AJSLP

Research Article

Insights Into Category Sorting Flexibility in Bilingual Children: Results of a Cognitive Lab Study

Gabriela Simon-Cereijido,[a] Lisa M. Bedore,[b] Elizabeth D. Peña,[c] and Aquiles Iglesias[d]

Purpose: The purpose of this study was to explore how bilingual children shift sets to gain flexibility when forming categories. Using a cognitive lab approach focused on understanding how learners approach problems, we asked children to sort 10 sets of pictures representing common objects in two different ways and to explain their rationale for the sort. We explored the relationship between age and language use on their performance.
Method: Forty-six typically developing Spanish–English bilingual children (25 girls, 21 boys) participated in the study. They ranged in age from 4;0 to 10;11 (years;months). Receptive and expressive responses to a novel category sorting task were collected.
Results: Forty-four of the 46 children tested were able to perform the category sorting task. Within language, receptive and expressive category sorting scores were positively and significantly correlated while only expressive scores were significantly associated across languages. There were significant correlations between the sorting scores and age and language output and input. Children's ability to provide expressive responses explaining their sort strategy was moderately correlated with their language experience, especially English output.
Conclusions: The category sorting task proved useful in eliciting sorting behaviors and naming from the children tested. The age effect suggests that sorting may reflect their general developmental experience rather than their language-specific experience. The cognitive lab approach allowed us to understand how children shift sets and verbalize their understanding of the categorization process. Knowing how children approach this task can inform future work to develop ways to strategically select language intervention goals and document progress.

The past 20 years have seen gains in our understanding of developmental language disorder (DLD) in bilingual children. A key effort in this regard has been to find procedures that can be employed to accurately identify DLD in bilingual children. As part of this work, we have identified language forms and behaviors that are especially challenging for children with DLD, are appropriate for bilingual children, and serve as clinical markers in assessment protocols (Bedore et al., 2018; Gutiérrez-Clellen & Simon-Cereijido, 2007; Sheng et al., 2012; Simon-Cereijido & Gutiérrez-Clellen, 2007). Clinical markers are forms or items that children with DLD are highly likely to fail and children with typical development are likely to pass (e.g., Bedore & Leonard, 1998). In the domain of grammar, many of the same forms that are difficult for monolingual children serve as clinical markers for bilingual children in each of their languages (Gutiérrez-Clellen & Simon-Cereijido, 2007; Simon-Cereijido & Gutiérrez-Clellen, 2007). Similarly, in the domain of semantics, tasks that challenge the breadth and depth of children's vocabulary are good markers of impairment (Peña et al., 2016). Such tasks include category generation, definitions, identification of characteristic properties, and functions of objects.

Building on this foundation, several assessment procedures have been developed and tested for Spanish–English bilinguals. These include standardized tests (Peña et al., 2018), language sampling (Bedore et al., 2010; Kapantzoglou et al., 2017), and dynamic assessment (Kapantzoglou et al.,

[a]Department of Communication Disorders, Rongxiang Xu College of Health and Human Services, California State University, Los Angeles
[b]Department of Communication Sciences and Disorders, Temple University, Philadelphia, PA
[c]School of Education, University of California, Irvine
[d]Department of Communication Sciences and Disorders, University of Delaware, Newark
Correspondence to Gabriela Simon-Cereijido: gsimonc@calstatela.edu

Editor-in-Chief: Julie Barkmeier-Kraemer
Editor: Mark Guiberson
Received July 3, 2019
Revision received October 11, 2019
Accepted December 9, 2019
https://doi.org/10.1044/2019_AJSLP-19-00008
Publisher Note: This article is part of the Forum: Innovations in Clinical Practice for Dual Language Learners, Part 2.

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

2012; Peña et al., 2008). Yet, for the gains that we have made in the domain of assessment to be applied to intervention, we must consider how these results inform selection of intervetion targets. Specific consideration should be given to the identification of appropriate intevention targets and for ways to monitor progress. Learning to use an evidence-based approach to intervention targets will support those children we identify with DLD and support amelioration of the consequences of language impairment. Here, we consider how what we have learned about DLD in one domain of language—semantics—can be leveraged to develop a means of selecting intervention targets that are appropriate for bilingual children in the future. In particular, we explore how bilingual children are able to shift sets to identify and define categories in both languages. The executive functions needed to do this task are documented weaknesses (e.g., Kapa & Plante, 2015) in children with autism and language development. However, any bilingual advantage associated with task switching seem unlikely to carry over to daily activities that are like the tasks presented here (Gonzalez-Barrero & Nadig, 2019).

In the United States today, 22% of children are English language learners, and the majority of those children (80%) speak Spanish at home (Uro & Lai, 2019). In states such as California, Texas, New Mexico, and New Jersey—some of the most diverse states—the Hispanic population ranges from 20% to nearly 44% of the population. However, a mere 5% of speech-language pathologists in the field report having training in bilingualism and fluency in a language other than English; only 5.8% of clinicians report fluency in Spanish (American Speech-Language-Hearing Association [ASHA], 2014). Recent graduates of speech-language pathology programs are well aware that they need to take into account both of the child's languages as they make treatment decisions (ASHA, 2014). Still, 20% of the respondents in the ASHA survey who had obtained their terminal degrees within the last 15 years identified lack of knowledge about assessing and treating bilingual children as one of their greatest challenges in the work setting. A recent survey found that only 38% of school-based speech-language pathologists reported that they had sufficient training to prepare them to assess English language learner students (Arias & Friberg, 2017). Clinicians report that when they work with bilingual children, they often default to what they know about English language development in monolinguals to guide treatment plans (Thordardottir, 2010; Williams & McLeod, 2012). To reduce the health disparities associated with DLD in bilingual children, it is critical that we develop materials that will support clinical decision making for bilinguals. To facilitate gains in our ability to provide interventions for bilingual children, it is crucial that we study how bilingual children, especially those that we consider successful bilingual learners, perform in each of the languages across the school years.

