UC Irvine
UC Irvine Previously Published Works

Title

Conceptual Scoring and Classification Accuracy of Vocabulary Testing in Bilingual Children

Permalink

Journal

ISSN

Authors

Anaya, Jissel B
Peña, Elizabeth D
Bedore, Lisa M

Publication Date

DOI

Peer reviewed

Conceptual Scoring and Classification Accuracy of Vocabulary Testing in Bilingual Children

Jissel B. Anaya,[a] Elizabeth D. Peña,[a] and Lisa M. Bedore[a]

Purpose: This study examined the effects of single-language and conceptual scoring on the vocabulary performance of bilingual children with and without specific language impairment. We assessed classification accuracy across 3 scoring methods.
Method: Participants included Spanish–English bilingual children (*N* = 247) aged 5;1 (years;months) to 11;1 with and without specific language impairment. Children completed the English and bilingual versions of the Expressive One-Word Picture Vocabulary Test–Third Edition (Brownell, 2000a, 2001). Six scores, 2 representing monolingual scores in English and Spanish and 4 conceptual scores, were derived. The conceptual scores included within-test conceptual scores, which credited language responses in the other language during test administration, and across-test

conceptual scores, which we compiled by examining responses across independent administrations of the test in each language.
Results: Across-test conceptual scoring resulted in the highest scores and better overall classification, sensitivity, and specificity than within-test conceptual scoring. Both were superior to monolingual scoring; however, none of the methods achieved minimum standards of 80% accuracy in sensitivity and specificity.
Conclusions: Results suggest that bilingual children are not always able to readily access their other language in confrontation naming tasks. Priming or inhibition may play a role in test performance. Across-test conceptual scoring yielded the highest classification accuracy but did not meet minimum standards.

C hildren with specific language impairment (SLI) have limited expressive vocabulary (Gray, Plante, Vance, & Henrichsen, 1999; Leonard, 2014; Rice, Buhr, & Nemeth, 1990) and have difficulty learning new words (Alt & Spaulding, 2011; Gray, 2004). They often demonstrate weaknesses in word retrieval and have poor expressive vocabulary skills (Kambanaros, Michaelides, & Grohmann, 2015). Because of these difficulties in the domain of word learning and word knowledge in children with SLI, single-word vocabulary tests are often used to qualify children for speech and language services (Betz, Eickhoff, & Sullivan, 2013). Documented dependence on single-word vocabulary tests is surprising in the face of their poor psychometric properties (Gray et al., 1999). Bilingual children may perform poorly on standardized vocabulary

tests when tested in one language and compared with their monolingual peers (Bialystok, Luk, Peets, & Yang, 2010; Hemsley, Holm, & Dodd, 2010; Umbel, Pearson, Fernández, & Oller, 1992). This low performance may be because their vocabulary knowledge is distributed across their two languages (Gollan & Ferreira, 2009; Ordóñez, Carlo, Snow, & McLaughlin, 2002; Patterson & Pearson, 2004; Sheng, Peña, Bedore, & Fiestas, 2012). Another factor affecting vocabulary naming may be the accessibility of the first and second language in the context of the test language (Meuter & Allport, 1999). The manner in which vocabulary words are elicited may affect children's performance. Conceptual scoring, where responses are accepted in either language and credited on a single-word test, has been demonstrated to mitigate some of the documented differences in vocabulary scores between monolingual and bilingual children (Bedore, Peña, Garcia, & Cortez, 2005; Pearson, Fernandez, & Oller, 1993; Peña, Bedore, & Kester, 2015). The numerous studies that demonstrate the utility of conceptual vocabulary scoring for capturing vocabulary knowledge may lead clinicians to think that such an approach is sufficient for informing diagnostic decision making. Because vocabulary tests are documented to have poor sensitivity and

---

[a]Department of Communication Sciences and Disorders, The University of Texas at Austin

Elizabeth D. Peña is now at the University of California, Irvine.

Correspondence to Jissel B. Anaya: jissel.anaya@gmail.com

---

specificity, however, we have no reason to expect high classification accuracy, even if conceptual scoring is used. Thus, we explore the diagnostic accuracy of conceptual vocabulary scoring to determine if using conceptual scoring with single-word vocabulary tests increased their sensitivity so they might contribute to the identification of SLI in bilingual children.

