Published in final edited form as:

Int J Lang Commun Disord. 2016 March ; 51(2): 192–202. doi:10.1111/1460-6984.12199.

Assessment of language impairment in bilingual children using semantic tasks: two languages classify better than one

Elizabeth D. Peña, Lisa M. Bedore, and Ellen S. Kester

Abstract

Background—Significant progress has been made in the identification of language impairment in children are bilingual. Bilingual children’s vocabulary knowledge may be distributed across languages. Thus, when testing bilingual children it is difficult to know how to weigh each language for diagnostic purposes. Even when conceptual scoring is used in vocabulary testing, bilingual children may score below that of their typical monolingual peers.

Aims—The primary aim was to evaluate the classification accuracy of two approaches (total semantics score and two-dimensional bilingual coordinate score) that combined lexical–semantic knowledge across two languages. We investigated the classification accuracy of the English and Spanish semantics subtest using the experimental version of the Bilingual English Spanish Assessment (BESA) with bilingual children with and without language impairment.

Methods—A total of 78 bilinguals with balanced exposure to English and Spanish (15 with language impairment, 63 with typical development) participated. Children were between 4;0 and 6;11 years old. Discriminant function analysis explored the extent to which these children were accurately classified when combining Spanish and English subtests.

Outcomes & Results—Discriminant analysis yielded above 85% correct classification for balanced bilingual children for both approaches.

Conclusions & Implications—For the most accurate assessment and diagnostic decision-making for bilinguals, approaches that consider both languages together are recommended.

Keywords

bilingual; language impairment; assessment; semantics

Introduction

Assessment in two languages is considered the standard for assessment of bilingual children referred for speech–language testing (Bedore and Peña 2008, American Speech–Language–Hearing Association 1985, Kohnert 2010). Generally, it is agreed that bilingual children with true language impairment (LI) would present patterns associated with the diagnosis in both their languages (Caesar and Kohler 2007, Thordardottir *et al.* 2006). To date, however, no

Address correspondence to: Elizabeth D. Peña; lizp@mail.utexas.edu.

Declaration of interest: Elizabeth Peña and Lisa Bedore are authors of Bilingual English Spanish Assessment (BESA) and receive royalties from its sales.

empirically validated procedures guide clinicians for how to combine information from two languages in order to make a diagnostic decision. In this analysis we investigated the feasibility of combining the scores from both the children's languages to assess language ability in bilingual children using the experimental version of the semantics subtest from the Bilingual English Spanish Assessment (BESA) (Peña *et al.* 2014) which incorporates semantics-depth tasks including description, comparisons, analogies and category generation.

Children with LI by definition demonstrate delays in language relative to their typically developing (TD) peers (Bishop 1992, Leonard *et al.* 2003). Much of the focus on identification of LI focuses on the morphosyntactic domain for both monolinguals and bilinguals (Bedore and Leonard 2005, Bortolini *et al.* 2006, Gutiérrez-Clellen *et al.* 2008, Jacobson and Schwartz 2002, Leonard *et al.* 2003). However, these children also have documented difficulties in the lexical-semantic domain. Specifically, children with LI demonstrate delays in vocabulary acquisition (Girolametto *et al.* 2001, McGregor 2009, Nash and Donaldson 2005). Given these documented delays, single-word expressive and/or receptive vocabulary tests are often used as a part of a test battery to identify LI (Betz *et al.* 2013). Yet, children with LI appear to catch up with their TD peers and often score within the normal range on such measures (McGregor 2009). Single-word tests most often do not yield good classification accuracy. Gray *et al.* (1999) compared the performance of 62 children with and without LI on four vocabulary tests (PPVT-III; Dunn and Dunn 1997; EOWPVT-R; Gardner 1990); ROWPVT; Gardner 1985; and EVT; Williams 1997). Although children with LI scored significantly below their age-matched TD peers, classification accuracy was poor. Sensitivity ranged between 71% and 77% and specificity was between 68% and 77%. While single-word vocabulary is a relative strength for children with LI, they appear to have more difficulty in the area of semantic depth.

With respect to semantic depth, children with LI have difficulty in understanding the meaning of words and relationships between words (Gray 2004, Jordaan *et al.* 2001, McGregor 2009). Children with specific language impairment produce semantic and phonological errors when asked to produce repeated associations (such as tell me a word that goes with 'car' repeated three times to elicit three different associations) (Sheng and McGregor 2010). Bilingual children with LI, like their monolingual peers, demonstrate low performance on the repeated associations task relative to their peers matched on age of first experience and amount of use of each language indicated reduced depth of semantic knowledge (Sheng *et al.* 2012, 2013). Given the extent of lexical-semantic difficulties, we focus on the utility of a semantic measure (rather than single-word vocabulary) for diagnostic decision-making.

Accounting for distributed knowledge in two languages

Lexical-semantic tasks are especially appealing for assessment of bilinguals because they are easily translated and take little time to administer. Among the more frequently used language measures by speech-language pathologists, four of the top 10 are vocabulary measures (Betz *et al.* 2013). These four include the PPVT-IV (Dunn and Dunn 2012); EOWPVT-III (Brownell 2010); ROWPVT-II (Brownell 2000); and EVT-II (Williams 2007).

Of the four, three (EOWPVT-III, PPVT-IV and ROWPVT-II) have Spanish translations or adaptations. Compared with adaptation of standardized measures of morphosyntax, vocabulary measures can be translated with less concern about retaining the meaning and word order, although word frequency may affect item difficulty (Arnold and Matus 2000, Patricacou *et al.* 2007, Peña 2007, Stansfield 2003).