To date, the most effective interventions identified for children with DLD are those that are manualized (Bleses et al., 2018) and directly target linguistic deficits (R. B. Gillam et al., 2008). Much of the intervention for children with DLD focuses on vocabulary development for both English speakers (Crystal, 1998) and bilingual Spanish–English speakers (Restrepo et al., 2013; Simon-Cereijido & Gutiérrez-Clellen, 2014). Effective interventions are those that utilize integrative approaches such as book-based treatment so that children can have opportunities to build language skills in context (S. L. Gillam et al., 2012; Justice et al., 2005). An integrated approach targets vocabulary, grammatical morphology, complex syntax, phonological awareness, and narrative skills. In the vocabulary domain, children need to develop deep semantic mappings that will enhance their selection and retrieval of appropriate words in context. For bilingual children, the most effective intervention outcomes are observed when targets reflect skills needed to support bilingual development, regardless of the language in which intervention is delivered (Thordardottir et al., 2015). For monolingual, English-speaking children, clinicians are guided by a rich understanding of language development sequences that inform their understanding of how linguistic targets are mastered (Owen Van Horne & Green Fager, 2015), but most clinicians lack this knowledge for bilingual children.

When clinicians are uncertain about the next steps, they may turn to the item types that challenge children in language assessments as possible intervention targets. This practice has been characterized as "symptom-driven practice" (Selin et al., 2019); clinicians appear to excessively rely on standardized testing information rather than on the communicative and linguistic symptomology and functions. Items on standardized tests are not intended for use as the basis of intervention targets. This is because tests often focus on clinical markers—specific linguistic forms that children with DLD are unlikely to produce. For example, in Spanish, only singular masculine article use (i.e., *el*) reliably differentiated Spanish–English children with DLD, and these were thus included on the Bilingual English Spanish Assessment. However in Spanish, children need to produce articles contrasting gender and number (i.e., *el, la, los, las*). Thus, while clinical markers are useful for identifying DLD, intervention goals are broader and require attention to patterns of mastery and nonmastery. A review of evidence-based research shows that goals associated with these approaches are not the markers themselves but skills that support broader changes in communication (McCauley et al., 2017).

Here, we will turn our attention to understanding how linguistic targets for intervention are mastered by bilingual children with the long-term goal in mind of developing criterion-referenced probes of language development. Criterion-referenced probes could serve to quantify our understanding of children's knowledge in order to identify intervention targets that they can use to make meaningful contrasts to express a full range of communicative functions.

A starting point for developing such criterion-referenced measures is the identification of functional tasks that reflect children's linguistic and academic needs as identified on language assessments. One such area is children's ability to organize and categorize information. This knowledge is reflected in category generation or semantic fluency tasks

on a number of standardized tests in which children are asked to identify as many items as they can in a category or talk about which items go together or belong in a category (e.g., Clinical Evaluation of Language Fundamentals; Wiig et al., 2013). This task is robust for children who are bilingual, because it can draw on their knowledge of both languages and it is included as an item type on the Bilingual English Spanish Assessment (Peña et al., 2018). Lack of semantic breadth and depth are central to vocabulary deficits in bilingual children with DLD (Peña et al., 2016; Sheng et al., 2012). Children with DLD have much more difficulty with category knowledge as evidenced by their difficulties with category generation and repeated associations tasks. To improve our clinical ability to provide interventions for bilingual children, it is crucial that we study how bilingual children, especially those that we consider successful bilingual learners, perform in each of the languages across the school years.

Outside the context of the standardized testing, there is developmental information about children's ability to organize information and how this relates to the size of their vocabulary (Nelson, 1993; Peña et al., 2003). The revised hierarchical model suggests that bilingual children must associate their conceptual base with first and second language–specific vocabulary in order to build lexical semantic knowledge within and across languages (Kroll et al., 2010; Potter et al., 1984). By drawing on conceptual knowledge across language-specific lexical items, children increase options for category sorting. As links across languages and with their conceptual base increase in strength, flexibility and vocabulary depth emerge. The links across conceptual, semantic, and lexical items within and across languages develop over time and are influenced by linguistic experience. For example, in a study of 778 Spanish–English typically developing children screened for prekindergarten, the gap between receptive and expressive language skills in each language was predicted by the current language experience in the language (Gibson et al., 2014). Studies with older bilingual participants also found a receptive–expressive gap in lexical and semantic tasks within each language associated to the linguistic experience. Interestingly, some studies found a larger gap in the first language and others in the second language (see Gibson et al., 2014).