Vocabulary Difficulties in Children With SLI

Lexical acquisition deficits are a commonly noted characteristic of SLI in research and clinical practice (Gray, 2004, 2005, 2006; Horohov & Oetting, 2004; Kiernan & Gray, 1998; Nash & Donaldson, 2005; Rice, Oetting, Marquis, Bode, & Pae, 1994; Weismer & Hesketh, 1998). Compared with peers with typically developing (TD) language skills, children with SLI exhibit slower vocabulary growth (Rescorla, Roberts, & Dahlsgaard, 1997), difficulty learning new words (Alt, Plante, & Creusere, 2004; Alt & Suddarth, 2012), and limited expressive vocabulary (Gray & Brinkley, 2011). Experimental studies comparing children with SLI and TD language skills document significant difficulties in word learning. For example, Weismer and Hesketh (1998) found that children with SLI require more exposures to a word to comprehend or produce it than their TD peers. In both fast mapping and quick incidental learning tasks, children with SLI learn fewer novel words (Alt, 2011; Alt et al., 2004; Gray, 2004, 2006; Oetting, Rice, & Swank, 1995; Rice, Cleave, & Oetting, 2000; Rice et al., 1994). They also demonstrate weaknesses in word retrieval (Gray, 2004; Gray & Brinkley, 2011; Kambanaros et al., 2015; McGregor, Newman, Reilly, & Capone, 2002), naming errors, and word approximation difficulties (Dollaghan, 1998; Sheng & McGregor, 2010; Spaulding, 2010).

Due to these documented deficits in word learning and perhaps to their ease of administration and scoring, vocabulary tests are widely used by clinicians to determine whether a child's language skills require further evaluation (Campbell, Bell, & Keith, 2001), as a method of identifying SLI in children for research studies (Rice et al., 1990; Rice, Buhr, & Oetting, 1992; Rice et al., 1994), or to document vocabulary growth (Rowe, Raudenbush, & Goldin-Meadow, 2012). While numerous comprehensive language batteries are commercially available to clinicians, vocabulary tests are frequently used as a component of diagnostic evaluations of children to determine if a child meets the criteria for SLI (Betz et al., 2013), even though test manuals may not recommend them for this purpose (Brownell, 2000a).

Vocabulary Tests as Indicators of SLI

To use vocabulary test scores for screening or identification purposes, clinicians must first be confident that the assessment has evidence of validity and reliability (McCauley & Swisher, 1984). Plante and Vance (1994) suggested that language tests that discriminate between children with and without SLI with an accuracy level above 90% are considered "good" discriminators, and those that

discriminate with an accuracy level between 80% and 90% are considered "fair" discriminators. Gray et al. (1999) evaluated the clinical utility of four single-word vocabulary tests (two expressive and two receptive). Their results indicated that sensitivity ranged from 71% to 77% and specificity ranged from 68% to 77% across these tests. When these authors combined scores across tests, sensitivity and specificity failed to improve; thus, their results did not support the use of vocabulary tests for SLI identification purposes in monolingual children (Gray et al., 1999). As such, we do not expect different results with a bilingual population.

A recent clinical survey administered to speech-language pathologists (SLPs) within the United States found that four single-word vocabulary tests were among the top 10 most frequently selected assessments for diagnostic purposes (Betz et al., 2013). These included the Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007), the Expressive One-Word Picture Vocabulary Test–Third Edition (EOWPVT-3; Brownell, 2000a), the Receptive One-Word Picture Vocabulary Test (Brownell, 2000b), and the Expressive Vocabulary Test–Second Edition (Williams, 1997). Regrettably, the quality of the standardized tests, as measured by their poor classification accuracy, does not coincide with their high frequency of use.

Testing Vocabulary of Bilingual Children

Using vocabulary tests with bilingual populations is particularly problematic given that bilingual children have consistently been found to score below their monolingual peers on vocabulary measures in one (Bialystok et al., 2010; Hemsley et al., 2010; Umbel et al., 1992) or both of their languages (Gollan & Ferreira, 2009; Ordóñez et al., 2002; Patterson & Pearson, 2004; Sheng et al., 2012). There are some concepts for which bilinguals know the corresponding word in one language but do not know the label in their other language. This pattern of distributed vocabulary has been observed in infants and toddlers, preschool children, and school-age children and can result in an underestimation of a child's lexical knowledge when considering only one of their languages.