Bilingual children learn vocabulary to meet demands associated with two languages. Because these linguistic demands may differ by context it is likely they have not learned the same vocabulary in both their two languages. Divided knowledge may lead to lower vocabulary scores in each of their two languages (Oller *et al.* 2007). Approaches that account for vocabulary knowledge across two languages may be better indices of language learning ability in bilinguals. For example, in a recent study of simultaneous bilingual toddlers, Marchman *et al.* (2010) found that speed of language processing was associated with vocabulary knowledge overall. Specifically they found that children who knew more words (combining English and Spanish) demonstrated faster processing speed in at least one of their languages compared with children who knew fewer words overall.

Approaches to assessment of two languages includes testing in one language but allowing for responses in either language and testing in two languages. Conceptual scoring is one option that combines children's vocabulary knowledge across languages (Pearson and Fernández 1994). The advantage of this approach is that bilingual children can respond in either of their two languages, or translated versions can be compared and merged to obtain a conceptual score. Conceptual scores, which give credit to word knowledge across a child's two languages, seem to capture better the bilingual's breadth of vocabulary. Studies comparing single-language vocabulary and conceptual vocabulary scores demonstrate that bilingual children's scores are more similar to monolingual comparisons when using conceptual scoring (Thordardottir *et al.* 2006, Junker and Stockman 2002, Core *et al.* 2013).

Conceptual scores have not been compared specifically among bilingual children with and without LI, but they potentially could reduce classification errors. Bedore *et al.* (2005) compared bilingual single-language and conceptual scores with single-language scores of monolinguals on a measure of semantics. A greater proportion of the TD bilingual children's scores fell below the range of scores achieved by TD monolingual children when single-language scores were used. Bilingual children's conceptual scores, however, were more comparable with the single-language scores of monolingual children. Thordardottir *et al.* (2006) evaluated performance of French–English bilingual toddlers on French (Frank *et al.* 1997, Trudeau *et al.* 1999) and English (Fenson *et al.* 1993) versions of the MacArthur Bates Communicative Developmental Inventories (CDI) and on the Peabody Picture Vocabulary Test (Dunn and Dunn 1997) and its French adaptation (Dunn *et al.* 1993). In general, conceptual scores were comparable with the French monolingual norm, but below those of their English norm. Examination of individual scores among these TD bilingual children showed considerable variation in each language. Therefore, the conceptual scoring approach of accepting responses in either language on a single-language administration may not be sufficient to overcome the limitations of single-language approaches. Similarly, de Abreu *et al.* (2012) compared conceptual and single-word vocabulary knowledge on a translated (from English) version of the EOWPVT. Findings demonstrated that Portuguese–

Luxembourgish bilinguals demonstrated lower single-language scores and conceptual scores on the EOW-PVT compared with monolingual Portuguese children. Words represented on a developmental vocabulary test typically reflect those words that most children speaking that language know at that age. If the translation equivalents are of different frequency or familiarity in the other language, it is possible that children may not know all words with the same familiarity in either language depending on the test's language. For example, on the Spanish (Jackson Maldonado *et al.* 2001) and English (Fenson *et al.* 1993) versions of the CDI, 'book' and 'egg' are known by 90% and 66% of 24-month-old English speakers respectively, but their translation equivalents 'libro' and 'huevo' are known by 41% and 81% of the 24-month-old Spanish speakers in the norm. Thus, if a Spanish–English bilingual child does not know the high-frequency word 'book' when tested in English it is less likely that they would know 'libro', which has a lower frequency in Spanish.

In a recent study, Peña *et al.* (2015) found that bilingual and functional monolingual (those with less than 20% daily exposure to another) children with and without LI scored significantly different on the experimental version of the BESA (Peña *et al.* 2014) semantics subtest. Classification accuracy was fair (80% or above) to good (90% or above) for both English and Spanish versions of the test for functional monolingual and bilingual children. However, classification accuracy was slightly lower for the bilingual children on the English version of the task even though conceptual scoring was used. These results are consistent with those reported by Thordardottir *et al.* (2006). Like Thordardottir *et al.*, Peña *et al.* combined scores by counting overlapping items only once in a conceptual score. They compared bilingual and functional monolingual children's performance using the conceptual score. As with Thordardottir *et al.*'s French–English bilinguals whose conceptual scores were most comparable with those of French monolinguals, Peña *et al.* found their bilingual participant's scores to be more like those of their functional monolingual Spanish group than those of the functional monolingual English group. In both studies, the children performed more similarly to functional monolinguals of their home language than they did to functional monolinguals of their second language.

The use of conceptual scoring to assess bilinguals is not without limitations. While it is used in part to not overinflate estimates of children's vocabulary knowledge (Pearson *et al.* 1993, 1995), conceptual scoring may underestimate bilinguals' vocabulary knowledge. For example, Bedore *et al.* (2005) speculated that even though bilinguals may have translation equivalents in their repertoire they may use these in very different ways and to mean different things. Some researchers have argued that discounting cross-linguistic synonyms fails to acknowledge that children must learn two form-meaning mappings for each concept (Core *et al.* 2013, Pearson 1998).

