is that using a measure focused on language output may be a better predictor of language production errors. Although language input and output (i.e., language experience) are strongly correlated, input and output differentially contribute to development of language knowledge across domains for bilinguals (i.e., semantics and morphosyntax; Bohman, Bedore, Peña, Mendez-Perez, & Gillam, 2010). While input may help young children develop lexical-semantic networks, allowing for the understanding of what they should say in response to seeing a pictured item, avoiding the production of errors may rely more on language output, where children benefit from practice producing the target words. Factors other than language exposure may also contribute to error frequency, including lexical priming (i.e., sociolinguistic account; Gibson, Oller, Jarmulowicz, & Ethington, 2012) and/or individual differences in processes related to picture naming (e.g., attention, suppression, working memory).

Regardless of language ability, we found a significant semantic blocking effect—where all children (TD and DLD combined) were more accurate and produced fewer errors in the Mixed-category as compared with the Same-category context. Thus, the semantic blocking effect, where lexical access and retrieval is hindered in the context of semantically-related items, occurs for all children despite differences in their lexical quality, where it is presumed that children with DLD have impoverished semantic knowledge. Intriguingly, we also found a significant interaction between task language and semantic context, indicating that the semantic context manipulation produced greater overall difficulty for the children in Spanish, their first language, than in English, their second. This interaction might be explained by assuming that children who are developing two lexicons have stronger lexical-semantic links in their first language than in their second (cf. Kroll & Stewart, 1994), as a result of their greater experience with it (cf. Gollan, Montoya, Cera, & Sandoval, 2008, and Gollan et al., 2011's Frequency Lag hypothesis). However, considering the experiment design, we remain cautious about generalizing from this cross-linguistic difference. Overall, our results pertaining to the main effect of semantic blocking are consistent with findings from investigations of monolingual children (e.g., Boelens & La Heij, 2017) and adults (e.g., Oppenheim, Dell, & Schwartz, 2010; Schnur, et al., 2006).

### Quantifying qualitative information: Evaluating error types

Although standardized and nonstandardized tests evaluating children's language ability are scored based on accuracy, tests often allow clinicians to credit the test item when the child self-corrects. For example, in response to being shown a picture of an apple, the child said, "banana… I mean apple," a test may allow the clinician score the response as correct, despite children's initial production of an error. Importantly, this type of production provides qualitative information about the child's ability to access, select, and retrieve lexical information, as well as their higher level cognitive abilities (i.e., self-monitoring). Because of this, we should attend to the types of error responses that the child produces and not just the child's overall final accuracy. By quantifying such qualitative data, we can discern patterns that point to break-downs in lexical processing and potentially language impairment.

Our prediction that children with DLD would produce quantitatively more and qualitatively different errors than their TD peers was accurate; however, this was dependent upon the task language. Children with DLD produced significantly more intrusions and semantic errors than their TD peers in Spanish; no significant differences emerged across ability groups in English. While differences by ability group were not evident in English, all children (both TD and DLD) were in the process of acquiring their second language (English). The emergence of differences in naming ability in their first language (Spanish) may reflect the underlying stability of their Spanish knowledge, while English word knowledge would be relatively less robust. Increased production of the semantic and intrusion error types by children with DLD points to potentially unrefined, sparser semantic networks inherent to deficient language ability. Deficient semantic knowledge would result in difficulty accessing and/or retrieving the lexical item associated with the target. Children with DLD continue to demonstrate poorer accuracy on semantic tasks in comparison to their TD peers, even after a conceptual scoring approach—as opposed to a single language scoring approach—is applied (e.g., Sheng, Peña, Bedore, & Fiestas, 2012). This finding, in conjunction with the types of errors produced on our task, support a position that sparse semantic networks are characteristic of DLD in all children, regardless of the numbers of languages they are learning and potentially constrain the spread of activation during lexical access and retrieval.

Another position, described by Martin, Lesch, and Bartha, (1999), is that intrusions can result from delayed decay of activation, increasing the activation of recently accessed competitors. In relation to our findings, children with DLD may lack efficient decay in Spanish, resulting in more intrusions or semantic errors compared with their TD peers. However, even in Spanish we did not observe substantial semantic context effects for intrusions that would support this explanation (see Schnur et al., 2006, for discussion distinguishing predictions of overactivation versus underactivation accounts of errors in blocked cyclic naming).

