Journal of Communication Disorders 53 (2015) 30–41

Contents lists available at ScienceDirect

Journal of Communication Disorders

Discriminant accuracy of a semantics measure with Latino English-speaking, Spanish-speaking, and English–Spanish bilingual children

Elizabeth D. Peña [a,*], Lisa M. Bedore [a], Ellen S. Kester [b]

[a] *Communication Sciences and Disorders, University of Texas at Austin, United States*
[b] *Bilinguistics, Austin, TX, United States*

ARTICLE INFO

Article history:
Received 6 December 2013
Received in revised form 18 October 2014
Accepted 23 November 2014
Available online 12 December 2014

Keywords:
Bilingual
Language impairment
Spanish
English
Diagnostic accuracy

ABSTRACT

We explored classification accuracy of English and Spanish versions of an experimental semantic language measure with functional monolingual–bilingual children with and without language impairment. A total of 441 children participated, including 78 balanced bilinguals (15 with language impairment, 63 with typical development); 179 monolingual Spanish (36 with language impairment, 143 with typical development); and 183 monolingual English (49 with language impairment, 134 with typical development) children between 4;0 and 6;11 years. Cut points derived for functionally monolingual children were applied to bilinguals to assess the predictive accuracy of English and Spanish semantics. Correct classification of English monolinguals and Spanish monolinguals was 81%. Discriminant analysis yielded 76% and 90% correct classification for balanced bilingual children in English and Spanish respectively. This semantics-based measure has fair to good classification accuracy for functional monolinguals and for Spanish–English bilingual children when one language is tested.

Learning outcomes: As a result of this study, the reader will describe advantages of lexical-semantic tasks for identification of language impairment. They will be able to describe procedures for conceptual scoring and identify its benefits. Readers will also gain an understanding of similarities and differences in bilingual and monolingual performance on a semantics task in Spanish and English.

1. Introduction

Bilingual children with language impairment (LI), like their monolingual counterparts, have difficulties in the lexical-semantic domain (Gray, 2004; Jordaan, Shaw-Ridley, Serfontein, Orelwitz, & Monaghan, 2001; McGregor, 2009; Sheng, Bedore, Peña, & Taliancich-Klinger, 2013; Sheng, Peña, Bedore, & Fiestas, 2012) but such measures have not been evaluated for their classification accuracy. Bilingual children may demonstrate low performance, indicated by lower than expected scores, in the lexical-semantic domain due to LI, level of exposure to each language, or both. This low performance may be due to divided input in each of their two languages or because of acquisition difficulties associated with language impairment. Therefore a significant difficulty that speech-language pathologists face in identifying LI in bilingual children is determining if lower than expected language performance should be attributed to typical bilingual development or LI.

* Corresponding author. Tel.: +1 512 471 2690.
E-mail address: lizp@mail.utexas.edu (E.D. Peña).

http://dx.doi.org/10.1016/j.jcomdis.2014.11.001
0021-9924/

A number of recent studies have focused on diagnostic accuracy of bilingual children's language ability on measures of morphosyntax (Gutiérrez-Clellen, Restrepo, & Simon-Cereijido, 2006; Gutiérrez-Clellen & Simon-Cereijido, 2007), nonword repetition (Gutiérrez-Clellen & Simon-Cereijido, 2010; Kohnert, Windsor, & Yim, 2006; Windsor, Kohnert, Lobitz, & Pham, 2010), and counting span (Danahy, Windsor, & Kohnert, 2007), as well as language sampling approaches (Bedore, Peña, Gillam, & Ho, 2010; Gutiérrez-Clellen & Simon-Cereijido, 2009; Simon-Cereijido & Gutiérrez-Clellen, 2007) but little work has focused on the lexical-semantic domain. A recent meta-analysis (Dollaghan & Horner, 2011) shows that there are a number of promising measures of language ability, particularly in the morphosyntactic domain, for Spanish–English bilinguals. Nevertheless, the authors noted that many of these measures had negative likelihood ratios with unacceptably large confidence intervals which limit the certitude of a negative diagnostic result.

Lexical-semantic measures of language are potentially useful for bilingual children in the process of learning two languages. First, bilingual children with LI are documented to have difficulty in the area of semantics (Sheng et al., 2012) as do monolingual children with LI (Gray, 2004; McGregor, 2009). Tasks that challenge the lexical-semantic system seem to be more robust indicators of impairment than single-word vocabulary knowledge. Such semantics tasks may have an advantage over assessment of basic vocabulary knowledge because they focus on use and organization of the lexicon rather than knowledge of a predetermined item set. Second, semantic knowledge is distributed across both of the bilingual's languages and can be expressed in either of two languages. Testing in one language at a time can allow for responding in either language, which one cannot account for when testing morphosyntax. These approaches to assessment of the lexical-semantic domain may help to level the playing field for children with varying experiences (Calderón et al., 2005; Hoff & Naigles, 2002; Hoff & Tian, 2005).

1.1. Vocabulary and semantic performance in language impairment

There are a number of commonly used measures of vocabulary development (Betz, Eickhoff, & Sullivan, 2013) including the *Peabody Picture Vocabulary Test* (Dunn & Dunn, 1981; Dunn & Dunn, 1997) and the *Expressive One Word Picture Vocabulary Test-2000* Edition (Brownell, 2000). These include Spanish-language versions: the *Test de Vocabulario en Imagenes Peabody* (TVIP, Dunn, Padilla, Lugo, & Dunn, 1986); and the *Expressive One Word Picture Vocabulary Test-Spanish Bilingual Edition* (Brownell, 2001). These measures focus on the production or comprehension of single words. Research demonstrates that children with LI have significantly lower single-word vocabulary skills than do their typically developing peers. However, many of these single-word-based tests have poor classification accuracy for English monolinguals (Gray, Plante, Vance, & Henrichsen, 1999; Restrepo et al., 2006; Spaulding, Plante, & Farinella, 2006) and for children from minority language backgrounds including African American and Latino American children (Peña & Quinn, 1997).