To translate this common assessment task into a task that would support children's linguistic competency, we considered what the corresponding linguistic skill might be as a potential intervention target. While we are rarely asked to list items in categories in daily conversations, we often have the need to categorize objects. In academic settings, learners need to be able to shift sets in order to flexibly access words and concepts in different contexts (topics, academic subjects). That is, learners must have the cognitive flexibility to shift sets in order to organize ideas, concepts, and words in different ways according to functional and concrete needs and to abstract demands. Daily examples of the need to categorize in a flexible manner might include sorting laundry (e.g., clean and dirty), in addition to sorting them into winter and summer clothes for storage. We may need to store categories of foods (e.g., fruits and vegetables) in different locations and also sort them as healthy or not healthy when we plan a meal. In the academic context, children might be initially taught about animals that live in different environments (water, land), then refine categories (farms, jungles, forests), and then regroup based on higher level taxonomies (e.g., reptiles, mammals, canines, felines). Multiple exemplars and experiences in different cultural and linguistic contexts may influence category sorting quality (taxonomic vs. slot filler), so it is important to study how children approach this task. Because this skill builds on children's language experience across multiple contexts, we expect that the performance of bilingual and monolingual children would be similar and dependent on their total language knowledge.

A task that can be used to address this skill is a category sorting task in which children are asked to sort items or pictures in multiple ways and to verbalize how the items are sorted. This task requires in-depth knowledge of items and categories. For example, children may be asked to sort pictures into the categories of fruits and vegetables (e.g., orange and banana vs. onion and cucumber), with a set of pictures that can also be sorted by shape (e.g., round items: orange and onion vs. long items: banana and cucumber). The beauty of this task is that it allows children to draw on concepts to which they have been exposed in multiple contexts within and across their languages and thus provides insight into organization rather than knowledge of specific semantic classes of items.

In the early stages of test development, a cognitive lab study can be employed to gain an understanding of how children might perceive a given task. This approach uses clinical interviewing to gain insights about a child's approach to a task (Brédart et al., 2014). After completing a task, the interviewer asks children to verbalize their reason for responding on the way they did. Follow-up questions might ask children to make judgments about whether there are other possible responses or how another child who is younger might respond. This provides developers with a range of possible responses. Also, these data are analyzed qualitatively to evaluate whether children's responses align with the intended nature of the construct.

With the goal of developing a functional task that could inform intervention, we undertook a cognitive lab study focusing on how bilingual Spanish–English-speaking children ages 4–10 years create semantic categories by shifting sets. By asking children to sort common items from prototypical categories without providing semantic cues as to the intended sort strategy, we hope to gain insights into children's knowledge of semantic categories. A qualitative and quantitative analysis of the children's responses will inform the cognitive and semantic strategies (e.g., perceptual and/or representational similarity) used across languages and ages.

Method

Participants

A total of 46 bilingual Spanish–English children participated in the study. The study had institutional review

board approvals from the three participating institutions (The University of Texas at Austin, University of Delaware, and California State University, Los Angeles). Parents gave informed consent, and children verbally assented to participate before taking part in the study. Children's age ranged from 4;0 to 10;11 (years;months); the mean age was 8;1 (SD = 22 months). Two families did not return the parent questionnaire, and thus, demographic data are partially complete. There were 25 girls and 21 boys. All children were reported to be Latino/a and lived with Spanish-speaking family members. Parents and children were assessed in a variety of locations, including community settings (e.g., Sunday school classrooms), homes, and university labs.

Inclusion criteria required child exposure and use of Spanish and English, without specifying a predetermined level. Participating children were exposed to and used English and Spanish at different levels. Parents completed the Bilingual Input Output Survey (BIOS) linguistic questionnaires (Peña et al., 2018), based on which input and output percentages and proficiency levels were calculated (see Table 1). English input ranged between 25% and 83%, whereas English output range was from 26% to 100%. For Spanish, input ranged from 17% to 75%, whereas output ranged from 0% to 74%.

Children completed the Bilingual English Spanish Oral Screener (BESOS; Peña et al., 2008), and both English and Spanish skills were assessed in the areas of semantics and morphosyntax. Children were expected to pass Morphosyntax and Semantics subtests in at least one of their languages in order to participate in the study. Children who failed two Morphosyntax and/or two Semantic subtests were excluded from the study.

Table 1. Means and standard deviations of participants' age and linguistic information, and outcome scores.

Variable	*M (SD)*
Age and linguistic information	
Chronological age in months	97 (22)
Spanish input	41.50 (16.17)
English input	58.50 (16.17)
Spanish output	37.83 (18.74)
English output	62.17 (18.74)
BESOS English Semantics	114.54 (32.72)
BESOS English Morphosyntax	102.18 (22.02)
BESOS Spanish Semantics	108.57 (26.11)
BESOS Spanish Morphosyntax	84.30 (26.30)
Outcomes	
Receptive category sorting	0.47 (0.21)
English receptive category sorting	0.40 (0.32)
Spanish receptive category sorting	0.45 (0.23)
Expressive category sorting	8.22 (4.14)
English expressive category sorting	4.02 (2.41)
Spanish expressive category sorting	4.20 (2.06)

Note. The average standard score is 100 for all subtests. BESOS = Bilingual English Spanish Oral Screener.

Measures

BIOS

The BIOS is used as an interview to document hour-by-hour exposure to English and Spanish. Parents are asked to report children's activities and the language used by the child and interlocutor for a typical weekday and a weekend day. These are projected out to calculate the percentage of Spanish and English the child hears and uses during a typical week.