In younger children, Pearson, Fernández, and Oller (1995) compared parent reports on the MacArthur Communicative Development Inventories (Fenson et al., 1994) and its Spanish adaptation, the MacArthur Inventario de Desarrollo de Habilidades Comunicativas (Jackson-Maldonado & Bates, 1988), and found that, among Spanish–English bilingual children between the ages of 0;8 (years; months) and 2;6, 70% of their vocabulary was language specific and did not overlap with their other language. These findings are comparable to studies of older bilingual children. Peña, Bedore, and Zlatic-Giunta (2002) found that a large proportion (over 65%) of the category items generated by bilingual (Spanish–English) children between the ages of 5;1 and 6;5 was unique to either language. This distribution is due, in part, to the context in which bilingual children learn and use concepts. For example, a child whose

dominant home language is Spanish but who attends an English instruction classroom may know math terms only in English but may know more food and household items in Spanish. Children may also choose to use different words depending on the setting, interlocutor, and context (Iglesias, 2001) as well as their cultural experiences (Peña, 2001). The total knowledge of bilinguals in a single language may not be directly comparable to that of monolinguals due to the nonoverlapping vocabulary knowledge. These findings underscore the importance of conducting assessments in both languages to document bilinguals' word knowledge.

One alternative to single-language scoring that has been used with young children is the use of total vocabulary scores (Core, Hoff, Rumiche, & Señor, 2013). These are derived by summing up raw vocabulary scores in each of a bilingual child's languages and interpreting the total raw score in reference to monolingual norms (Pearson et al., 1993). While some researchers have found that using total vocabulary scores with bilingual children under the age of 3;0 results in similar vocabulary sizes to monolingual children (Core et al., 2013; Marchman & Martinez-Sussmann, 2002; Pearson et al., 1993), other researchers have suggested it may result in an overestimation of the bilingual child's lexical knowledge (Junker & Stockman, 2002). The children in the present study are over the age of 3;0. Therefore, total vocabulary scores were not considered.

Conceptual scoring, which considers the number of concepts for which a child has a word in any language, has been demonstrated to reduce some of the documented differences between monolingual and bilingual children (Bedore et al., 2005; Core et al., 2013; Pearson et al., 1993; Peña et al., 2015). To achieve their communication goals and to convey concepts, bilinguals will make use of their vocabulary resources in both languages by sometimes switching between their two languages at the single-word level (i.e., code-mix) or at the conversational or narrative level (i.e., code-switch; Greene, Peña, & Bedore, 2013). Conceptual scores can be gathered from analyzing a child's responses during spontaneous and elicited code-mixing.

Counting Conceptual Vocabulary

Although conceptual scoring can give us a more complete picture of a bilingual child's lexical knowledge than single-language scoring, conceptual scores can be obtained in different ways. One method is to ask the child to code-mix immediately after an incorrect or no response is given in the target language. For example, if the target word is *dog* and the bilingual child does not respond or produces an incorrect response (e.g., *cat*), the administrator would then ask the child to name the picture in Spanish. A conceptual score can be derived after a single test administration in either language. Another method of obtaining a conceptual score is to administer all items in the dominant language first and then readminister all incorrect items in the second language. The Preschool Language Scale–Fifth Edition: Spanish (Zimmerman, Steiner, & Pond, 2011) is a language

assessment that uses this conceptual scoring method. A third method of obtaining a conceptual score is to test each of the child's languages independently of each other (e.g., English administration on one day, Spanish administration on a different day) and then add the total unique concepts that the child knows across their two languages. While this method has not been recommended in standardized tests, perhaps because it is more time consuming, it has been implemented in research studies. For example, both Pearson (2001) and Marchman and Martinez-Sussmann (2002) measured bilingual children's "Total Conceptual Vocabulary" on a parent report measure by counting cross-language synonyms, such as *dog* and *perro*, only once. Although these three different conceptual scoring methods have been used in various standardized tests and research studies, possible differences in children's conceptual vocabulary scores have not been evaluated.

These different methods of elicitation require different levels of activation of the first and second language(s) during testing, potentially leading to switch costs that result in variations in vocabulary scores across testing methods. Results from some studies suggest that switching into the weaker language presents more challenges than switching into the stronger language because dominant language suppression requires more cognitive resources (Bialystok, Craik, & Luk, 2008; Kroll, Bobb, Misra, & Guo, 2008; Misra, Guo, Bobb, & Kroll, 2012). For example, Greene et al. (2013) found that only about half of bilingual preschool-age children code-mixed spontaneously when responding to expressive semantics items on an English–Spanish language screening measure. Of those who code-mixed, only a small minority (7%) code-mixed in both directions. Among the children who code-mixed in one direction, children were more likely to code-mix into their dominant language when tested in their nondominant language. This is consistent with work finding that when a bilingual is speaking their dominant language, less effort is required to suppress the less dominant language (Costa & Santesteban, 2004).