In addition to smaller vocabulary, children with LI often have significant semantic processing difficulties compared to their typically developing peers. These difficulties often occur even if children have relatively good vocabulary comprehension (e.g., McGregor & Windsor, 1996). The documented processing limitations of children with LI can lead to delays in word learning and in making connections among lexical-semantic entries. These difficulties can affect performance on a number of lexical-semantic tasks. For example, Nippold, Erskine, and Freed (1988) documented that children, ages 6–8 with language impairment, had particular difficulty on tasks requiring analogical reasoning in comparison to their typical peers. Children with language impairment have also been shown to perform more poorly on tasks of description which required them to describe a picture making comparisons to other pictures (Bishop & Adams, 1991). Children with language impairments also have documented difficulties with verbal fluency. Weckerly, Wulfeck, and Reilly (2001) compared children of ages 8–12 with and without language impairment on a set of verbal fluency tasks. There was a semantic condition in which children generated as many exemplars in a category in 90 s and a phonetic condition where they generated as many items beginning with a given sound in 60 s. Overall, children with language impairment generated fewer items in both conditions and made more errors compared to children with typical development. Children with LI also have difficulties naming action. Sheng and McGregor (2010b) examined object and action naming in children with LI as well as their age and verbal matched peers. While children with LI were less accurate and slower in naming as compared to their age-matched peers, they had particular difficulty with action naming. In a recent study Sheng and McGregor (2010a) showed that children with LI demonstrated more tenuous, shallow connections between words on a repeated associations task (where children provide a semantically related word three times in a row in response to an item such as "dinner"), in comparison to children with typical development. In this study, children of ages 5–8 provided three associations when provided with a stimulus word. Children with language impairment made more out of class (e.g., "rainbow" in response to what goes with the word "dinner") and phonological errors (e.g., "linner" in response to what does with the word "dinner") compared to children with typical development. Consistent with other reports comparing vocabulary and semantic tasks, children with language impairment demonstrated different patterns of responses compared to their vocabulary matched peers. Their pattern of response indicates difficulty making semantic links among related concepts not predicted by level of vocabulary knowledge. In contrast to single-word vocabulary testing, these semantic tasks require deeper knowledge and organization of the semantic system making them particularly challenging for children with LI. Brackenbury and Pye (2005) reviewed and summarized the semantic difficulties documented for children with LI. They concluded that assessment practices should include not only documentation of vocabulary size but also of children's ability to make connections among lexical entries through tasks such as picture description, and tasks that assess knowledge of category, part-whole, and thematic relationships.

There are few studies documenting lexical-semantic performance in bilingual children with LI. Thus far however, there are similar patterns of semantic difficulties for bilingual and monolingual children with LI. For example, on a picture-naming task, Simonsen (2002) found that when bilingual children did not know an item, children with LI were more likely to provide no response, whereas the typically developing controls provided a description of the target. On a repeated-associations task bilingual children with TD demonstrated higher semantic depth scores than children with LI (Sheng et al., 2012). Children with LI were more likely than their TD peers to repeat words, give no response, or make errors in both their languages. These differences were observed even after controlling for differences in vocabulary size. Bilingual children with language impairment also demonstrated difficulty on a definitions task. Specifically, Gutiérrez-Clellen and DeCurtis (2001) noted that when providing definitions of words, children with LI used non-specific vocabulary, limited responses, and were less likely to generate multiple meanings of words.

1.2. Accounting for distributed knowledge in two languages

A well-documented finding in assessment of bilinguals' vocabulary is that their single-language scores are low compared to monolingual norms (Bialystok, Luk, Peets, & Yang, 2010; Carlo et al., 2004). These low single-language scores are likely due to children's divided input in each of their two languages (Oller, Pearson, & Cobo-Lewis, 2007). Thus, for bilinguals, it is especially important to account for knowledge in each of their two languages.

A strategy for accounting for knowledge in two languages is conceptual scoring. With this method, children are credited for use of vocabulary in either of their two languages, with correct responses credited once (Pearson & Fernández, 1994). Conceptual scores are generated by combining data from independent testing in two languages (e.g., Umbel, Pearson, Fernández, & Oller, 1992) or by crediting responses in either language on a single administration of a test (Brownell, 2001). The logic of these conceptual scoring approaches is that they accurately represent the child's language knowledge because they account for concepts represented in either language and they do not double count concepts represented in both languages.

When children's conceptual knowledge is accounted for by crediting items produced or responded to in the non-test language these scores are often comparable to that of monolinguals. Pearson, Fernández, and Oller (1993) first applied this idea to examine bilingual toddler's expressive and receptive vocabularies. Using the English (Dale, Bates, Reznick, & Morisset, 1989) and Spanish (Jackson-Maldonado, 2003) versions of the MacArthur Communicative Development Inventory (MCDI), Pearson and colleagues found that while bilingual's single-language scores were lower compared to a monolingual comparison group, their conceptual scores were more comparable. This finding has been replicated with the more recent versions of these instruments in English (Fenson et al., 2006) and Spanish (Jackson-Maldonado, 2003) by Mancilla-Martinez, Pan, and Vagh (2011) for both comprehension and production.

Conceptual scores have also been successfully applied is semantic measures to reduce monolingual–bilingual differences including category generation (Peña, Bedore, & Zlatic-Guinta, 2002) and repeated associations (Sheng, McGregor, & Marian, 2006). In a recent study of repeated associations, Sheng et al. (2012) examined individual bilingual children's performance as compared to group means derived in Spanish and English. Use of conceptual scoring was most accurate in determining both TD and LI than either Spanish scoring or English scoring.

Peña, Bedore, and Rappazzo (2003) developed and evaluated a set of expressive and receptive items organized around six task types including: analogies, characteristic properties, categorization, functions, linguistic concepts, and similarities and differences. For each task type in each language 12 items were tested (6 expressive and 6 receptive) with the exception of categorization, which had 18 expressive and 8 receptive items. The item sets were not translation equivalents but rather tested parallel concepts in different contexts. For English, the item set was organized around a birthday party theme and for Spanish the items were organized around a picnic (or "día de campo") theme. Children who were dominant speakers of Spanish, dominant speakers of English, and balanced, Spanish–English bilinguals were compared. There were no differences for total scores across the two versions of the task. Nor were there differences between bilingual and Spanish dominant and English dominant in each language. There were however differences by language of testing and task. This pattern of findings demonstrated that across languages, similar items had different levels of difficulty. A reduced set of items was selected from the larger item set for Spanish and English and tested with a group of 40 children with typical development. Bedore, Peña, García, and Cortez (2005) compared single-language scores to conceptual scores using these sets of lexical-semantics items. With single-language scores, more of the typical bilingual children fell below the range for monolingual children in each language; but when conceptual scores were used, the range of performance was more similar among monolingual and bilingual children.

The norm of comparison (for example, English-language norms or Spanish-language norms) may affect interpretability of conceptual scores. Thordardottir, Rothenberg, Rivard, and Naves (2006) found that French–English bilinguals scored lower than their monolingual counterparts on English and French MCDIs using single-language scores in each language. Conceptual scores used with the French version of the measure were more comparable to their French monolingual counterparts'. In contrast, conceptual scores on the English version of the measure were still lower than their English monolingual peers'. Examination of individual scores among these typical children showed considerable variation in each language. Similarly, when Mancilla-Martinez et al. (2011) compared single-language and conceptual scores to English-language norms and Spanish-language norms they found that fewer children fell below −1SD when conceptual scores were used. However, more children fell below −1SD when compared to English-language norms vs. Spanish-language norms even with conceptual scores. It is possible that between-language word frequency differences led to between-language differences, even when conceptual scoring was used. Frequency differences between languages may lead to item difficulty differences (Restrepo & Silverman, 2001). Difficulty levels may be affected by frequency (Tamayo, 1987) tied to cultural-linguistic experience. It may be that a vocabulary or lexical-semantic test developed for a monolingual context may not tap into the words a bilingual child knows in their other language. For bilinguals, who have divided input word frequency differences between their two languages may affect their performance on tests designed for monolinguals.