BESOS

The BESOS consists of two subtests each in two languages. Semantics is composed of items that test knowledge of semantic relationships such as functions, associations, and definitions. Both expressive and receptive items are included. Morphosyntax targets items that are difficult for children with DLD in the target language (e.g., past tense in English and direct object clitics in Spanish). Morphosyntax is elicited through sentence repetition and cloze items in each language. There are different versions for preschool and kindergarten, first and second grades, and third and fourth grades (with approximately 20%–30% of overlap among items). Raw scores are converted to standard scores. Comparing between languages, we use the best score for each subtest in screening. Using a cutscore of −1 SD from the mean, sensitivity ranges from 80% to 93%, and specificity ranges from 90% to 94%, depending on age. The group means and standard deviations for the four subtests are listed in Table 1.

Procedure

Experimental Category Sorting Task

Children completed an experimental novel category sorting task. They were required to sort a set of pictures into two groups of approximately the same size and explain the category supporting the sorting. Then, the children were asked to do it again by sorting the pictures into two different groups. After doing this, they were asked to explain the category of the two new groups. No cues were given as to the intended categories or sort strategies.

The sets were developed using a dual-focus approach (Erkut et al., 1999). This approach is used to guard against cultural and linguistic bias from translation in only one direction. Half of the sets were conceived in English first, and then the target categories and item words were translated into Spanish. The other half was developed in Spanish and then translated into English. A team of experts provided feedback on the appropriateness of the semantic categories, pictures, and target vocabulary. Ten sets were created in total. Due to its experimental and novel nature and depending on the categories, each set of pictures included eight, nine, or 10 pictures. For example, the first set included animals belonging to the taxonomic categories of mammals and reptiles. These same animals could also be sorted by habitat (i.e., land or water). The set included two land mammals, two aquatic mammals, two land reptiles, and two aquatic reptiles. For all sets, the task pictures were

colored drawings, presented in pictures sized 3 in. × 3 in., positioned on a grid.

The 10 experimental sets in English and 10 in Spanish were distributed across three versions of the category sorting task, and each version had a total of 10 sets. These shorter versions allowed us to evaluate the 20 available sets across children and prevent children's fatigue or lack of engagement. Each child completed one of the three versions. Children were presented with five sets in English and five in Spanish. In every version, there were two overlapping sets that were administered in both English and Spanish. Thus, children were presented with three sets in English only, three sets in Spanish only, and two sets in both languages (the same set delivered in English and a second time in Spanish). English sets were always presented before Spanish sets. Children heard the following instructions before each sort: "Show me how these things go in two groups. [*Pon estos dibujos en dos grupos.*]" and "Show me how these things go in two different groups. [*Pon estos dibujos en dos grupos diferentes.*]". The grouping sorts were considered receptive responses. Immediately after sorting the pictures, children were presented with the following prompt: "Tell me how these things go together [*Dime cómo estos dibujos van juntos en estos grupos*]". Children's explanations were the expressive responses. The instructions were not literal translations but adaptations of the same prompt reached by consensus across sites. They contain different key words that were selected, taking into consideration language-specific word frequencies. For example, the word "show [*enseñar*]" has a relative higher frequency than "put [*poner*]" in English than in Spanish (Davies, 2008, 2016). These instructions were also vetted relative to the forms that children would be likely to hear in Spanish at home and at school.

Two practice or demonstration sets were presented to all children. The first demonstration set asked children to sort triangles and circles of two different colors in two different ways—by shape and by color. Similarly, the second set asked children to sort numbers and letters of two different colors in two different ways. If a child had difficulties with this task, the evaluators modeled the sorting and the explanations to them. Most school-age children found the task easy and appeared to quickly grasp the task. Some of the youngest children (4 years of age) required support.

Bilingual research assistants administered and recorded the children's responses in paper and in audio files. Audio recordings were reviewed in order to corroborate written notes. Responses were then entered into a database. The bilingual assistants scored the task under the supervision of the project's investigators.

Two main scores were calculated from these data. First, a "receptive category sorting score" consisted of the proportion of sorted pictures that matched or "hit" the set target categories. For example, a child received full points for the fruits and vegetables set if he sorted the eight pictures by fruits and vegetables in one sort in addition to by shape (four round and four long objects) in another sort. If the child sorted the pictures according to a semantic category that was not expected, the child did not receive points for that sort. For example, if the child sorted the fruits and vegetables set by color, the child did not receive points for that sort. The proportion of correctly sorted category pictures was calculated for each child (i.e., the number of correctly sorted pictures divided by the total number of pictures to be sorted). Children's average receptive category sorting scores ranged from 0 to .99. The mean score was .40 (SD = .32). We also calculated the proportion of correct sorts for the children's English subset and the Spanish subset, resulting in an English receptive category sorting score and a Spanish receptive category sorting score (see Table 1).

Second, the verbal explanations for the groupings were judged to be an expressive measure of semantic skills, the "expressive category sorting score." Answers were accepted in any language. Dialectical differences were recorded and not penalized. If the child's explanations matched the target categories, the child received 2 points. If only one category in the set was matched, the child received 1 point, and if the child did not hit any of the target categories, he received 0 points. Points for the 10 sets were added for a total score. The maximum possible score was 20 (i.e., two target sorts in 10 sets). Children's scores ranged from 0 to 16; the mean score was 8.22 (SD = 4.14). For this measure, we also calculated English and Spanish expressive category sorting scores, for the subset of items presented in one or the other language. A list of children's responses across sites was built in order to increase reliability across scorers. In cases of discrepancies across scorers, agreement was met by consensus and ensured that the same responses were scored in the same ways across sites.