Switching costs are context dependent. The language suppression hypothesis posits that the linguistic context of the immediate environment influences switching (Green, 1998). For example, exposure to an English-dominant environment may lead to higher switch costs for the other language, even if the speaker is dominant in the other language. In immersion (L2) contexts, it may be more difficult to switch into the more dominant (L1) language because resources are allocated to inhibiting L1 to allow for second language (L2) learning. In this context, when L2 learners are then asked to uninhibit L1, L1 may be more difficult to access, as compared with those learning an L2 in a non-immersion context (Linck, Kroll, & Sunderman, 2009). For English language learners in the U.S. context, immersion is often the method by which they learn English. Thus, when being tested in English, they may not readily switch into the native language. Yan and Nicoladis (2009) suggest that low expressive vocabulary scores in bilinguals may not reflect small expressive vocabularies but rather a difficulty

accessing and retrieving the appropriate word in the moment of the assessment. These findings suggest that lexical representations in the nontarget language are suppressed, thereby supporting separate administration of expressive vocabulary tests in each language rather than requiring children to respond in both languages during a single administration. It is hypothesized that this method of obtaining a conceptual score would yield higher scores, which better represent their lexical knowledge. Given that vocabulary tests applied to monolinguals are known to have poor sensitivity and specificity, however, we expect that similar patterns will emerge even if conceptual scores are used with bilinguals.

The purposes of this study were to compare performance of bilingual (Spanish–English) school-age children with and without SLI on the EOWPVT-3 (Brownell, 2000a) and the Expressive One-Word Picture Vocabulary Test–Third Spanish-Bilingual Edition (EOWPVT-3: SBE; Brownell, 2001) and to determine which method of scoring yielded the best diagnostic accuracy. Specific questions included the following:

1. Are there differences in EOWPVT scores by language ability, the language of testing, and scoring method?

2. Which method of scoring—(a) single-language, (b) conceptual scoring within a single administration, or (c) conceptual scoring across independent administrations of the EOWPVT-3 and the EOWPVT-3: SBE—results in higher diagnostic accuracy of SLI in bilingual children?

Method

Participants

We obtained data from an ongoing longitudinal study examining the cross-linguistic outcomes of bilingual children with and without SLI (Bedore, Peña, Griffin, & Hixon, 2016). The data included here are from the second phase of the study. There were 290 bilingual children (at least 20% combined input and output in both languages) in the data set. Of these, 33 had missing EOWPVT data (19 were missing data in both languages, 14 were missing data in Spanish, and 10 were missing data in English). Thus, the participants in the present study included 247 Spanish–English bilingual children enrolled in kindergarten, second grade, and fourth grade aged 5;1 through 11;1. Participant demographics are shown in Table 1. The average combined language input and output was 51.19% (*SD* = 26.37%) in English and 48.81% (*SD* = 26.37%) in Spanish. All children attended bilingual campuses implementing dual-language programs (Spanish/English). Socioeconomic status was based on the mother's education. The average was a 2.61 (*SD* = 1.63) Hollingshead score, which corresponds to a junior high school (2) or partial high school (3) education (Hollingshead, 1975). We identified 38 children with SLI (15%) and 209 as TD. We oversampled our SLI group in order to reach adequate statistical power

for our analyses, resulting in an SLI sample that exceeded the 7% prevalence reported by Tomblin (1997).

Sampling Procedure

Each year for 4 years, preschool, first grade, and third grade children were screened using the Bilingual English–Spanish Oral Screener (BESOS; Peña, Bedore, Gutiérrez-Clellan, Iglesias, & Goldstein, 2010b), an experimental version of a semantics and morphosyntax screener with high classification accuracy (Lugo-Neris, Peña, Bedore, & Gillam, 2015). Parents also responded to a questionnaire and reported on their child's hour-by-hour exposure to and use of Spanish and English (Gutiérrez-Clellen & Kreiter, 2003; Gutiérrez-Clellen, Restrepo, & Simón-Cereijido, 2006). Because the purpose of the larger study was to follow bilingual children at risk for SLI, participants were invited to the longitudinal portion of the study if their highest score on the morphosyntax or semantics subtest of the BESOS fell below the 25th percentile in either language. We required children to have at least 20% combined exposure to and use of both English and Spanish and to have been exposed to English by the age of 5 years. Moreover, TD age-, gender-, socioeconomic status–, and language use–matched peers who scored above the 25th percentile were invited at the rate of two matches for every child at risk for SLI.

Testing Procedure

In the confirmatory phase of the study, we administered a comprehensive battery of standardized and experimental measures in English and Spanish to confirm the presence or absence of SLI. All examiners were Spanish–English bilinguals and were trained by a certified bilingual SLP on all confirmatory measures. Training consisted of reading test manuals, observing several sessions of testing from an experienced examiner, and administering the measures at least two times under the supervision of an experienced examiner.