It is clear that conceptual scoring reduces discrepancies when bilinguals are compared to monolingual norms, potentially leading to reduction in the rate of false positives. But, the (monolingual) norm children are compared to may also affect diagnostic decisions (Pearson, 2010). However, to date there are no studies comparing children with and without language impairment using conceptual scores. It is unknown if the use of conceptual scoring could be used to reduce classification error in bilingual relative to monolingual classification.

1.3. Motivation for the current study

Lexical-semantic tasks have the potential to be good indicators of language impairment. They are more demanding than single-word vocabulary tasks and children with language impairment have particular difficulty with these tasks even after controlling for vocabulary knowledge. For bilinguals, it is additionally important to account for knowledge of two languages. In the semantic domain it is possible to do so through use of conceptual scoring. Thus, we focus on two questions in the current study:

1. What is the diagnostic accuracy on the English and Spanish semantics subtests of the BESA for monolingual children of each language?
2. For BB children, how does diagnostic accuracy compare with that of monolingual children when conceptual scoring is used?

2. Methods

2.1. Participants

Data for this study were drawn from those gathered in development of the Bilingual English Spanish Assessment (BESA), which includes measures of semantics, morphosyntax, phonology, and pragmatics in Spanish and English. A total of 785 children participated in the tryout and normative phases of development of the semantics subtest. Children were recruited to this study from Texas, California, and Pennsylvania. They ranged in language experience from 100% exposure to English and 0% exposure to Spanish to 100% exposure to Spanish and 0% exposure to English. Participants were grouped by language experience, age, and language ability (procedures are described below). The current study focuses on 78 balanced bilingual children who completed testing in Spanish and English and their functional monolingual counterparts (183 participants in Spanish and 179 participants in English) (see Table 1).

2.2. Instrument development

The experimental version of the BESA semantics subtest (Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2008) was designed to be neutral to sources of variability related to bilingual exposure while retaining sensitivity to LI. Specifically, items were designed so that children with different levels of exposure to English and Spanish could still respond successfully. Item types were those targeting semantic development, including description of characteristics (e.g., tell me about *X*), categorization (e.g., tell me all the animals you can think of), functions (e.g., what is an *X* used for?), linguistic concepts (e.g., what shape is this *X*?) verbal analogies (e.g., *X* is to *Y* as *A* is to ____), and similarities and differences (comparisons; how are these two *X* different?) (Bedore et al., 2005; Peña et al., 2003).

Instrument development was completed in two phases. The initial semantics item set included a minimum of 12 items for each of the types indicated above (6 receptive and 6 expressive) for a total of 86 items in Spanish and English each (Peña et al., 2003). This item set was administered to a group of 72 Latino-American children with varying exposure to English and Spanish. Children ranged in age from 3;7 to 7;2 with a mean of 5;7 and were further divided into three age groups (23, 3;7–4;11; 30, 5–5;11; and 19, 6;0–7;2). Of the 72 children, 55 were identified with typical development and 17 with language impairment. The children with language impairment were clustered in the 5–5;11 age range. Children were identified as functionally monolingual or bilingual based on parent and teacher report of input and output of English and Spanish (Bedore et al., 2005; Gutiérrez-Clellen & Kreiter, 2003; Peña et al., 2003; Peña et al., 2002). Eighteen participants were identified as functionally monolingual English speakers, 29 were functionally monolingual Spanish speakers, and 25 were bilingual. Refinement and selection of 48 items for Spanish and 49 items for English from this larger set of items were completed using a classical test theory approach (Allen & Yen, 1979). Specifically, items were selected if the 5-year-old children with typical development were at least 20% more accurate than those with language impairment; or if older typical children performed were more accurate than younger typical children. The selected item sets were psychometrically equivalent between the

Table 1
Number of participants by age, language ability, and language exposure.

Age	FME	BB	FMS	All
LI				
4	20	4	19	43
5	22	4	14	40
6	7	7	3	17
Total	49	15	36	100
TD				
4	40	6	34	80
5	51	22	63	138
6	43	35	46	126
Total	134	63	143	344

Note: LI, language impaired; TD, typically developing; FME, functional monolingual English; BB, balanced bilingual; FMS, functional monolingual Spanish.

two languages with similar numbers of items in the easy, medium and high-difficulty ranges based on the children with typical development divided into three age ranges. For English there were 16 easy, medium and difficult items. In Spanish there were 17 easy, 15 medium, and 17 difficult items. Note that the final items were not designed to be translation equivalents. Thus, the experimental version of the semantics measures employed in the current study is the 49 Spanish and 48 English item sets. Coefficient alpha for the English item set was .88 and .85 for the Spanish item set, indicating good internal consistency (DeVellis, 1991).

2.3. Norming

We calculated standard scores for each language based on the means and standard deviations of 561 typical children at 6-month intervals, with a mean of 100 and standard deviation of 15. Table 2 (upper panel) displays the parent and teacher language ratings, percent exposure to English overall, and percent exposure to English in school. Percentage of children who participated in the free or reduced lunch program is included as an index of socioeconomic status.

Spanish means include scores of children were used Spanish at least 60% of the time. Data from the children in the 40–60% use range were included in this normative sample if their Spanish score was higher than English. Similarly, English means included children who used English at least 60% of the time. Those in the 40–60% use range were included if their English test score was higher than Spanish. Finally, for typical children who completed testing in both languages, if their semantics scores were within 10% of each other they were included in both comparison norms. The decision to exclude cases in the weaker language from the norm was to avoid conflating low scores due to limited linguistic experiences with those due to low language ability. There were 329 children in the English norm and 332 children in the Spanish norm; 100 children were included in both languages.

2.4. Procedures

Testing included elicitation of 100-utterance language samples which were used as an external measure of language ability. Semantics testing using the reduced item sets was completed as part of a battery of tests including morphosyntax,

Table 2
Demographic information: parent and teacher reports, SES (standard deviations in parentheses).

		FME	BDE	BB	BDS	FMS
TD	Teacher rating of English	3.92 (0.31)	3.53 (0.94)	2.63 (1.25)	2.20 (1.13)	1.55 (1.18)
	Teacher rating of Spanish	0.36 (0.67)	1.74 (1.57)	3.00 (1.21)	3.41 (1.19)	3.65 (0.81)
	Parent rating of English	3.96 (0.20)	3.63 (0.56)	3.55 (0.70)	2.46 (0.86)	1.55 (1.06)
	Parent rating of Spanish	1.41 (1.00)	2.68 (1.08)	3.52 (0.72)	3.85 (0.36)	3.95 (0.30)
	Percent English at school	87.07% (28.27)	79.33% (32.92)	51.79% (36.11)	37.66% (28.15)	39.32% (29.90)
	Free reduced lunch	35.82%	41.67%	35.48%	26.37%	35.17%
	English input/output composite[a]	92.07% (5.98)	70.18% (5.37)	49.92% (4.59)	30.06% (5.76)	7.94% (6.71)
LI	Teacher rating of English	3.89 (0.31)		2.71 (1.25)		1.47 (1.06)
	Teacher rating of Spanish	0.38 (0.31)		3.29 (0.76)		3.63 (0.81)
	Parent rating of English	3.82 (0.39)		3.31 (0.75)		1.72 (0.85)
	Parent rating of Spanish	0.82 (0.92)		3.31 (0.63)		3.94 (0.24)
	Percent exposure to English at school	77.05% (37.39)		30.00% (26.48)		31.56% (28.15)
	Free reduced lunch	54.17%		26.67%		24.32%
	English input/output composite[a]	95.70% (5.85)		50.69% (5.07)		7.39% (6.70)

Note: FME, functional monolingual English; BDE, bilingual dominant English; BB, balanced bilingual; BDS, bilingual dominant Spanish; FMS, functional monolingual Spanish. TD, typically developing normative groups; LI, language impaired.