Finaly, we calculated receptive and expressive scores for the set categories that were presented to children twice: once in English and once in Spanish. These set level scores—the proportion of pictures sorted in the target set categories and the number of target explanations—were compared within children.

Analysis

Descriptive statistics were computed for the variables of interest. Bivariate correlational analyses were run across the outcomes, age, and language input and output, and univariate analyses of covariance were used to explore the children's receptive and expressive category sorting performance by version, after controlling for age. Finally, paired t tests were conducted to determine whether children had similar performance across languages for the sets presented in English and Spanish. Expressive category responses were evaluated qualitatively, as well.

Results

Task Difficulty

Table 1 shows the mean scores and standard deviations of all measures to give the reader an overview of the children's performance. Both receptive and expressive category sorting score averages were below the potential midpoint for the measures (i.e., below 0.50 for the receptive

score and below 10 for the expressive score). The task appears to provide variability around the mean and opportunities for differences in relative performance. Mean scores in each language were comparable for both the receptive and the expressive task.

Within language, receptive and expressive category sorting scores were positively and significantly correlated (see Table 2). Across languages, Spanish and English expressive sorting scores were also significantly associated. In contrast, there was no significant association between Spanish and English receptive sorting scores. In considering this relationship, it is important to keep in mind that the large majority of responses were either nonspecific in nature or displayed knowledge of one of the categories only. Once children could specifically identify the categories used for sorting, they were likely to apply this strategy in both languages.

Bivariate correlational analyses also showed statistically significant correlations between age, language output and input, and the outcome variables in most of the cases (see Table 2). Age was positively and significantly correlated to all receptive and expressive category sorting scores. These correlations ranged from .327 (age and expressive receptive category sorting) to .667 (age and receptive category sorting). Correlations between age and Spanish sorting scores were stronger than between age and English sorting scores. These positive significant correlations with age suggest that these tasks are developmental. Age was also significantly associated to language input and output. Specifically, English input and output were positively correlated to age, indicating that the English input and output of children tend to increase with age.

Linguistic input and output also showed positive and significant correlations with some but not all outcomes. English input was significantly associated to expressive category sorting and Spanish expressive category sorting scores, while English output was significantly associated to receptive category sorting and all three expressive category sorting scores. Overall, linguistic variables were more frequently associated to the expressive scores rather than the receptive scores.

The three experimental category sorting versions did not present the same challenge level to children (see Table 3). There was no significant difference across scores for the receptive category sorting task, $F(2, 40) = 1.894$, $p = .164$, and the Spanish receptive category sorting task, $F(2, 40) = 0.435$, $p = .650$, across versions. In contrast, the children's English receptive category sorting scores were significantly different, $F(2, 40) = 25.104$, $p < .001$. Children obtained higher scores for Version 2. Age, the covariate, was significantly related to all receptive scores: receptive category sorting, $F(1, 40) = 30.457$, $p < .001$; Spanish receptive category sorting, $F(1, 40) = 25.346$, $p < .001$; English receptive category sorting, $F(1, 40) = 25.104$, $p < .001$.

Similarly, there were significant differences across versions for the expressive category sorting scores. The English expressive category sorting task score, $F(2, 40) = 6.146$, $p = .005$, was significantly different across versions, indicating that Version 2 resulted in higher scores. The Spanish expressive category sorting task scores were not significantly different across versions, $F(2, 40) = 0.185$, $p = .831$. The same was found for the combined expressive category sorting scores, $F(2, 40) = 2.768$, $p = .075$. Age was also a significant covariate for the three expressive scores: expressive category sorting, $F(1, 40) = 20.064$, $p < .001$; Spanish expressive category sorting, $F(1, 40) = 19.332$, $p < .001$; English receptive category sorting, $F(1, 40) = 14.367$, $p < .001$.

Paired t tests for the items that were administered in both English and Spanish were only significant for two sets of items across the English and Spanish receptive category sorting (see Table 4). In both cases, English scores were significantly higher than Spanish scores. The other comparisons did not reveal a significant difference across languages.

Qualitative Analysis

A qualitative evaluation of children's responses showed that there were multiple levels of potential responses at the expressive level. The highest level was to provide a specific label representing each of the sorts (e.g., labeling fruits and vegetables and the corresponding shape characteristics). This level of specificity was rarely produced and only by the oldest children. Most (47%) responses were of the type "these are X and these are not," where X might refer to a specific label related to the categories in question. These responses

Table 2. Correlations between age, English input and output, and receptive and expressive category sorting results.

Variable	1	2	3	4	5	6	7	8	9
1. Age	—								
2. Receptive category sorting	.661**	—							
3. English receptive category sorting	.327*	.622**	—						
4. Spanish receptive category sorting	.615**	.862**	.255	—					
5. Expressive category sorting	.569**	.877**	.699**	.721**	—				
6. English expressive category sorting	.486**	.836**	.763**	.592**	.940**	—			
7. Spanish expressive category sorting	.579**	.789**	.516**	.761**	.916**	.724**	—		
8. English input	.434**	.269	.147	.238	.311*	.251	.340*	—	
9. English output	.400**	.313*	.258	.261	.408**	.376*	.392*	.911**	—

Note. English input and output were collected. Spanish input and output are the inverse of English.

*$p < .05$. **$p < .01$.

Table 3. Receptive and expressive category sorting scores by version.