Two options for scoring and comparing children's lexical–semantic knowledge in two languages include use of a *total semantics score* and what we call a *two-dimensional bilingual coordinate score*. Total vocabulary is the sum of items known across two (or more) languages regardless of whether they represent translation equivalents. The two-dimensional bilingual coordinate score represents scores from each language as a coordinate in an x – y plane. This two-dimensional bilingual coordinate score is explained below.

Total vocabulary has the advantage of representing a bilingual child's vocabulary across their two languages regardless of the degree of overlap in meaning. Total vocabulary may better capture what bilingual children know and the different contexts in which they use translation equivalents (Pearson 1998). In a recent study comparing young (22–30 months) bilingual children's total and conceptual vocabularies with a monolingual control over time, total vocabulary was the more comparable with the monolingual scores at all three time points (Core *et al.* 2013). Additionally, the proportion of children falling below the 25th percentile was most similar when the total vocabulary score was used. For children in this age range it seems that the total vocabulary score is most analogous to monolingual vocabulary. These findings are consistent with Thordardottir *et al.* (2006) who found that for French–English bilinguals total vocabulary was most comparable with that of English monolinguals using the English version of the CDI.

The two-dimensional bilingual coordinate score focuses on simultaneous analysis of the two languages. This method allows for consideration of the relative strength in each language while accounting for the contribution of the weaker language. This approach can also incorporate use of the non-target language via a conceptual score in each of the languages tested and the scores obtained for each language version are compared. Bilinguals' performance can be examined in each of their two languages (A and B), using an x – y graph so that performance in both languages is represented by a point or coordinate on a graph as an ordered pair $AB_n(x, y)$ (figure 1). In bilingual acquisition research, Uccelli and Pérez (2007) charted performance in both languages in two-dimensional space to show how the patterns of performance in two languages change over time. Additionally, this notion of testing in both languages and examining cut points in both simultaneously has been used to verify LI in bilinguals (Paradis *et al.* 2003) and to identify TD bilingual children (Kohnert *et al.* 2006). But this approach has not been empirically validated. As seen in figure 1, we can plot children's A and B language scores as ordered pairs, where language A is represented on the x -axis and language B is represented on the y -axis. We envision that TD children with uneven levels of language proficiency in each language would fall in the upper left and lower left quadrants (e.g., AB_1 and AB_3 on figure 1) and that TD children with balanced proficiency would fall in the upper right quadrant (e.g., AB_2). Bilingual children with LI would demonstrate low performance in both languages (e.g., AB_4).

In the current study we follow-up on our finding in Peña *et al.* (2015) that classification accuracy for balanced bilingual children was slightly lower than for monolinguals. Specifically, we explore whether combining performance in two languages, using a total score and a two-dimensional bilingual coordinate score improves classification accuracy using the 2008 experimental version of the semantics subtest from the BESA (Peña *et al.* 2014). It is possible that the total score across both languages is more representative of bilingual children's knowledge and experiences. Such an approach may improve classification accuracy because it considers both languages and language context(s). Similarly a two-dimensional bilingual coordinate approach samples from both the L1 and the L2. In this approach we reasoned that low scores in one language might be offset by higher scores in the other language as compared with using a single-language approach as in our previous study. Specific research questions included:

- What is the diagnostic accuracy on the English and Spanish semantics subtests of the BESA experimental version for bilingual children when combined as a total score?
- What is the diagnostic accuracy on the English and Spanish semantics subtests of the BESA experimental version for bilingual children when combined as a simultaneous two-language score?

Methods

Participants

The current study focuses on 78 bilingual children (15 with LI) who completed testing in both Spanish and English. These children had between 40% and 60% exposure to English and Spanish, as reported by parents and teachers. These 78 bilingual cases were selected from the normative and validation samples ($n = 785$) used for development of the semantics subtest of the BESA. Children were from Texas, California and Pennsylvania. Language exposure for the normative sample ranged from 0% to 100% in Spanish and English. We focus on this group of children who hear and use both languages 40–60% of the time because they are likely to offer the most robust test of the utility of dual language testing. Unlike their Spanish or English dominant bilingual peers, if they have used both languages from birth in a balanced manner then their performance in both languages should be informative. If they have more recently moved into this balanced range they are likely to demonstrate asynchronies in their language skills. It may not be known which language will be the most informative for decision making regarding language status.

Instrument

The BESA semantics subtest has six item types: characteristic properties (e.g., tell me three things about X), categorization (e.g., tell me all the foods you can think of), functions (e.g., what is X used for?), linguistic concepts (e.g., what size is the X?), verbal analogies (e.g., X is to Y as A is to), comprehension of passages (e.g., who, what and where questions), and similarities and differences (e.g., what is different about these two Xs?) (Peña *et al.* 2003, Bedore *et al.* 2005). The experimental version of the semantics measures employed in the current study consisted of 48 Spanish and 49 English items. For each language, we selected items so that they were similarly distributed by item difficulty using classical test theory (Allen and Yen 1979). Items were those that demonstrated the best item discrimination based on a preliminary analysis of a larger item set of 86 items in each language. The item sets were psychometrically similar and were not translation equivalents. That is, distribution of item difficulty was similar across the two languages. Table 1 displays the distribution of items by type and difficulty level. Cronbach's alpha for the English item set was .88 and .85 for the Spanish item set, indicating good internal consistency for each target language (DeVellis 1991).