[a] Spanish input/output composite is the inverse.

phonology, and pragmatics. Children were tested over a two- to three-week period, in 30-min increments. On the semantics subtest, items were presented in the target language, but children were allowed to respond in either language. Testing was conducted in Spanish or English consistent with percentage of greater language exposure. Children who presented as bilingual were tested in both languages. Testing was conducted in one language at a time, randomizing order of first language presentation by different examiners on different days.

At the time of testing and scoring, testers and authors were blind to child language ability. Testing was completed in the target language (i.e., English or Spanish) and the child was allowed to use either Spanish or English in responding. All child responses were recorded in the language in which the child responded for later scoring. Responses were entered into a spreadsheet and scored as correct or incorrect according to guidelines established during the first phase of testing. Responses in either language were considered acceptable. A second scorer checked the responses and item scores. Scoring disagreements were resolved by a third person.

Parent interviews (Gutiérrez-Clellen & Kreiter, 2003) were completed by phone or in person. Parents were asked hour-by-hour what language the child heard and spoke during a typical weekday and weekend. From this report, the percentage of Spanish and English input and output was calculated. Parents were also asked if they had concerns about their child's language development. They were asked to rate their child's grammar, comprehension, and vocabulary skills in each language using a scale from 1 (low proficiency) to 5 (high proficiency). Teacher questionnaires (Gutiérrez-Clellen & Kreiter, 2003) were completed by teachers for children in their classes. These parallel the parent interview and targeted amount of input and output in English and Spanish, concerns about language development, and teacher ratings of language skill (see Table 2).

2.4.1. Determination of language ability

We define language ability as underlying capacity for language learning regardless of language being learned or level of proficiency in a language. This definition is in contrast to language performance which is defined as how well a child scores on a language test in a given language independent of ability. Because there is no gold standard for identification of LI in Latino-bilingual children, we used converging evidence, consistent with current best practice recommendations, to classify children as typical or language impaired (Bedore & Peña, 2008; Kohnert, 2010). Language impairment is considered to be a complex disorder that affects various domains of language (Bartlett et al., 2002; Bishop, 2006; The SLI Consortium, 2002). Adding to this complexity is the variability in children's bilingual language experiences (O'Toole & Hickey, 2013). For such complex disorders, a composite reference standard is thought to be superior to a single-reference standard used in isolation (Rutjes, Reitsma, Coomarasamy, Khan, & Bossuyt, 2007). Thus, a composite reference standard was utilized that included (1) language sampling (Gutiérrez-Clellen, Restrepo, Bedore, Peña, & Anderson, 2000; Hewitt, Hammer, Yont, & Tomblin, 2005; Simon-Cereijido & Gutiérrez-Clellen, 2007); (2) parent; (3) teacher report (Bedore, Peña, Joyner, & Macken, 2011; Gutiérrez-Clellen & Kreiter, 2003); and (4) clinical observation during elicitation of language (Peña, Reséndiz, & Gillam, 2007). Children were identified as having LI if three of four of these primary indicators were consistent with LI.

Language samples were elicited from two wordless picture books (see Gutiérrez-Clellen & Kreiter, 2003; Miller & Iglesias, 2005), and conversational samples were elicited during play with toys. An indicator of LI was more than 20% ungrammatical utterances in the better language (Gutiérrez-Clellen & Kreiter, 2003; Gutiérrez-Clellen et al., 2006; Gutiérrez-Clellen & Simon-Cereijido, 2007; Restrepo, 1998).

Parent or teacher language concern was indicated when the respondent reported that the child had difficulty with language comprehension or language expression that was not attributed solely to articulation difficulties. Previous studies demonstrate significant associations between these parent and teacher questionnaires and children's grammatical productivity in Spanish- and English-speaking school-age (Gutiérrez-Clellen & Kreiter, 2003; Restrepo, 1998) and preschool age children (Bedore et al., 2011).

Clinician observation focused on child responsiveness to the task and transfer from a model to an independent narration during collection of the language and narrative samples. We used a 5-point (from 1 = needs constant support to complete tasks to 5 = needs little to no support to complete tasks) Likert scale based on Peña et al. (2006). Responsivity focused on how verbally responsive children were while telling a story after a model and during the conversational language sample. Transfer focused on the ability to tell a second story independently after the first story task was completed. An indicator of LI was flagged if the child scored a 2 (needed prompts more than 50% of the time) or less during elicitation of stories and conversation.

In 28 cases, primary indicators were missing, so we used up to two secondary measures consistent with standard practice in the field. Mean length of utterance (MLU) in the better language more than 1SD below the mean for a child's age and language (based on Miller and Iglesias' narrative retell SALT database) was a secondary indicator of LI. Parent proficiency rating of 2 (limited proficiency) or lower in the best language was considered a secondary indicator of LI. Identification as language impaired by the school-based speech-language pathologist was also considered a secondary indicator of LI. If three or more of the primary indicators were missing, the case was excluded from the current analysis.

2.4.2. Language experience groups

Children were grouped according to language exposure on the basis of parent and teacher reports. Percentages of input and output were averaged together to yield percentage of language exposure (Bedore et al., 2012; Peña, Gillam, Bedore, & Bohman, 2011). Children were divided into five language groups. Functional monolingual Spanish (FMS) and functional monolingual English (FME) groups had at least 80% exposure to Spanish or English respectively. We consider these two groups to be functional monolingual children because they have limited exposure to the non-dominant language. Children with 60–80% exposure to Spanish (20–40% exposure to English) were classified as bilingual Spanish dominant (BDS). Children with 60–80% exposure to English (20–40% exposure to Spanish) were classified as bilingual English dominant (BDE). The BB children were those who had at least 40% exposure to both Spanish and English.

2.5. Current analysis

For the current analysis we drew all FME, FMS, and BB cases between the ages of 4;0 and 7;0 who had complete testing, language grouping, and language ability data. This included a total of 183 FME and 179 FMS children who were tested in English and Spanish respectively and 78 balanced bilingual speakers who had completed testing in both languages. Distribution by language exposure and age is displayed in Table 1. Table 2 contains parent and teacher proficiency ratings for children in Spanish and English, and Spanish-language exposure data. Individual raw scores were converted to standard scores with a mean of 100 and standard deviation of 15 according to test language and age.