Variable	Version 1 (n = 15) M (SD)	Version 2 (n = 14) M (SD)	Version 3 (n = 17) M (SD)
Receptive category sorting	0.41 (0.20)	0.58 (0.21)	0.44 (0.20)
English receptive category sorting	0.39 (0.21)	0.71 (0.23)*	0.14 (0.21)
Spanish receptive category sorting	0.43 (0.24)	0.44 (0.21)	0.48 (0.24)
Expressive category sorting	7.60 (3.70)	10.43 (5.03)*	6.94 (3.05)
English expressive category sorting	3.80 (2.21)	5.64 (2.59)*	2.88 (1.70)
Spanish expressive category sorting	3.80 (2.01)	4.79 (2.55)	4.06 (1.60)

*$p < .05$.

demonstrate knowledge of categorization and differing degrees of knowledge of specific vocabulary. Another 38% of the responses represented nonspecific responses that did not indicate that the child could verbalize specific knowledge that guided their sorting strategy.

Discussion

In this cognitive lab study, we aimed to develop a functional task that could inform intervention in the area of semantics. We focused on how bilingual Spanish–English-speaking children ages 4–10 years create semantic categories in a flexible manner. Typically developing children appeared interested in and able to shift sets and sort pictures corresponding to different categories in a flexible way, suggesting that this task has the potential to assess this functional semantic skill. Their performance for both the receptive and expressive category sorting task was positively and significantly correlated to age, indicating that this is a developmental task. In addition, age was a significant covariate for the receptive and expressive tasks. On average, children sorted and explained categories on half of the opportunities. Only two children did not sort any of the 10 sets in a flexible manner. These children were ages 4;0 and 5;6. Five children could only sort up to 25% of the pictures; in this case, four of them were 5 years old, and one was 9 years old. In contrast, the three children who sorted at least 75% of the pictures were 9 and 10 years old. The average age of the children who flexibly sorted between 25% and 75% of pictures was 8;5. Thus, this type of task appears to be sensitive to semantic development.

Without any semantic cues as to the intended sort strategy, children showed a variety of responses. No child sorted the 10 sets twice according to the target categories, and only five children sorted the 10 sets in at least one intended category. Based on the difference across versions, it is clear that there were certain categories that were more reliable than others. For example, the distinction between dirty and clean clothes and household items was a set that presented difficulties to few children: Only two children failed the two sorts when the prompts were presented in English, and two other children failed in Spanish. In contrast, there were other sets that tended to elicit variable responses, such as categories about food and sports. Less than half of the children sorted individual and team sports that are played with or without a ball according to at least one of the prototypical categories when the set was presented in English. Fifty-eight percent of the children sorted this set into one prototypical category when presented in Spanish; however, no child showed the flexibility to sort this set into the two categories.

We also explored this novel category sorting task in relationship to language input and output. Taking into account that age was correlated to both English input and output and to the category sorting scores, it is not surprising that the category sorting scores were positively correlated with English input and output. English output, in particular, is associated with increased performance on the expressive

Table 4. Means and standard deviations for sets sorted twice (once in each language) by the same children.

Set	n	Receptive category sorting				Expressive category sorting			
		English		Spanish		English		Spanish	
		M	SD	M	SD	M	SD	M	SD
Mammals vs. reptiles; land vs. water	15	0.40	0.36	0.44	0.34	0.67	0.62	0.67	0.72
Pets vs. wild animals; land vs. water	7	0.36	0.24	0.36	0.24	0.87	0.52	0.73	0.59
Fruits vs. vegetables; long vs. round	13	0.59*	0.29	0.44	0.25	1.07	0.73	0.86	0.53
Sweet vs. savory food; healthy vs. unhealthy food	14	0.48*	0.25	0.27	0.24	0.71	0.61	0.50	0.52
Individual vs. team sports; with or without a ball	17	0.17	0.22	0.26	0.28	0.41	0.51	0.59	0.62
Emergency vs. daily vehicles; big and small	16	0.45	0.37	0.45	0.37	0.94	0.66	0.76	0.75

*$p < .05$.

category sorting, that is, in the ability of children to explain the categories. This finding underscores the importance of current linguistic output (or language use) as a parent or teacher report measure that predicts a child's linguistic ability (Bohman et al., 2010; Gutiérrez-Clellen et al., 2012).

Bilingual children appeared to sort the sets in similar ways across their two languages. In every version of the task, there were two sets that were flexibly sorted in response to English and Spanish prompts, in that order. Two out of six sets received significantly higher scores in English than in Spanish. Due to the small sample size and the previously reported version differences, it is important to use caution when interpreting this finding. The other four sets elicited similar answers in both languages. Evaluators noted that children were aware of the fact that the set was repeated once it was presented in Spanish for the second time. It is possible that their motivation decreased responsivity in these cases. Overall, both receptive and expressive English and Spanish were similar, suggesting that children approached the task in a bilingual manner. As a matter of fact, children's explanations included answers in the nontarget language and utterances with code switches.

This cognitive lab study also revealed different strategies to explain the categories. Children explained them by labeling the overarching category, by labeling the two subcategories, or by labeling one category and stating the other group was not a member of the category. For example, for land versus water animals, some of children's responses were as follows:

> These are sea animals; and these are land animals.
> These go in the water; and these don't go in the water.
> These know how to swim; and these don't know how to swim.
> *Son del océano; son de tierra* (they are from the ocean; they are from the land).
> **Estos viven en la agua; y estos en la tierra* (these live in the water; and these in the land).
> *Nadan en el agua; no nadan* (they swim in the water; they don't swim).