In line with studies of error production patterns in monolingual children with and without DLD (Lahey & Edwards, 1999; McGregor, Newman, Reilly, & Capone, 2002), we also found that bilingual children with DLD produced significantly more phonological and no response errors than their TD peers in both languages. Descriptively, children from both groups in our sample produced a large number of phonological errors across languages and contexts. These types of errors point to unrefined phonological representations and do not necessarily indicate impaired language.

Although gaps in lexical-semantic information would impose insufficient constraint for accurate lexical activation and selection, phonological encoding errors may also occur. For example, if a child with DLD produced a rhyming error such as "hook" for the pictured item *book*, the substitution of /h/ for /b/ changes the semantic meaning. Alternatively, children with DLD might be slower to form refined phonological representations of words. Fuzzy phonological representations of this type, are likely to increase sound-based error production on naming tasks. For instance, Bedore, Peña, & Boerger (2010), noted that Spanish-English bilinguals produced errors such as "bark" for *barco* (boat) and "rinocornio" for *rhinoceros* during a naming task. The phonological similarities between the errors and target names suggest that the children are using their semantic knowledge but experience difficulty with accurate retrieval of the phonological form. Inaccurate storage and/or fuzzy representations would likely result in inaccurate (or inefficient) word retrieval. In our study, phonological errors predominated for both groups of children, demonstrating the challenge of accurately retrieving the phonological form of lexical entries. For children learning two languages, underspecified phonological representations could be secondary to limited language exposure in either the first or second language, a deficit inherent to DLD, or both. Further investigation is required to tease the source of these errors apart.

Although the participants in this study attended schools with bilingual educational programs, English was often the dominant language in classrooms. In an environment where children sense that one language is dominant, or subtly preferred, the lexicon of that language (i.e., English) might be primed. Additionally, according to Green's (1998) Inhibitory Control model, a suppression mechanism for the non-target language (i.e., Spanish) could facilitate access to English lexical-semantic information relative over Spanish. This sociolinguistic account has been explored in the context of vocabulary knowledge of school-aged bilingual children (Gibson, Oller, Jarmulowicz, & Ethington 2012). In the current study, relatively higher English activation combined with suppression of the Spanish lexicon could interfere with or impede lexical retrieval for both groups of children.

Similarly to the results for accuracy and error frequency, a semantic context effect was observed for the semantic error types across languages, where children produced significantly more in the Same-category context than the Mixed-category context. Regardless of language ability, children had difficulty with lexical retrieval secondary to increased lexical competition for semantically-related items. Additionally, in Spanish no responses were significantly more frequent in the Same-category context. This demonstrates that children—regardless of language ability—experienced more difficulty accessing the target lexical item for the picture when semantically-related pictures are presented. Words included on this blocked cyclic naming task were specifically chosen because of their high frequency and high agreement in each language; thus, children would have these words stored in each of their lexicons (i.e., translation equivalents), and naming errors point to breakdowns in the lexical retrieval process.

**Limitations and Future Directions**

In this study, we used different stimuli in each language. This minimized the opportunity for cross-language transfer or interference, but made it more difficult to balance other constraints when choosing our stimuli. For instance, we cannot rule out the possibility that the Spanish stimuli were more semantically similar than the English stimuli. Therefore, it seems important replicate our apparent cross-linguistic differences in semantic context effects with other materials. On a related point, although laboratory studies of blocked cyclic naming often focus on how the semantic context effect increases with repetition (e.g. Oppenheim et al, 2010; Schnur et al., 2006; Fink, Oppenheim, & Goldrick, 2018), we did not consider repetition in this analysis.

Another limitation of our stimuli is the inclusion of some cognates, which may have been named more easily than non-cognates. Cognates were included because they were highly familiar words with high convergence for naming in young Spanish-English bilinguals. The ability to name cognates in one language is strongly related to bilingual children's accuracy for naming the cognate in another language regardless of language ability (TD or DLD; Grasso, Peña, Bedore, Hixon, & Griffin, 2018). Additionally, bilinguals demonstrate an advantage for cognate production during un-timed naming tasks without a repeated design (Leacox, Wood, Sunderman, & Schatschneider, 2016). Evaluating potential facilitation effects for naming cognates using a timed picture naming task with a repeated design, such as semantic blocking, would be another avenue of investigation to clarify cross-language transfer, incremental learning, and/or lexical competition in developing bilinguals.