Procedures

Determination of language ability—Language ability was determined on the basis of language sample analysis, parent and teacher interviews, and clinical observation during

elicitation of language. 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 (Gutiérrez-Clellen and Kreiter 2003, Miller and Iglesias 2005), and conversational samples were elicited during play with toys. An indicator of LI was given when there were more than 20% ungrammatical utterances in the language in which they had a higher percentage of grammaticality based on 100-utterance samples combining narrative and conversational samples (Restrepo 1998, Gutiérrez-Clellen and Kreiter 2003, Gutiérrez-Clellen *et al.* 2006, Gutiérrez-Clellen and Simon-Cerejido 2007).

Parent and teacher interviews (Gutiérrez-Clellen and Kreiter 2003) were completed by telephone or in person to determine possible concerns about language development and language exposure. The concern questions were treated as an indicator of LI. Specifically, parents and teachers were asked if they were concerned about their child's language development or comprehension. Follow-up questions were used to rule out second language acquisition or articulation difficulties as the source of concern.

During elicitation of the language sample, clinicians focused on child responsiveness and transfer. A five-point Likert scale (1 = needs constant support to complete tasks, 5 = needs little to no support to complete tasks) based on Peña *et al.* (2006a) was completed immediately after eliciting the sample. Responsiveness focused on whether the child was verbally responsive in telling a story after a model and during conversation. Transfer focused on the children's facility in telling a second story independently after retelling a story after a model. 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. Table 2 displays the means for LI and TD groups on each of these indicators.

If one or two of the primary indicators were missing, additional measures were used. If three or more of the primary indicators were missing, the case was excluded from the current analysis. Additional indicators of impairment included mean length of utterance (MLU), parent and/or teacher proficiency rating, and identification by the school-based speech-language pathologist as having LI.

Specifically, MLU in the two languages was compared. If the higher MLU of the two languages was more than 1 SD (standard deviation) below the mean for a child's age and language (based on Miller and Iglesias's 2005 narrative retell SALT database) this was considered an indicator of LI. On the questionnaires, parents and teachers were asked to rate children's grammar, comprehension and vocabulary skills in each language using a scale from 1 (low proficiency) to 5 (high proficiency). Parent and/or teacher proficiency ratings were considered an indicator of LI if they scored child proficiency as 2 (limited proficiency with grammatical errors, limited vocabulary, understands the general idea) or below. Previous studies have found that parents and teachers are able to make good judgments of children's language proficiency (depending on language) in school-age (Gutiérrez-Clellen and Kreiter 2003) and preschool-age children (Bedore *et al.* 2011). Identification as

language impaired by the school-based speech–language pathologist was also considered a secondary indicator of LI.

Language use and exposure—Parents were asked hour by hour what language the child heard and spoke during a typical weekday at school and home and weekend at home. Percentage of Spanish and English input and output was calculated on the basis of this report. If the use and exposure average was between 40% and 60% in both languages, they were identified as balanced bilingual and were selected into the current study. Indexing language use and exposure rather than a direct measure of proficiency helps to decouple proficiency and ability.

Individual testing—Testing included individualized testing of each child in Spanish and English. Order of language testing (Spanish versus English) was randomized. Both the gold standard measures (language sampling) and the experimental measures were given to each child over a 2–3-week period. Testing was completed in one language before testing in the other language began. Language sampling included conversational samples and two narratives in each language for a total of 100 utterances. Experimental measures included semantics, morphosyntax, phonology and pragmatics. On the semantics subtest, items were presented in the target language, but children were allowed to respond in either language for conceptual scoring.

Testers and authors were blind to child language ability at the time of testing and scoring. Children’s responses were recorded verbatim in the language of response. Responses were entered into a spreadsheet and each item was scored as correct or incorrect according to established guidelines. A second scorer checked the responses and item scores item by item. Scoring disagreements were resolved by a third person.

Conversion to standard scores—Raw scores were converted to standard scores on the basis of a normative sample of 544 TD children at 6-month intervals, with a mean of 100 and SD of 15. Spanish norms included scores of children who were functional monolingual Spanish (less than 20% exposure to English) and bilingual dominant in Spanish (between 20% and 40% exposure to English). Children who demonstrated balanced bilingualism (between 40% and 60% exposure to both English and Spanish) were included in the norm if their Spanish score was higher than English. English norms were derived in the same way, including children who were functionally monolingual English speakers, and dominant English speakers. Those who were balanced bilingual were included if their English test score was higher than Spanish. Balanced bilinguals were included in both the English and Spanish norms if their scores were within 10% of each other.

Analysis—We used discriminant functions analysis to evaluate the classification accuracy of the English and Spanish measures together in two analyses. For the first, we summed the standard scores across English and Spanish and entered these scores into a discriminant analysis. Standard scores were used in this analysis because they control for age differences and they are equal interval scores (Salvia *et al.* 2009). For the two-dimensional bilingual coordinate approach, we used the cut scores derived for functional monolingual children in Spanish and English in Peña *et al.* (2015) and entered these together. The purpose was to test

the discriminant accuracy of using the empirically derived cut points in both languages simultaneously. We wanted to know if classification accuracy improved with an approach in which balanced bilingual children had to score below cut points in both languages to be considered as having LI.

Likelihood ratios from the sensitivity and specificity results were calculated for both sets of analyses. Positive likelihood ratios of 10 or greater and negative likelihood ratios less than .1 are considered large, conclusive and 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 (Hanley and McNeil 1982). When positive likelihood ratios fall below 2 or when negative ratios are over .5, the test results are uninformative.