For some children, the Spanish explanations appeared to use additional lexical strategies such as circumlocutions or descriptions (Greene et al., 2013). For example, see the following explanations for land and water animals:

> *En el agua; En un lugar que es seco* (in the water; in a place that is dry).
> *Pueden nadar y pueden caminar* (they can swim and they can walk).

In addition, some children code-switched to the nontarget language in their responses. These code switches were observed in both English and Spanish sets:

> *Ir en la piscina* (in the pool), they swim; they can't swim.

One source of difference in children's response patterns is likely to reflect children's gradual shift to English as reflected here in the correlation of English language use and age. Furthermore, many of the themes selected reflect topics children use at school so they may be more likely to know that vocabulary in the language of schooling and respond by describing different sets when using English. However, it is worth noting that, when children performed by shifting sets to form different categories in Spanish, their responses were of equally high quality. Consistent with the idea that children draw on their cumulative language experience when children classified the same items in each language, their strategy and explanations were similar across languages. This suggests that the differences in performance are more likely tied to language knowledge (e.g., semantic mappings) than they would be to children's monolingual or bilingual status.

Study Limitations and Future Directions

Several methodological issues present limitations to the study. For example, socioeconomic status and school settings were not controlled. In addition, the experimental task was always initiated in English without any counterbalancing. Thus, there is the possibility of language order bias. In addition, there is a possibility that the slight variations in instruction prompts across languages, strategically designed to facilitate comprehension in each language, may have had unforeseen effects on the set shifting task, a cognitive task.

The future direction of this project is to continue investigating the development of semantic skills in bilingual children and increase our ability to distinguish difference from disorder. In addition, this type of knowledge will help us understand how developmental information around tasks, such as the category sorting task, could fruitfully be employed in intervention. While it has long been known that we need to focus on the causes of language impairment rather than the symptoms, much research has focused on the nature and evolution of symptoms. By looking at the ways that children's responses change to tasks that would address their underlying knowledge, we can make decisions about the level at which children's language knowledge would warrant attention to a skill such as the one explored in this study. We can then determine a sequence of expected changes and have externally validated responses that would indicate if the child is making progress.

Acknowledgments

This work was funded by National Institute on Deafness and Other Communication Disorder Grant R01DC015588 to Lisa Bedore (principal investigator). We thank our participants and their families for supporting this research.

References

American Speech-Language-Hearing Association. (2014). *2014 School surveys.*

Arias, G., & Friberg, J. (2017). Bilingual language assessment: Contemporary versus recommended practice in American

schools. *Language, Speech, and Hearing Services in Schools, 48*(1), 1–15. https://doi.org/10.1044/2016_LSHSS-15-0090

Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research, 41*(5), 1185–1192. https://doi.org/10.1044/jslhr.4105.1185

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., Gillam, R. B., & Ho, T.-H. (2010). Language sample measures and language ability in Spanish–English bilingual kindergarteners. *Journal of Communication Disorders, 43*(6), 498–510. https://doi.org/10.1016/j.jcomdis.2010.05.002

Bleses, D., Højen, A., Justice, L. M., Dale, P. S., Dybdal, L., Piasta, S. B., Markussen-Brown, J., Clausen, M., & Haghish, E. F. (2018). The effectiveness of a large-scale language and preliteracy intervention: The SPELL randomized controlled trial in Denmark. *Child Development, 89*(4), e342–e363. https://doi.org/10.1111/cdev.12859

Bohman, T. M., Bedore, L., 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

Brédart, A., Marrel, A., Abetz-Webb, L., Lasch, K., & Acquadro, C. (2014). Interviewing to develop patient-reported outcome (PRO) measures for clinical research: Eliciting patients' experience. *Health and Quality of Life Outcomes, 12*(1), Article no. 15. https://doi.org/10.1186/1477-7525-12-15

Crystal, D. (1998). Sense: The final frontier. *Child Language Teaching and Therapy, 14*(1), 1–27. https://doi.org/10.1177/026565909801400101

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 560 million words, 1990–present*. https://www.english-corpora.org/coca/

Davies, M. (2016). *Corpus del Español: Two billion words, 21 countries*. http://www.corpusdelespanol.org/web-dial/

Erkut, S., Alarcón, O., García Coll, C., Tropp, L. R., & García, H. A. V. (1999). The dual-focus approach to creating bilingual measures. *Journal of Cross-Cultural Psychology, 30*(2), 206–218. https://doi.org/10.1177/0022022199030002004

Gibson, T. A., Peña, E. D., & Bedore, L. M. (2014). The relation between language experience and receptive-expressive semantic gaps in bilingual children. *International Journal of Bilingual Education and Bilingualism, 17*(1), 90–110. https://doi.org/10.1080/13670050.2012.743960

Gillam, R. B., Loeb, D. F., Hoffman, L. M., Bohman, T., Champlin, C. A., Thibodeau, L., Widen, J., Brandel, J., & Friel-Patti, S. (2008). The efficacy of Fast ForWord language intervention in school-age children with language impairment: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 51*(1), 97–119. https://doi.org/10.1044/1092-4388(2008/007)