2.5.1. Analysis

We used discriminant functions analysis to evaluate the classification accuracy of the English and Spanish measures. Linear discriminant analyses evaluating group classifications (LI and TD children) were performed on the standard scores in two phases for each language. For each language, test scores from the FME and FMS children were entered into a discriminant analysis to determine optimal cut scores. In the second, cross-validation phase, we tested the extent to which the functions derived from the exploratory datasets predicted group membership of BB children in each language.

We calculated likelihood ratios from the sensitivity and specificity results. According to Hanley and McNeil (1982) and Jaeschke, Guyatt, and Sackett (1994) positive likelihood ratios of 10 or greater and negative likelihood ratios less than .1 are considered large and conclusive and are highly informative; positive likelihood ratios of 5–10 and negative likelihood ratios between .1 and .2 are moderately informative. Positive likelihood ratios between 2 and 5 and negative likelihood ratios between .2 and .5 are modestly informative. When positive likelihood ratios fall below 5 or when negative ratios are over .5, the test results are uninformative.

3. Results

3.1. English semantics measure

Table 3 displays the English BESA semantics means and standard deviations of the TD and LI children divided by language exposure group. We conducted a preliminary ANOVA comparing the means of children from the two ability groups by language group. There was a significant main effect by Ability, $F(1,257) = 104.53$, $p < .001$, $\eta_p^2 = .289$. In addition, the ANOVA indicated significant differences by Language Exposure Group $F(1, 257) = 9.955$, $p = .0018$, $\eta_p^2 = .037$, but the effect size was marginal. The Ability × Language Exposure Group interaction was non-significant $F(1.257) = 0.273$, $p = .602$, $\eta_p^2 = .001$.

Discriminant analysis was conducted with the FME in the exploratory phase, followed by two cross-validation phases (see Table 4). Box's M indicated that the assumption of equality of covariance matrices was violated, although the log determinants were similar (LI = 5.99, TD = 5.35). With large samples this problem is not considered serious (Burns & Burns, 2008; Harris, 2001). The analysis was rerun with separate covariance matrices in the classification. Because the new results did not improve the accuracy rate of the discriminant model, we report the original results. The results of the exploratory analyses yielded significant differences between children with and without LI, $F(1,181) = 112.32$, $p < .001$, and a significant canonical correlation of .619, $p < .001$. This result indicates a large and significant association between the BESA-Semantics English score and language ability. We empirically derived the cut score (87.81) using the mid-point between the two ability group means. The test cut score classified 81% of the cases accurately with 78% sensitivity and 82% specificity. The cross-validated classification with the BB children demonstrated 76% accuracy, with 100% sensitivity and 70% specificity.

A positive likelihood ratio of 4.33 (CI = 2.92–6.41) for the FME group was modestly to moderately informative. The negative likelihood ratio of 0.27 is informative and indicated modest to moderate evidence for the accuracy of a negative result. For the BB group, a positive likelihood ratio of 3.31 was slightly lower than that for the FME group and suggests a modest level of informativeness, but it is less definitive by comparison to the FME group. The negative likelihood ratios of <0.001 indicated very strong evidence for the accuracy of a negative result.

Table 3
Semantics task performance by ability and language exposure group: standard score means (SD).

	English				Spanish			
	FME		BB		BB		FMS	
LI	73.55	(19.98)	66.56	(14.55)	68.54	(10.36)	72.77	(9.94)
TD	102.07	(14.48)	93.31	(15.57)	100.27	(17.42)	97.26	(15.36)

Note: LI, language impaired, TD, typically developing; FME, functional monolingual English; BB, balanced bilingual; FMS, functional monolingual Spanish.

Table 4
Classification accuracy English and Spanish semantics measures by language exposure group.

Language group	Test language						
	Sensitivity (LI as LI)	Specificity (TD as TD)	Correct classification	+Likelihood ratio	95% CI	−Likelihood ratio	95% CI
English							
FME	78%	82%	81%	4.32	2.92–6.41	0.27	0.16–0.46
BB	100%	70%	76%	3.31	2.27–4.83	0.00	Infinite
Spanish							
FMS	83%	80%	81%	4.26	2.96–6.11	0.21	0.09–0.43
BB	93%	89%	90%	8.40	4.12–17.11	0.08	0.01–0.50

Note: LI, cut scores English < 87.81, Spanish < 85.01; LI, language impaired; TD, typically developing; CI, confidence interval.

3.2. Spanish semantics measure

The same procedures were followed for the Spanish semantics measure as described for the English semantics measure. Table 3 displays the means and standard deviations of the TD and LI children divided by language exposure group. Preliminary ANOVA comparing the language ability groups by language exposure was conducted. Results indicate a significant main effect by Ability, $F(1, 253) = 120.423$, $p < .001$, $\eta_p^2 = .322$. There were no differences by Language Exposure group $F(1, 253) = 0.091$, $p = .763$, $\eta_p^2 < .001$ and no Ability by Language Exposure Group interaction $F(1, 253) = 2.163$, $p = .143$, $\eta_p^2 = .008$.

As before, discriminant analysis was conducted in two phases (see Table 4). First the FMS group was used in the exploratory phase, followed by the cross-validation phase. Box's M indicated that the assumption of equality of covariance matrices was violated. Log determinants were similar (LI = 4.59, TD = 5.46). We reran the analysis with separate covariance matrices in the classification. As before, the new results did not improve the classification rate of the discriminant model, so we report the original results. The results of the exploratory analyses yielded significant differences between children with and without LI $F(1,177) = 82.59$, $p < .001$, and a significant canonical correlation of .564, $p < .001$. This result indicates a large and significant association between the BESA-Semantics Spanish score and language ability. We derived the cut score (85.02) empirically, using the mid-point between the two ability group means. The test cut score classified 81% of the FMS cases accurately, with 83% sensitivity and 80% specificity. The cross-validated classification with the BB children demonstrated 90% accuracy with 93% sensitivity and 89% specificity.

A positive likelihood ratio of 4.26 (CI = 2.96–6.11) for the FMS group is modestly to moderately informative. The negative likelihood ratio of 0.21 (CI = .09–.43) is informative and indicates modest evidence for the accuracy of a negative result. The positive likelihood ratio of 8.40 (CI = 4.12–17.11) suggests a high level of accuracy of a positive result for the BB. The negative likelihood ratio of 0.08 (CI = .01–.50) is similarly highly informative.

4. Discussion

The most critical evidence of a measure's validity is the extent to which it accurately classifies clinical and non-clinical groups (Friberg, 2010; Spaulding et al., 2006). For comparison of different groups of children, the degree to which a measure classifies children with similar accuracy is also used as a measure of bias (Valencia & Suzuki, 2001). Thus, the current study had two main goals: (1) to evaluate the diagnostic accuracy of the English and Spanish semantics subtests with monolingual English and Spanish speakers, respectively and (2) to explore whether the same cut points derived for monolingual speakers could be utilized for bilingual Spanish–English children.