Results

Total (Spanish and English) semantics score

Table 3 displays the means and SDs of the TD and LI children. We conducted a preliminary analysis of variance (ANOVA) comparing children with and without LI using the total scores. Results indicate a significant main effect for Ability, $F(1,76) = 69.049$, $p < .001$, $\eta_p^2 = .610$.

Discriminant analysis was conducted entering the sum of Spanish and English standard scores, followed by a leave-one-out cross-validation phase, a model validation technique that assesses how results will generalize to an independent dataset. Box's M indicated that the assumption of equality of covariance matrices was met. The results of the exploratory analyses yielded a significant canonical correlation of .690, $p < .001$. This result indicates a large and significant association between the total BESA-Semantics scores and language ability. The test cut score classified 87.2% of the cases accurately with 93.3% sensitivity and 85.7% specificity. The leave one-out cross-validated classification yielded the same results. In the leave-one-out phase the model is repeatedly refit leaving out a single observation, and the model is used to derive a prediction for the left out observation. The classification results yielded a positive likelihood ratio of 6.53 (CI = 3.52–12.00), which is informative, and a negative likelihood ratio of 0.08 (CI = .01–.52), which is highly informative.

English and Spanish two-dimensional bilingual coordinate score approach

We conducted preliminary ANOVAs comparing children with and without language ability in Spanish and English (see table 1 for means by group). There was a significant main effect for Ability, in Spanish $F(1,76) = 46.595$, $p < .001$, $\eta_p^2 = .380$; and English $F(1,76) = 32.910$, $p < .001$, $\eta_p^2 = .302$.

We used the derived the cut scores of 87.81 for English 85.02 for Spanish. These cut scores were empirically derived on the basis of 179 monolingual Spanish (36 with LI, 143 with typical development) and 183 monolingual English (49 with LI, 134 with typical development) comparison children from Peña *et al.* (2015). If children scored below the cut point on both measures, they were classified as LI. Children were classified as TD if either

of the two scores was above the cut point. Exploratory discriminant analysis was followed by a leave-one-out cross-validation phase.

Box's M indicated that the assumption of equality of covariance matrices was met. The results of the exploratory analyses yielded significant differences between children with and without LI, $F(1,76) = 260.801$, $p < .001$, and a significant canonical correlation of .880, $p < .001$. This result indicates a large and significant association between the two BESA-Semantics scores and language ability. The test cut score classified 96.2% of the cases accurately with 93.3% sensitivity and 96.8% specificity. The leave one-out cross-validated classification yielded the same results. A positive likelihood ratio of 29 (CI = 7.47–116.00) and a negative likelihood ratio of 0.07 (CI = .01–.46) are highly informative.

The results of the two-language classification are displayed in figure 2. The standard scores for the 78 children are displayed. Note that most of the LI children fall in the lower left-hand quadrant, and most of the TD children fall in the upper right-hand quadrant. The upper left-hand quadrant shows children who would be incorrectly identified with LI if they were tested in Spanish only. That is, their Spanish scores fell below the cut point. Similarly, the cases in the lower right-hand quadrant are those who would be misidentified as having LI if they were tested in English only—these children scored below the cut point in English.

Discussion

The current study evaluated two ways of combining semantics test scores in bilinguals who were tested in each language. We examined the diagnostic accuracy of using a total semantics score as well as a two-dimensional bilingual coordinate score. While several researchers have advocated testing in both languages, until now there has not been an empirical evaluation of such approaches. Testing the accuracy of these types of combined two-language approaches is important to establish whether appropriate classification accuracy can be obtained. Such data-based approaches should lead to improved diagnostic outcomes.

In the current study, both the total semantics score and the two-dimensional bilingual coordinate score resulted in acceptable accuracy levels, but the two-dimensional bilingual coordinate score had a higher classification rate. As compared with single-language approaches (i.e., Peña *et al.* 2015) both the two-language approaches yielded more accurate classification compared with testing in English or Spanish alone. We also found higher positive and lower negative likelihood ratios when the children were classified on the basis of both languages. Thus, one can have a great deal of confidence when making a diagnostic decision on the basis of two languages using a test of semantic depth such as that included on the BESA.

Semantic language testing of bilinguals is more challenging than for monolinguals because bilinguals vary greatly in the amount of L1 and L2 exposure, its timing and content. The question of language dominance is additionally complicated because bilinguals do not demonstrate similar dominance in all domains (Bedore *et al.* 2010, 2012). The focus on a dominant language ignores skills that a bilingual may possess solely in the non-dominant

language. A benefit to a two-language approach is that the entire language system is accounted for, and there is no risk of ignoring skills in the non-dominant language that could be informative. This is especially true for children who are in the process of shifting dominance from one language to another. A two-dimensional bilingual coordinate approach alleviates the need to rely on measures of dominance, accounts for skills and knowledge in both languages, and increases diagnostic accuracy.

From a practical perspective, classification accuracy is critical during the early school years. This is when diagnostic decisions about LI are frequently made. Children may be changing in their relative exposure to the home versus school language, resulting in instability of language performance. By classifying children as LI only when they scored below the cut-off in both their languages, we improved classification accuracy and reduced the percentage of false-positives for a test of semantic development. These findings provide empirical validation for approaches used by researchers to classify children with LI or to rule out LI (Paradis *et al.* 2003, Kohnert *et al.* 2006).