**Clinical Implications**

Clinically, speech-language pathologists should attend to the number and types of errors that bilingual children produce during assessments, because these errors may be indicative of deficient lexical processing efficiency and potentially impoverished semantic knowledge. Targeting the development of semantic depth during intervention may decrease error production in bilingual children with DLD. Further investigation of error production in this context is needed to support this proposal.

While our findings support the position that a characteristic of DLD is deficient semantic networks, not all children with DLD appear to have deficits in vocabulary knowledge on force-choice recognition tests (see Gray, Plante, Vance, & Heinrichsen, 1999, for examples with monolingual children). Clinically, grammatical structures may be better indicators of potential DLD, particularly for bilingual children with a range of second language experience (see Bedore, Peña, Anaya, Nieto, Lugo-Neris, & Baron, 2018). While patterns of error production may be easily perceptible for clinicians (i.e., semantic errors), not all error types necessarily point to impaired language (i.e., phonological errors). Careful consideration of the interpretation for the frequency and types of errors is warranted when disentangling deficit versus disorder for bilingual children with varying degrees of language exposure.

**Conclusion**

For bilingual children with DLD, the observed errors point to lexical processing deficits that extend beyond those exhibited in typical development. Monolinguals (e.g., McGregor, Newman, Reilly, & Capone, 2002) and bilinguals with DLD produce a greater number and, more importantly, different error types than their TD peers. Our finding support that this remains true even after accounting for language exposure. Robust word representations within a rich, interconnected semantic network support efficient word retrieval for naming. Because children with DLD have weaker word representations due to impoverished semantic networks, insufficient semantic knowledge may cause the diffuse spread of activation to nontarget lexical entries. This impediment results in strong activation of several lexical alternatives, increasing the chance of error production during picture naming.

In our study, patterns in the quantity and types of errors that occurred across both languages demonstrated that—even after accounting for language exposure—qualitative differences emerge in picture naming for children with DLD relative to their TD peers. Specifically, those with DLD were less accurate and produced more errors than their TD peers. Additionally, children with DLD produced more intrusion and semantic errors than their TD peers in Spanish. We believe that the difference between these groups' error patterns is therefore not due to second language exposure, but rather deficits inherent to DLD, which inhibits lexical-semantic processing and organization in their first language. Given that this pattern did not occur in their second language, English, more research in this line of inquiry is needed to understand the link between error production, semantic knowledge, and breakdowns in lexical processing in bilingual children with and without DLD. Future studies should evaluate the relationships between error production and semantic knowledge to clarify whether or not children with DLD experience a lack of semantic constraint based on their limited semantic knowledge. Additionally, investigating individual changes over time may inform theoretical and practical information on individual differences in processing ability. As linguistic knowledge is constructed through cultural experiences, understanding how differences in language ability and language exposure impact lexical processing in children developing multiple languages will lead to a broader understanding as to why a child might name a picture of a *strawberry* as "cherry".

# References


Alario, F. X., & del Prado Martín, F. M. (2010). On the origin of the "cumulative semantic inhibition" effect. *Memory & Cognition*, *38*(1), 57-66.

Archibald, L. M. (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.

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.

Bedore, L. M., Peña, E. D., & Boerger, K. (2010). Ways to words: Learning a second language vocabulary. In M. Schatz & L. Wilkinson (Eds.), *The education of English language learners: Research to practice* (pp. 87-107). New York, NY: Guildford Press.

Bedore, L. M., Pena, 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.

Belke, E. (2013). Long-lasting inhibitory semantic context effects on object naming are necessarily conceptually mediated: Implications for models of lexical-semantic encoding. *Journal of Memory and Language*, *69*(3), 228–256. https://doi.org/10.1016/j.jml.2013.05.008

Belke, E. (2017). The Role of Task-Specific Response Strategies in Blocked-Cyclic Naming. *Frontiers in Psychology*, *7*, 1955. https://doi.org/10.3389/fpsyg.2016.01955

Boelens, H., & La Heij, W. (2017). The development of semantic blocking in children. *The British Journal of Developmental Psychology*, *35*(2), 310–315. https://doi.org/10.1111/bjdp.12178

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.

Brackenbury, T., & Pye, C. (2005). Semantic deficits in children with language impairments. *Language, Speech, and Hearing Services in Schools*, 36(1), 5-16. https://doi.org/10.1044/0161-1461(2005/002)

Charest, M. (2017). Cumulative semantic interference in young children's picture naming. *Applied Psycholinguistics*, *38*(04), 835-853. https://doi.org/10.1017/S0142716416000461

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-428. https://doi.org/10.1037//0033-295X.82.6.407

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language*, *45*(4), 721-736.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283-1296.