4.1. Classification accuracy in Spanish and English: monolingual children

Classification accuracy for the English and Spanish versions of the measure was in the fair range for children in the FME and FMS groups (81% correct classification overall). For English, specificity of 78% is near an acceptable level of 80% and a sensitivity of 82% is considered fair (Dollaghan & Horner, 2011; Plante, 2004; Spaulding et al., 2006). Spanish sensitivity and specificity of 83% and 80% are also considered fair. This level of classification accuracy represents a 3–10% improvement over that which has been reported for most single-word vocabulary tests of English for both receptive and expressive domains alone or used in combination (Gray et al., 1999; Spaulding et al., 2006). While the classification accuracy of our measure did not reach the 90% level, it is possible that continued refinement of these kinds of tasks and additional item analysis will lead to selection of a more robust set of items. Nonetheless, it is an important step in development of a semantics test that can eventually be used diagnostically.

The difficulties that children with LI have on the BESA semantic subtests is consistent with findings that children with LI have difficulty learning new words (Alt & Plante, 2006; Alt & Suddarth, 2012) which may be related to identified delays in development of semantic networks (Sheng & McGregor, 2010; Sheng et al., 2012). This set of semantic items seems to differentially tax the semantic processing systems of children with and without LI.

4.2. Classification accuracy of bilinguals with and without LI

While there were large, significant differences in the standardized scores for children with and without LI, there were also small, but significant differences for FME and BB standard scores in English. Children in the BB group scored slightly lower on the semantics measure compared to FME children. Additionally, accurate classification of BB children using the English-semantics measure was slightly lower for the BB group (76%). The lower accuracy rate for the BB group was due to an unacceptably high false positive rate (30%). These differences in scores between BB and FME on the English measure, although small, appear to have had practical effects on classification accuracy. For the Spanish measure, there were significant differences between children with and without LI, but there were no significant standard score differences between children in the BB and FMS groups. Further, sensitivity and specificity were within acceptable levels for the Spanish semantics measure for the BB and FMS groups.

Recall that children were allowed to respond in either Spanish or English and these responses were taken into account in their (conceptual) score. Thus, they were given the opportunity to demonstrate knowledge in either language regardless of the test language. Nonetheless, at least for the English measure, BB children scored slightly lower than their FME counterparts and these lower scores seem to affect classification accuracy. These findings are consistent with Thordardottir, Rothenberg, Rivard, and Naves (2006) and Mancilla-Martinez et al. (2011) who reported that for typically developing children, conceptual scores improved over single-language scores more in the home language compared to the second language.

We wondered whether, given the similar magnitude of differences for the FME and BB children for TD and LI children if resetting the cut scores would yield increased accuracy. An independent discriminant analysis for the BB group on the English semantics measure resulted in a similar overall classification rate this time with an unacceptably high number of false negatives vs. the high false positive rate when using the FME cut point. Clearly there was a high number of typically developing bilingual children whose scores were low in English. For this group of children, these results indicate that English testing alone is not sufficient to accurately make a diagnosis of language impairment. Further, local norming as has been suggested in the literature (Bracken & Barona, 1991) based on the range of performance is unlikely to be an effective solution because of the overlap in performance between children with and without LI on the English measure.

The higher classification accuracy rate of bilinguals on the Spanish measure compared to the English measure may be due to children's sociolinguistic experiences as indexed to language of exposure. According to teacher and parent reports (see Table 2), the BB group had similar amounts of *current* exposure to English and Spanish. But it is likely that the exposure was distributed between home (Spanish) and school (English). Examination of parent education data shows that 63% of the fathers and 60% of the mothers of children in the BB group received their education in Latin America, which is likely to be associated with Spanish use at home. Therefore it is possible that the BB children as a group had more cumulative exposure to Spanish. It is likely that children's performance in Spanish was more stable relative to their English language performance allowing for Spanish language testing to be more diagnostically accurate in this developmental window.

Another related possibility is that the BB children may have been more familiar with the content presented on the Spanish measure than on the English measure (Bedore et al., 2011). Bilingual children have different experiences in each of their two languages that lead to learning words and concepts unique to each language (Pearson, Fernández, & Oller, 1995). Indeed, children's home cultural experiences influence how they respond linguistically to various tasks (Greenfield, Keller, Fuligni, & Maynard, 2003; Luykx et al., 2007). Thus, conceptual scoring alone narrowed but did not entirely close the gap between monolingual and bilingual performance in English the way it did in Spanish.

4.3. Clinical implications

Examination of the group means by language exposure and ability shows similarities in Spanish, but differences in English, for monolingual and bilingual groups. Nonetheless, overall classification accuracies were generally fair to good across language exposure groups with the exception of specificity for the BB on the English measure. This finding provides evidence of the clinical utility of the BESA.

For bilinguals, consistent with findings for English, testing may be able to rule out LI, but cannot rule it in (Kohnert et al., 2006). The use of conceptual scoring seems to equalize performance for monolingual and bilingual children in the home language, though not completely in English.

On the basis of these findings we recommend use of conceptual scoring in testing semantics bilinguals. English testing can be used to rule out language impairment if the child scores within normal limits. But, if they score in the impairment range, the home language (using conceptual scoring) would need to additionally be tested in order to make an accurate diagnostic decision. Additional testing for making a diagnostic decision should, consistent with recommended practices, include assessment of morphosyntax, spontaneous conversational and narrative samples, as well as parent and teacher report in addition to examination of children's semantic skill.

Acknowledgements

Data collection for this project was supported in part by a grant from the National Institutes of Health (NIDCD) Contract # N01-DC-8-2100. We thank all of the interviewers and testers for their assistance in collecting the data for this project and the school districts for allowing us access to collect the data. Data for this project were contributed by Aquiles Iglesias, Vera Gutierrez-Clellen, and Brian Goldstein.

Appendix A. Continuing Educations Questions

1. Bilingual children may show below-average performance on vocabulary tests because:
   a. Lack of exposure
   b. Language impairment
   c. Lack of exposure and/or language impairment
   d. Semi-lingualism
2. Lexical-semantic measures of language for bilinguals are useful because:
   a. they are widely available to SLPs
   b. they test one language at a time
   c. they can test both languages at once
   d. they go beyond testing single word knowledge
3. Single word vocabulary tests may be of limited value for identification of language impairment in bilinguals because:
   a. They tend to have poor classification accuracy with monolinguals and it's likely that this would be the case for bilinguals
   b. Bilinguals can be tested in either their home language or English and the scores are interchangeable
   c. Translation of the vocabulary items allow for direct comparison between the two test languages
   d. They are highly reliable indicators of performance
4. Conceptual scoring is
   a. Adding up correct items in either language
   b. Counting all known concepts (but translation equivalents only once)
   c. A good way to compare what a child knows in one or both languages
   d. A good way to determine language of dominance
5. In the current study, the semantics subtest of an experimental measure of bilingual language impairment:
   a. Showed poor classification accuracy for bilinguals but not monolinguals
   b. Showed good classification accuracy for bilinguals and monolinguals on the Spanish measure
   c. Showed good classification accuracy for bilinguals and monolinguals on the English measure
   d. Showed poor classification accuracy for monolinguals but not bilinguals

References

Allen, M., & Yen, W. (1979). *Introduction to measurement theory*. Belmont, CA: Wadsworth.

Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 49*, 941–954.

Alt, M., & Suddarth, R. (2012). Learning novel words: Detail and vulnerability of initial representations for children with specific language impairment and typically developing peers. *Journal of Communication Disorders, 45*, 84–97. http://dx.doi.org/10.1016/j.jcomdis.2011.12.003

Bartlett, C. W., Flax, J. F., Logue, M. W., Vieland, V. J., Bassett, A. S., Tallal, P., et al. (2002). A major susceptibility locus for specific language impairment is located on 13q21. *American Journal of Human Genetics, 71*, 45–55.

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism, 11*(1), 1–29.

Bedore, L. M., Peña, E. D., García, M., & Cortez, C. (2005). Conceptual versus Monolingual Scoring: When does it make a difference? *Speech, Language, Hearing Services in Schools, 36*, 188–200.

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. http://dx.doi.org/10.1016/j.jcomdis.2010.05.002

Bedore, L. M., Peña, E. D., Joyner, D., & Macken, C. (2011). Parent and teacher rating of bilingual language proficiency and language development concerns. *International Journal of Bilingual Education and Bilingualism, 14*(5), 489–511. http://dx.doi.org/10.1080/13670050.2010.529102

Bedore, L. M., Peña, E. D., Summers, C., Boerger, K., Resendiz, M., Greene, K., et al. (2012). The measure matters: Language dominance profiles across measures in Spanish/English bilingual children. *Bilingualism: Language and Cognition, 15*, 616–629.

Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language, Speech, and Hearing Services in Schools, 44*, 133–146. http://dx.doi.org/10.1044/0161-1461(2012/12-0093)

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*, 525–531. http://dx.doi.org/10.1017/s1366728909990423

Bishop, D. V. (2006). What causes specific language impairment in children? *Current Directions in Psychological Science, 15*, 217–221.

Bishop, D. V. M., & Adams, C. (1991). What do referential communication tasks measure? A study of children with specific language impairment. *Applied Psycholinguistics, 12*, 199–215.

Bracken, B. A., & Barona, A. (1991). State of the art procedures for translating, validating and using psychoeducational tests in cross-cultural assessment. *School Psychology International, 12*, 119–132.

Brackenbury, T., & Pye, C. (2005). Semantic deficits in children with language impairments: Issues for clinical assessment. *Language, Speech, and Hearing Services in Schools, 36*, 5–16.

Brownell, R. (2000). *Expressive One Word Picture Vocabulary Test – 2000 edition*. Novato, CA: Academic Therapy Publications.

Brownell, R. (2001). *Expressive One Word Picture Vocabulary Test – Spanish-Bilingual edition*. Novato, CA: Academic Therapy Publications.

Burns, R., & Burns, R. (2008). *Business research methods and statistics using SPSS*. London: Sage.

Calderón, M., August, D., Slavin, R., Duran, D., Madden, N., Cheung, A., et al. (2005). Bringing words to life in classrooms with English-language learners. In *Teaching and learning vocabulary: Bringing research to practice* (pp. 115–136). Mahwah, NJ, USA: Lawrence Erlbaum Associates Publishers.

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., et al. (2004). Closing the gap: Addressing the vocabulary needs of English language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39*, 188–215.

Dale, P., Bates, E., Reznick, J. S., & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language, 16*, 239–249.

Danahy, K., Windsor, J., & Kohnert, K. J. (2007). Counting span and the identification of primary language impairment. *International Journal of Language & Communication Disorders, 42*(3), 349–365. http://dx.doi.org/10.1080/13682820600940083

DeVellis, R. (1991). *Scale development: Theory and applications* (Vol. 26). Newbury Park, CA: SAGE Publications.

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research, 54*, 1077–1088. http://dx.doi.org/10.1044/1092-4388(2010/10-0093)

Dunn, L., & Dunn, L. (1981). *Peabody Picture Vocabulary Test – Revised*. Circle Pines, MN: American Guidance Service.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test – Third edition*. Circle Pines, MN: American Guidance Service.

Dunn, L., Padilla, R., Lugo, S., & Dunn, L. (1986). *Test de vocabulario en imagenes peabody*. Circle Pines, NM: American Guidance Service.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2006). *The Macarthur-Bates Communicative Inventories*. Baltimore: Brookes.

Friberg, J. C. (2010). Considerations for test selection: How do validity and reliability impact diagnostic decisions? *Child Language Teaching and Therapy, 26*, 77–92. http://dx.doi.org/10.1177/0265659009349972

Gray, S. (2004). Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research, 47*, 1117–1132.

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*, 196–206.

Greenfield, P. M., Keller, H., Fuligni, A., & Maynard, A. (2003). Cultural pathways through universal development. *Annual Review of Psychology, 54*, 461–490.

Gutiérrez-Clellen, V. F., & DeCurtis, L. (2001). Examining the quality of children's stories: Clinical applications. *Seminars in Speech and Language, 22*(01), 079–088.

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics, 24*, 267–288.

Gutiérrez-Clellen, V. F., Restrepo, M. A., Bedore, L. M., Peña, E. D., & Anderson, R. T. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech & Hearing Services in the Schools, 31*(1), 88–98.

Gutiérrez-Clellen, V. F., Restrepo, M. A., & Simon-Cereijido, G. (2006). Evaluating the discriminant accuracy of a grammatical measure with Spanish-speaking children. *Journal of Speech, Language, and Hearing Research, 49*, 1209–1223.

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*, 968–981.

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2009). Using language sampling in clinical assessments with bilingual children: Challenges and future directions. *Seminars in Speech and Language, 30*, 234–245. http://dx.doi.org/10.1055/s-0029-1241722

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2010). Using nonword repetition tasks for the identification of language impairment in Spanish–English-speaking children: Does the language of assessment matter? *Learning Disabilities Research & Practice, 25*(1), 48–58. http://dx.doi.org/10.1111/j. 1540-5826.2009.00300.x

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29–36.

Harris, R. J. (2001). *A primer of multivariate statistics* (3rd ed.). Mahway, NJ: Erlbaum.

Hewitt, L. E., Hammer, C. S., Yont, K. M., & Tomblin, J. B. (2005). Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. *Journal of Communication Disorders, 38*, 197–213.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development, 73*(2), 418–433.

Hoff, E., & Tian, C. (2005). Socioeconomic status and cultural influences on language. *Journal of Communication Disorders, 38*, 271–278.

Jackson-Maldonado, D. (2003). *MacArthur inventarios del desarrollo de habilidades comunicativas: User's guide and technical manual*. Paul H Brookes Pub Co.