Between the two approaches presented here, both the total score approach and the two-dimensional bilingual coordinate approach had acceptable classification accuracy. However, only the two-dimensional bilingual coordinate approach had classification accuracy above 90%. The total score approach was an advantage over a single-language approach and had a small amount of misclassification (14.3% false-positives). The two-dimensional bilingual coordinate approach retained the high true positive rate of 93.3% and reduced the false-positive rate to 3.2%. Why is the two-dimensional bilingual coordinate approach more accurate? Examination of the false-positive cases showed similar patterns among the children's profiles of performance. The subgroup of children scored within the normal range in one language (standard score $M = 95.85$, range = 84–112) and more than 2 SDs below the mean in the other language (standard score $M = 56.48$, range = 43–68). Their very low score in one of their two language disadvantaged them relative to the rest of the group whose lower score was typically no more than 1.5 SDs below the mean. A two-dimensional bilingual coordinate approach helped to offset a very low score with a score above the cut point in the stronger language. This approach is efficient because it does not require separate norms for monolingual and bilingual children with different levels of exposure to each language. Further, this approach additionally has the advantage of working even in absence of detailed knowledge of a children's level of exposure to each language.

With respect to the total score, one might ask whether a bilingual norm in which the raw scores are added together and then standardized might have improved the classification. The answer to this question requires empirical testing. Here, however, we speculate that if, like the typical balanced bilingual children in this study, there were a large number with uneven scores across the two languages in a norm, the SDs would be large relative to that of a monolingual norm. A large SD in a normative sample would likely have the effect of masking the differences between children with and without LI. Mathematically this would be similar to the effect of having norms that include children with and without LI where the variability of the performance of children with LI lowers the mean and increases the SD thus reducing overall classification accuracy (Peña *et al.* 2006b). A related question is whether including bilingual children's Spanish and English scores in the norms improves

classification accuracy. Recall that for bilingual children tested in both languages the corresponding norm only includes the higher of the Spanish or English score. Including lower scores in the norms for each language would also have had the effect of increasing the SDs in each language leading to reducing potential differences between children with and without LI. Note, however, that the sample of 78 tested here included a relatively small sample of 15 children with LI. Replication of this study with a larger sample will need to further confirm the findings presented here. Furthermore while we propose that this approach should work well with children who are less balanced in their current use of two languages, such an application needs to be empirically tested.

Clinical implications

We found that a bilingual approach of testing in both languages was more accurate than testing only in one language. However, such an approach may be challenging given the lack of bilingual speech–language pathologists available to test in two languages. The two-dimensional bilingual coordinate approach suggests that if single-language testing (e.g., English) demonstrates performance within normal limits, the other language would not need to be formally tested. That is, English language testing could be used to rule out LI consistent with the notion that LI would affect both languages (Kohnert 2010, Bedore and Peña 2008, Bedore *et al.* 2010). From a clinical perspective, however, it is critical to be able to respond to the concerns that led to the evaluation, and thus conducting language samples to speak to areas of concern in both languages is good practice. If however, test results in one language indicate performance in the impaired range, it would be important to conduct formal testing in the other language to make a more accurate diagnosis.

References

- Allen, M., Yen, W. Introduction to Measurement Theory. Belmont, CA: Wadsworth; 1979.
- American Speech–Language–Hearing Association. Clinical Management of Communicatively Handicapped Minority Language Populations [Position Statement]. Rockville: American Speech–Language–Hearing Association; 1985.
- Arnold, BR., Matus, YE. Test translation and cultural equivalence methodologies for use with diverse populations. In: Cuellar, I., Paniagua, FA., editors. Handbook of Multicultural Mental Health. San Diego, CA: Academic Press; 2000.
- Bedore LM, Leonard LB. Verb inflections and noun phrase morphology in the spontaneous speech of Spanish-speaking children with specific language impairment. *Applied Psycholinguistics*. 2005; 26:195–225.
- Bedore LM, Peña ED. Assessment of bilingual children for identification of language impairment: current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*. 2008; 11:1–29.
- Bedore LM, Peña ED, García M, Cortez C. Conceptual versus monolingual scoring: when does it make a difference? *Speech, Language, Hearing Services in Schools*. 2005; 36:188–200.
- Bedore LM, Peña ED, Gillam RB, Ho TH. Language sample measures and language ability in Spanish–English bilingual kindergarteners. *Journal of Communication Disorders*. 2010; 43:498–510. [PubMed: 20955835]
- Bedore LM, Peña ED, Joyner D, Macken C. Parent and teacher rating of language proficiency and concern. *International Journal of Bilingual Education and Bilingualism*. 2011; 14:489–511.
- Bedore LM, Peña ED, Summers C, Boerger K, Resendiz M, Greene K, Bohman T, Gillam RB. The measure matters: language dominance profiles across measures in Spanish/English bilingual children. *Bilingualism: Language and Cognition*. 2012; 15:616–629.