Damian, M. F., Vigliocco, G., & Levelt, W. J. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, *81*(3), 77-86. https://doi.org/10.1016/S0010-0277(01)00135-4

Degani, T., Kreiser, V., & Novogrodsky, R. (2019). The joint effects of bilingualism, DLD and item frequency on children's lexical-retrieval performance. *International Journal of Language & Communication Disorders*, *54*(3), 485-48. https://doi.org/10.1111/1460-6984.12454

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1-3), 287-314.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801-838.

Fink, A., Oppenheim, G. M., & Goldrick, M. (2018). Interactions between lexical access and articulation. *Language, Cognition and Neuroscience*, *33*(1), 12-24.

Gibson, T. A., Oller, D. K., Jarmulowicz, L., & Ethington, C. A. (2012). The receptive–expressive gap in the vocabulary of young second-language learners: Robustness and possible mechanisms. *Bilingualism: Language and Cognition*, *15*(1), 102-116.

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).* Unpublished test.

Gollan, T. H., & Acenas, L. A. R. (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*(1), 246-269. http://dx.doi.org/10.1037/0278-7393.30.1.246

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787-814.

Gollan, T. H., 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: General*, *140*(2), 186-209. http://dx.doi.org/10.1037/a0022256

Grasso, S. M., Peña, E. D., Bedore, L. M., Hixon, J. G., & Griffin, Z. M. (2018). Cross-linguistic cognate production in Spanish–English bilingual children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *61*(3), 619-633.

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.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67-81.

Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology*, *23*(4), 574-586.

Harvey, D. Y., & Schnur, T. T. (2016). Different Loci of Semantic Interference in Picture Naming vs. Word-Picture Matching Tasks. *Frontiers in Psychology*, *7*, 710. https://doi.org/10.3389/fpsyg.2016.00710

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.

Kail, R., & Leonard, L. B. (1986). Sources of Word-Finding Problems in Language. *Handbook of cognitive, social, and neuropsychological aspects of learning disabilities*, *1*, 185-202.

Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: a meta-analysis. *Journal of Speech, Language, and Hearing Research*, *53*(3), 739-756. https://doi.org/10.1044/1092-4388(2009/08-0248)

Kroll, J. F., Dijkstra, A., Janssen, N., & Schriefers, H. (2000, November). Selecting the language in which to speak: Experiments on lexical access in bilingual production. In *41st annual meeting of the Psychonomic Society, New Orleans, LA*.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149-174.

Lahey, M., & Edwards, J. (1996). Why do children with specific language impairment name pictures more slowly than their peers? *Journal of Speech and Hearing Research*, *39*(5), 1081-1098. https://doi.org/10.1044/jshr.3905.1081

Lahey, M., & Jan, E. (1999). Naming errors of children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *42*, 195-205.

Ladányi, E., & Lukács, Á. (2016). Lexical conflict resolution in children with specific language impairment. *Journal of Communication Disorders*, *61*, 119-130.

Leacox, L., Wood, C., Sunderman, G., & Schatschneider, C. (2016). Young Spanish--English Language Learners' Cognate Facilitation on Picture Naming. *Contemporary Issues in Communication Science & Disorders*, *43*, 115-128.

Leonard, L. B. (2014). *Children with specific language impairment* ($2^{nd}$ ed.). Cambridge, MA: The MIT press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1-38.

Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, *41*(1), 3-29.

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.

McGregor, K. K. (1997). The nature of word-finding errors of preschoolers with and without word-finding deficits. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1232-1244.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2007). Cumulative semantic interference as learning. *Brain and Language*, *103*(1-2), 175-176. https://doi.org/10.1016/j.bandl.2007.07.102

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227-252. https://doi.org/10.1016/j.cognition.2009.09.007

Oppenheim, G. M., Griffin, Z. M., Peña, E. D., & Bedore, L. M. (2019). *Longitudinal evidence for simultaneous bilingual language development with shifting language dominance, and how to explain it.* Manuscript submitted for publication.

Oppenheim, G., Wu, Y. J., & Thierry, G. (2018). Found in translation: Late bilinguals do automatically activate their native language when they are not using it. *Cognitive science*, *42*(5), 1700-1713.

Peña, E. D., Bedore, L. M., & Griffin, Z. (2010). Cross-Language Outcomes of Typical and Atypical Development in Bilinguals: National Institute of Deafness and other Communication Disorders.