Jaeschke, R., Guyatt, G. H., & Sackett, D. L. (1994). Users' guides to the medical literature. *Journal of the American Medical Association, 271*, 703–707.

Jordaan, H., Shaw-Ridley, G., Serfontein, J., Orelwitz, K., & Monaghan, N. (2001). Cognitive and linguistic profiles of specific language impairment and semantic–pragmatic disorder in bilinguals. *Folia Phoniatrica et Logopaedica, 53*, 153–165.

Kohnert, K. J. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders, 43*, 456–473. http://dx.doi.org/10.1016/j.jcomdis.2010.02.002

Kohnert, K. J., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research & Practice, 21*, 19–29.

Luykx, A., Lee, O., Mahotiere, M., Lester, B., Hart, J., & Deaktor, R. (2007). Cultural and home language influences on children's responses to science assessments. *Teachers College Record, 109*, 897–926.

Mancilla-Martinez, J., Pan, B. A., & Vagh, S. B. (2011). Assessing the productive vocabulary of Spanish–English bilingual toddlers from low-income families. *Applied Psycholinguistics, 32*, 333–357. http://dx.doi.org/10.1017/s0142716410000433

McGregor, K. K. (2009). Semantics in child language disorders. In *Handbook of child language disorders* (pp. 365–387). New York, NY, USA: Psychology Press.

McGregor, K. K., & Windsor, J. (1996). Effects of priming on the naming accuracy of preschoolers with word-finding deficits. *Journal of Speech & Hearing Research, 39*, 1048–1058.

Miller, J., & Iglesias, A. (2005). *Systematic analysis of language transcripts – Salt v9. Language analysis Laboratory Waisman Center*. Madison: University of Wisconsin.

Nippold, M. A., Erskine, B. J., & Freed, D. B. (1988). Proportional and functional analogical reasoning in normal and language-impaired children. *Journal of Speech and Hearing Disorders, 53*, 440–448.

O'Toole, C., & Hickey, T. M. (2013). Diagnosing language impairment in bilinguals: Professional experience and perception. *Child Language Teaching and Therapy, 29*, 91–109. http://dx.doi.org/10.1177/0265659012459859

Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics, 28*, 191–230.

Pearson, B. Z. (2010). We can no longer afford a monolingual norm. *Applied Psycholinguistics, 31*, 339–343. http://dx.doi.org/10.1017/s014271640999052x

Pearson, B. Z., & Fernández, S. C. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning, 44*, 617–653.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*, 93–120.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of Child Language, 22*, 345–368.

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

Peña, E. D., Bedore, L. M., & Zlatic-Guinta, R. (2002). Category-generation performance of bilingual children: The influence of condition, category, and language. *Journal of Speech, Language & Hearing Research, 45*, 938–947.

Peña, E. D., Gillam, R. B., Bedore, L. M., & Bohman, T. M. (2011). Risk for poor performance on a language screening measure for bilingual preschoolers and kindergarteners. *American Journal of Speech-Language Pathology, 20*(4), 302–314. http://dx.doi.org/10.1044/1058-0360(2011/10-0020)

Peña, E. D., Gillam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., et al. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research, 49*(5), 1037–1057.

Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B., & Bedore, L. M. (2008). *Bilingual English Spanish Assessment-Experimental Version*.

Peña, E. D., & Quinn, R. (1997). Task familiarity: Effects on the test performance of Puerto Rican and African American children. *Language, Speech, and Hearing Services in Schools, 28*(4), 323–332.

Peña, E. D., Reséndiz, M., & Gillam, R. B. (2007). The role of clinical judgments of modifiability in the diagnosis of language impairment. *Advances in Speech Language Pathology, 9*(4), 332–345.

Plante, E. (2004). Evidence based practice in communication sciences and disorders. *Journal of Communication Disorders, 37*, 389–390.
Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 1398–1411.
Restrepo, M. A., Schwanenflugel, P. J., Blake, J., Neuharth-Pritchett, S., Cramer, S. E., & Ruston, H. P. (2006). Performance on the PPVT-III and the EVT: Applicability of the measures with African American and European American preschool children. *Language Speech Hearing Services in Schools, 37*, 17–27. http://dx.doi.org/10.1044/0161-1461(2006/003)
Restrepo, M. A., & Silverman, S. W. (2001). Validity of the Spanish preschool language scale-3 for use with bilingual children. *American Journal of Speech-Language Pathology, 10*, 382–393.
Rutjes, A. W., Reitsma, J., Coomarasamy, A., Khan, K., & Bossuyt, P. (2007). Evaluation of diagnostic tests when there is no gold standard: A review of methods. *Health Technology Assessment, 11*. ix-51.
Sheng, L., Bedore, L. M., Peña, E. D., & Taliancich-Klinger, C. (2013). Semantic convergence in Spanish-English bilingual children with primary language impairment. *Journal of Speech, Langauge and Hearing Research, 56*, 766–777.
Sheng, L., & McGregor, K. K. (2010a). Lexical-semantic organization in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 53*, 146–159.
Sheng, L., & McGregor, K. K. (2010b). Object and action naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 53*, 1704–1719.
Sheng, L., McGregor, K. K., & Marian, V. (2006). Lexical-semantic organization in bilingual children: Evidence from a repeated word association task. *Journal of Speech, Language, and Hearing Research, 49*, 572–587.
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 & Hearing Research, 55*, 1–15.
Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2007). Spontaneous language markers of Spanish language impairment. *Applied Psycholinguistics, 28*, 317–339. http://dx.doi.org/10.1017/s0142716407070166
Simonsen, A. (2002). *Naming amongst 6 year old children. Working papers* (Vol. 50, pp. 79–84). Department of Linguistics, Lund University.
Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Speech, Language, Hearing Services in Schools, 37*, 61–72.
Tamayo, J. (1987). Frequency of use as a measure of word difficulty in bilingual vocabulary test construction and translation. *Educational & Psychological Measurement, 47*(4), 893–902.
The SLI Consortium (2002). A genomewide scan identifies two novel loci involved in specific language impairment. *American Journal of Human Genetics, 70*, 384–398.
Thordardottir, E., Rothenberg, A., Rivard, M.-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders, 4*, 1–21. http://dx.doi.org/10.1080/14769670500215647
Umbel, V., Pearson, B. Z., Fernández, M. C., & Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development, 63*, 1012–1020.
Valencia, R., & Suzuki, L. (2001). *Intelligence testing and minority students: Foundations, performance factors, and assessment issues*. Thousand Oaks, CA: Sage.
Weckerly, J., Wulfeck, B., & Reilly, J. (2001). Verbal fluency deficits in children with specific language impairment: Slow rapid naming or slow to name? *Child Neuropsychology, 7*, 142–152.
Windsor, J., Kohnert, K. J., Lobitz, K. F., & Pham, C. T. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology, 19*, 298–310. http://dx.doi.org/10.1044/1058-0360(2010/09-0064)