Gillam, S. L., Gillam, R. B., & Reece, K. (2012). Language outcomes of contextualized and decontextualized language intervention: Results of an early efficacy study. *Language, Speech, and Hearing Services in Schools, 43*(3), 276–291. https://doi.org/10.1044/0161-1461(2011/11-0022)

Gonzalez-Barrero, A. M., & Nadig, A. S. (2019). Can bilingualism mitigate set-shifting difficulties in children with autism spectrum disorders. *Child Development, 90*(4), 1043–1060. https://doi.org/10.1111/cdev.12979

Greene, K. J., Peña, E. D., & Bedore, L. M. (2013). Lexical choice and language selection in bilingual preschoolers. *Child Language Teaching and Therapy, 29*(1), 27–39. https://doi.org/10.1177/0265659012459743

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2007). The discriminant accuracy of a grammatical measure with Latino English-speaking children. *Journal of Speech, Language, and Hearing Research, 50*(4), 968–981. https://doi.org/10.1044/1092-4388(2007/068)

Gutiérrez-Clellen, V. F., Simon-Cereijido, G., & Sweet, M. (2012). Predictors of second language acquisition in Latino children with specific language impairment. *American Journal of Speech-Language Pathology, 21*(1), 64–77. https://doi.org/10.1044/1058-0360(2011/10-0090)

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergarteners. *Language, Speech, and Hearing Services in Schools, 36*(1), 17–32. https://doi.org/10.1044/0161-1461(2005/003)

Kapa, L. L., & Plante, E. (2015). Executive function in SLI: Recent advances and future directions. *Current Develomental Disorders Report, 2*(3), 245–252. https://doi.org/10.1007/s40474-015-0050-x

Kapantzoglou, M., Fergadiotis, G., & Restrepo, M. A. (2017). Language sample analysis and elicitation technique effects in bilingual children with and without language impairment. *Journal of Speech, Language, and Hearing Research, 60*(10), 2852–2864. https://doi.org/10.1044/2017_JSLHR-L-16-0335

Kapantzoglou, M., Restrepo, M. A., & Thompson, M. S. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in Schools, 43*(1), 81–96. https://doi.org/10.1044/0161-1461(2011/10-0095)

Kroll, J. F., van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism (Cambridge, England), 13*(3), 373–381. https://doi.org/10.1017/S136672891000009X

McCauley, R. J., Fey, M. E., & Gillam, R. G. (2017). *Treatment of language disorders in children* (2nd ed.). Brookes.

Nelson, K. (1993). Events, narratives, memory: What develops. In C. A. Nelson (Ed.), *Memory and affect in development* (Vol. 26, pp. 1–24). Erlbaum.

Owen Van Horne, A. J., & Green Fager, M. (2015). Quantifying the relative contributions of lexical and phonological factors to regular past tense accuracy. *International Journal of Speech-Language Pathology, 17*(6), 605–616. https://doi.org/10.3109/17549507.2015.1034174

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. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders/ Royal College of Speech & Language Therapists, 51*(2), 192–202. https://doi.org/10.1111/1460-6984.12199

Peña, E. D., 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., Gillam, R. B., Resendiz, M., Fiestas, C. E., & Bedore, L. M. (2008). *Dynamic assessment of narratives: Improving assessment accuracy in diverse learners*. Presentation at the 2008 American Speech Language and Hearing Association Convention, Chicago, IL.

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

Potter, M. C., So, K.-F., Eckardt, B. V., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 23–38. https://doi.org/10.1016/S0022-5371(84)90489-4

Restrepo, M. A., Morgan, G. P., & Thompson, M. S. (2013). The efficacy of a vocabulary intervention for dual-language learners with language impairment. *Journal of Speech, Language, and Hearing Research, 56*(2), 748–765. https://doi.org/10.1044/1092-4388(2012/11-0173)x

Selin, C. M., Rice, M. L., Girolamo, T., & Wang, C. J. (2019). Speech-language pathologists' clinical decision making for children with specific language impairment. *Language, Speech, and Hearing Services in Schools, 50*(2), 283–307. https://doi.org/10.1044/2018_LSHSS-18-0017

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

Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2007). Spontaneous language markers of Spanish language impairment. *Applied Psycholinguistics, 28*(2), 317–339. https://doi.org/10.1017/S0142716407070166

Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2014). Bilingual education for all: Latino dual language learners with language disabilities. *International Journal of Bilingual Education and Bilingualism, 17*(2), 235–254. https://doi.org/10.1080/13670050.2013.866630

Thordardottir, E. (2010). Towards evidence-based practice in language intervention for bilingual children. *Journal of Communication Disorders, 43*(6), 523–537. https://doi.org/10.1016/j.jcomdis.2010.06.001

Thordardottir, E., Cloutier, G., Ménard, S., Pelland-Blais, E., & Rvachew, S. (2015). Monolingual or bilingual intervention for primary language impairment? A randomized control trial. *Journal of Speech, Language, and Hearing Research, 58*(2), 287–300. https://doi.org/10.1044/2014_JSLHR-L-13-0277

Uro, G., & Lai, D. (2019). *English language learners in America's great city schools: Demographics, achievement, and staffing*. Council of the Great City Schools.

Wiig, E., Semel, E. H., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5)*. Pearson.

Williams, C. J., & McLeod, S. (2012). Speech-language pathologists' assessment and intervention practices with multi-lingual children. *International Journal of Speech-Language Pathology, 14*(3), 292–305. https://doi.org/10.3109/17549507.2011.636071