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., Gutiérrez-Clellen, V. F., Iglesias, A., & Goldstein, B. A. (2006). *Bilingual English Spanish Oral Screener (BESOS).*

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

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, *15*(1), 36-57.

Schnur, T., Schwartz, M., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*(2), 199-227. https://doi.org/10.1016/j.jml.2005.10.002

Sheng, L. (2014). Lexical-semantic skills in bilingual children who are becoming English-dominant: A longitudinal study. *Bilingualism: Language and Cognition*, *17*(3), 556-571.

Sheng, L., & McGregor, K. K. (2010). Lexical–semantic organization in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *53*(1), 14-159. https://doi.org/10.1044/1092-4388(2009/08-0160)

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)

Sheng, L., Bedore, L. M., Peña, E. D., & Fiestas, C. (2013). Semantic development in Spanish–English bilingual children: Effects of age and language experience. *Child Development*, *84*(3), 1034-1045.

Snyder, H. R., & Munakata, Y. (2013). So many options, so little control: abstract representations can reduce selection demands to increase children's self-directed flexibility. *Journal of Experimental Child Psychology*, *116*(3), 659-673. https://doi.org/10.1016/j.jecp.2013.07.010

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1245-1260.

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition*, *85*, B61-B69.

# Appendix A

Table 4: English items used in the semantic blocking task.

| | Animals | Fruits & Vegetables | Food | Clothing |
|---|---|---|---|---|
| Set 1 | Penguin | Orange | Pizza | Shoe |
| Set 2 | Elephant | Pumpkin | Donut | Pajamas |
| Set 3 | Snake | Banana | Fish | Hat |
| Set 4 | Dog | Apple | Eggs | Socks |

Table 5: Spanish items used in the semantic blocking task.

| | Animals | Fruits & Vegetables | Furniture | Vehicles |
|---|---|---|---|---|
| Set 1 | Tortuga | Sandia | Cama | Motocicleta |
| Set 2 | Caballo | Tomate | Espejo | Avion |
| Set 3 | Oso | Fresa | Silla | Bicicleta |
| Set 4 | Gato | Uvas | Lampara | Tren |

# Appendix B

Table 6: Descriptive data for the number of items double-coded.

| Error Code Type | Number Double-Coded | Percent |
|---|---|---|
| Phonological only | 572 | 8.72% |
| Semantic-Phonological | 491 | 7.48% |
| Phonological-Miscellaneous | 3189 | 48.59% |
| Semantic-Miscellaneous | 817 | 12.45% |
| Miscellaneous only | 1366 | 20.81% |
| Phonological-Semantic-Miscellaneous | 128 | 1.95% |
| Total number of items double-coded | 6563 | |

# Appendix C

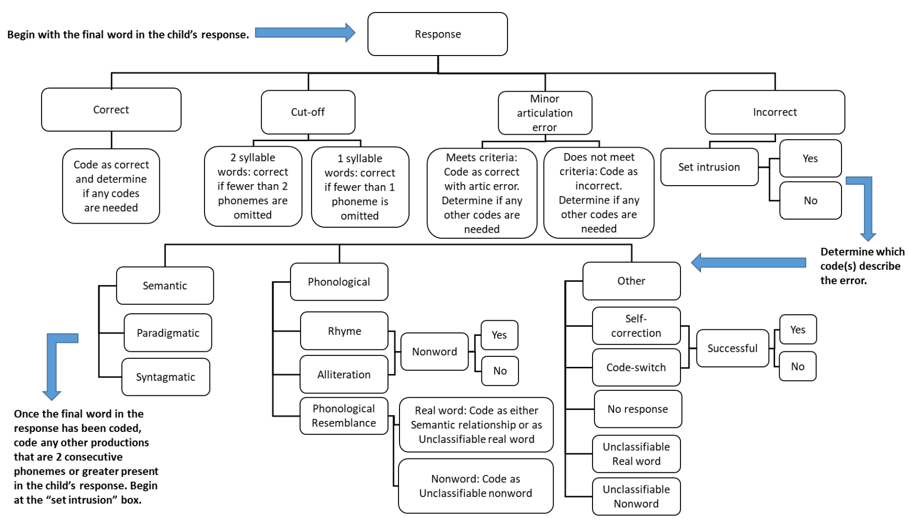


Figure 4: Coding Scheme.