- Betz SK, Eickhoff JR, Sullivan SF, Nippold M, Schneider P. Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language, Speech and Hearing Services in Schools*. 2013; 44:133–146.
- Bishop DVM. The underlying nature of specific language impairment. *Journal of Child Psychology and Psychiatry*. 1992; 33:3–66. [PubMed: 1737831]
- Bortolini U, Arfé B, Caselli MC, Degasperi L, Deevy P, Leonard LB. Clinical markers for specific language impairment in Italian: the contribution of clitics and non-word repetition. *International Journal of Language and Communication Disorders*. 2006; 41:695–712. [PubMed: 17079223]
- Caesar LG, Kohler PD. The state of school-based bilingual assessment: actual practice versus recommended guidelines. *Language, Speech, and Hearing Services in Schools*. 2007; 38:190–200.
- Core C, Hoff E, Rumiche R, Senor M. Total and conceptual vocabulary in Spanish–English bilinguals from 22 to 30 months: implications for assessment. *Journal of Speech, Language and Hearing Research*. 2013; 56:1637.
- DeVellis, R. *Scale Development: Theory and Applications*. Newbury Park, CA: SAGE; 1991.
- Dunn, LM., Dunn, LM. *Peabody Picture Vocabulary Test—Third Edition*. Circle Pines, MN: American Guidance Service; 1997.
- Dunn, LM., Thériault-Whalen, C., Dunn, C. *échelle de vocabulaire en images Peabody: Adaptation française du Peabody Picture Vocabulary Test—Revised*. Toronto, ON: Psy-Can; 1993.
- Fenson, L., Dale, PS., Reznick, JS., Thal, DJ., Bates, E., Har-tung, JP., Pethick, S., Reilly, JS. *MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. Baltimore, MD: Paul H. Brookes; 1993.
- Frank, I., Poulin-DUBOIS, D., Trudeau, N. *Inventaire MacArthur du développement de la communication: Mots et énoncés*. Montreal, QC: Concordia University; 1997.
- Girolametto L, Wiigs M, Smyth R, Weitzman E, Pearce PS. Children with a history of expressive vocabulary delay: outcomes at 5 years of age. *American Journal of Speech–Language Pathology*. 2001; 10:358–369.
- Gray S. Word learning by preschoolers with specific language impairment: predictors and poor learners. *Journal of Speech, Language and Hearing Research*. 2004; 47:1117–1132.
- Gutiérrez-Clellen VF, Kreiter J. Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*. 2003; 24:267–288.
- Gutiérrez-Clellen VF, Restrepo MA, Simon-Cerejido G. Evaluating the discriminant accuracy of a grammatical measure with Spanish-speaking children. *Journal of Speech, Language and Hearing Research*. 2006; 49:1209–1223.
- Gutiérrez-Clellen VF, Simon-Cerejido G. The discriminant accuracy of a grammatical measure with Latino English-speaking children. *Journal of Speech, Language and Hearing Research*. 2007; 50:968–981.
- Gutiérrez-Clellen VF, Simon-Cerejido G, Wagner C. Bilingual children with language impairment: a comparison with monolinguals and second language learners. *Applied Psycholinguistics*. 2008; 29:3–19. [PubMed: 22685359]
- Hanley JA, Mcneil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; 143:29–36. [PubMed: 7063747]
- Jacobson PF, Schwartz RG. Morphology in incipient bilingual Spanish-speaking preschool children with specific language impairment. *Applied Psycholinguistics*. 2002; 23:23–41.
- Jordaan H, Shaw-Ridley G, Serfontein J, Orelwitz K, Monaghan N. Cognitive and linguistic profiles of specific language impairment and semantic–pragmatic disorder in bilinguals. *Folia Phoniatrica et Logopaedica*. 2001; 53:153–165. [PubMed: 11316942]
- Junker DA, Stockman IJ. Expressive vocabulary of German–English bilingual toddlers. *American Journal of Speech–Language Pathology*. 2002; 11:381–394.
- Kohnert KJ. Bilingual children with primary language impairment: issues, evidence and implications for clinical actions. *Journal of Communication Disorders*. 2010; 43:456–473. [PubMed: 20371080]
- Kohnert KJ, Windsor J, Yim D. Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research and Practice*. 2006; 21:19–29.

- Leonard, LB., Levy, Y., Schaeffer, J. Language Competence across Populations: Toward a Definition of Specific Language Impairment. Mahwah, NJ: Lawrence Erlbaum Associates; 2003. Specific language impairment: characterizing the deficits.
- Marchman VA, Fernald A, Hurtado N. How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish–English bilinguals. *Journal of Child Language*. 2010; 37:817–840. [PubMed: 19726000]
- Mcgregor, KK. Semantics in child language disorders. In: Schwartz, R., editor. *Handbook of Child Language Disorders*. New York, NY: Psychology Press; 2009.
- Miller, JF., Iglesias, A. Systematic Analysis of Language Transcripts—SALT V9. Language Analysis Laboratory, Waisman Center. Madison, WI: University of Wisconsin; 2005.
- Oller DK, Pearson BZ, Cobo-Lewis AB. Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*. 2007; 28:191–230. [PubMed: 22639477]
- Paradis J, Crago M, Genesee F, Rice ML. French–English bilingual children with SLI: how do they compare with their monolingual peers? *Journal of Speech, Language and Hearing Research*. 2003; 46:113.
- Patricacou A, Psallida E, Pring T, Dipper L. The Boston Naming Test in Greek: normative data and the effects of age and education on naming. *Aphasiology*. 2007; 21:1157–1170.
- Pearson BZ. Assessing lexical development in bilingual babies and toddlers. *International Journal of Bilingualism*. 1998; 2:347–372.
- Pearson BZ, FernÁndez SC. Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning*. 1994; 44:617–653.
- Pearson BZ, FernÁndez SC, Oller DK. Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Language Learning*. 1993; 43:93–120.
- Pearson BZ, FernÁndez SC, Oller DK. Cross-language synonyms in the lexicons of bilingual infants: one language or two? *Journal of Child Language*. 1995; 22:345–368. [PubMed: 8550727]
- Peña ED. Lost in translation: methodological considerations in cross-cultural research. *Child Development*. 2007; 78:1255–1264. [PubMed: 17650137]
- Peña ED, Bedore LM, Kester ES. Discriminant accuracy of a semantics measure with Latino English-speaking, Spanish speaking, and English–Spanish bilingual children. *J Commun Disord*. 2015; 53:30–41. [PubMed: 25573288]
- Peña ED, Bedore LM, Rappazzo C. Comparison of Spanish, English, and bilingual children’s performance across semantic tasks. *Language, Speech and Hearing Services in the Schools*. 2003; 34:5–16.
- Peña ED, Gillam RB, Malek M, Ruiz-FELTER R, Resendiz M, Fiestas C, Sabel T. Dynamic assessment of school-age children’s narrative ability: an experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research*. 2006a; 49:1037–1057.
- Peña, ED., Gutierrez-Clellen, VF., Iglesias, A., Goldstein, B., Bedore, LM. *Bilingual English Spanish Assessment*. San Rafael, CA: A-R Clinical Publ; 2014.
- Peña ED, Spaulding TJ, Plante E. The composition of normative groups and diagnostic decision making: shooting ourselves in the foot. *American Journal of Speech, Language Pathology*. 2006b; 15:247–254. [PubMed: 16896174]
- Restrepo MA. Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language and Hearing Research*. 1998; 41:1398–1411.
- Salvia, J., Ysseldyke, J., Bolt, S. *Assessment: In Special and Inclusive Education*. Cengage Learning; 2009.
- Sheng L, Bedore LM, Peña ED, Taliancich-Klinger C. Semantic convergence in Spanish–English bilingual children with primary language impairment. *Journal of Speech, Language and Hearing Research*. 2013; 56:766–777.
- Sheng L, Mcgregor KK. Lexical–semantic organization in children with specific language impairment. *Journal of Speech, Language, and hearing Research*. 2010; 53:146–159.
- Sheng L, Peña ED, Bedore LM, Fiestas CE. Semantic deficits in Spanish–English bilingual children with language impairment. *Journal of Speech, Language and Hearing Research*. 2012; 55:1–15.

- Stansfield CW. Test translation and adaptation in public education in the USA. *Language Testing*. 2003; 20:189–207.
- Thordardottir ET, Rothenberg A, Rivard ME, Naves R. Bilingual assessment: can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*. 2006; 4:1–21.
- Trudeau N, Frank I, Poulin-Dubois D. Une adaptation en français québécois du MacArthur Communicative Development Inventory. *Journal of Speech–Language Pathology and Audiology*. 1999; 23:31–73.
- Uccelli P, PÁez MM. Narrative and vocabulary development of bilingual children from kindergarten to first grade: developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*. 2007; 38:225–236.

What this paper adds**What is already known on this subject?**

Bilingual children have lexical–semantic systems that are distributed across their two languages because their everyday experiences occur in different languages and they hear different words across these contexts. Thus, they often learn words associated with each language that they do not express in the other. This distributed lexical–semantic system makes it difficult to assess bilinguals accurately for the purpose of making diagnostic decisions. Approaches that combine semantics scores across languages seem to improve scores for bilingual children with typical development so that they are more comparable with typical monolingual performance. But it is unknown whether or not this will lead to improved classification of bilingual children with and without language impairment.

What this study adds

Two approaches, total semantics (represented as a sum of standard scores in Spanish and English) and two-dimensional bilingual coordinate score (ordered pairs represented in an x–y graph) yielded diagnostic accuracy above 85%.

Figure 1.

Two language bilingual ability models: 1, 2 and 3 depict typically developing cases with different levels of dominance in languages A and B; 4 depicts a language impaired case.

Figure 2.
 x - y scatter plot: balanced bilingual children.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1
Table of elements: English and Spanish versions of the semantics subtests of the BESA

Item difficulty		Item type						Total
		SD	LC	CP	CT	AN	FN	
English	Easy	2	4	2	3	1	4	
	Med	5	–	2	6	1	2	
	Difficult	1	2	4	4	3	2	
	Total	8	6	8	13	5	8	48
Spanish	Easy	1	–	4	3	–	9	
	Med	1	5	2	4	1	2	
	Difficult	2	4	–	6	4	–	
	Total	5	9	6	13	5	11	49

Note: SD, similarities and differences; LC, linguistic concepts; CP, characteristic properties; CT, category generation; AN, analogies; FN, functions.

Table 2

Indicators of language impairment and language exposure

Measure	TD	LI	Significance (<i>p</i>)
Parent Concern Rating	.19 (.40)	.46 (.52)	.042
Teacher Concern Rating	.03 (.17)	.50 (.55)	< .001
Clinician Concern Rating	.09 (.30)	.33 (.49)	.031
Percent Ungrammatical English	23.67% (17.27%)	31.67% (17.41%)	.213
Percent Ungrammatical Spanish	13.63% (6.84%)	16.67% (4.65%)	.336
Percent Input and Output English	49.31% (4.59%)	49.93% (5.07%)	.586

Note: Concern ratings: 1 = concern, 0 = no concern.

Table 3

Means and standard deviations for the TD and LI groups

	TD		LI	
	Mean	SD	Mean	SD
Total score	192.58	(24.92)	134.76	(20.85)
English	92.31	(15.85)	66.56	(14.56)
Spanish	100.27	(17.42)	68.20	(10.36)

Note: Total score is the sum of standard scores; English and Spanish scores are standard scores.