We tested participants at their schools in locations school administrators designated for testing. Examinations were conducted across three to four sessions lasting 30 to 45 min, depending on the child's attention span and time constraints resulting from testing in a school environment. We randomized the order of tests and language of testing across participants. Children completed the following tests in both English and Spanish: two subtests of the Bilingual English-Spanish Assessment (BESA; Peña, Gutierrez-Clellen, Iglesias, Goldstein, & Bedore, 2014) or the Bilingual English-Spanish Assessment–Middle Extension, Experimental Version (BESA-ME; Peña et al., 2010b), the Test of Narrative Language (TNL; Gillam & Pearson, 2004), and an experimental version of the TNL adapted to Spanish (TNL-S; Gillam, Peña, Bedore, & Pearson, in development). All children also completed the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998) and passed an initial hearing screening

Table 1. Participant demographic information.

Variable	Minimum	Maximum	*M*	*SD*
Age in months	61	133	92.37	18.29
Nonverbal IQ	76	129	101.10	13.92
Percent English input/output	0	100	51.19	26.37
Year of first exposure to English	0	6	2.76	1.77
Mother's education (Hollingshead score)	0	7	2.61	1.63

Note. *N* = 247 (127 girls, 120 boys).

or a follow-up hearing test conducted by the schools' nurses. Research assistants scored all tests following training that consisted of reading the relevant test manuals, one-on-one meetings with experienced project staff including certified SLPs, and follow-up with a project coordinator. Parents and teachers also completed the Bilingual Input–Output Survey (BIOS) and the Inventory to Assess Language Knowledge (ITALK) from the BESA (Peña et al., 2014).

Confirmatory Measures

UNIT (Bracken & McCallum, 1998)

The UNIT was used to exclude participants who met criteria for intellectual disability. The abbreviated battery of the UNIT is administered nonverbally and includes two subtests: Symbolic Memory and Cube Design. The test is appropriate for children ages 5 to 17 years and takes approximately 15 min to administer. The subtests of the abbreviated battery all have low linguistic demand and low-to-medium cultural loading. The reliability coefficient for the abbreviated battery, as reported in the manual, is .96. Standard administration and scoring procedures were followed to yield a standard score for a nonverbal intelligence quotient. Consistent with previous research on SLI (Rice, Tomblin, Hoffman, Richman, & Marquis, 2004; Tomblin, Zhang, & Tager-Flusberg, 1999), children whose scores on the UNIT were below 1.3 *SD*s from the mean were excluded from further participation. We also excluded children with reported history of brain injury, severe social–emotional problems, diagnosis of an autism spectrum disorder, or hearing loss from the study.

BIOS (Peña et al., 2014)

Language history information was obtained through a detailed parent and teacher interview using the BIOS. The interview was administered by phone to parents and in person to teachers. Parents were asked about their child's history of exposure to both languages at home and school since birth to calculate the child's age of first exposure to English. Teachers responded to questions about the child's language use at school.

ITALK (Peña et al., 2014)

Parents and teachers rated the participants' ability in the following areas: vocabulary use, speech production

(intelligibility), sentence production (utterance length), grammatical proficiency, and comprehension proficiency. Each area was independently rated in both languages on a 5-point scale, where 1 = *minimal proficiency in this area* and 5 = *high proficiency in this area*. The five scores in each language were averaged to yield two separate Spanish and English scores based on parent and teacher reports, resulting in a total of four scores. Only the higher average (between Spanish and English) from the parent and the teacher was considered, yielding a total of two scores per child. We collected data from both parents and teachers to get a more complete indicator of relative language use in both the school and home environments. This helped ameliorate differences in the ratings due to variations in the amount of time parents and teachers heard the children speaking each of their languages or variations in the language demands between the home and school. If the parent or teacher did not have knowledge of an area or of one of the child's two languages, this was marked as "unknown" and not included in that average (Gutiérrez-Clellen & Kreiter, 2003).

BESA (Peña et al., 2014) or BESA-ME (Peña, Bedore, Gutiérrez-Clellen, Iglesias, & Goldstein, 2010a)

We administered the BESA to children of ages 4;0–6;11 and the BESA-ME to children of ages 7;0–10;11. Normative data are based on 381 children for English and 295 children for Spanish between the ages of 7;0 and 9;11. We additionally have a small set of participants between the ages of 10;0 and 11;11 (134 English; 94 Spanish). Examination of their scores (divided into 6-month intervals) shows no differences among them or in comparison to those in the 9;6–9;11 age range. Thus, we used the means and standard deviations of the oldest age group (9;6–9;11) to derive standard scores for the 19 children between the ages of 10;0 and 11;1.

In general, items for the BESA-ME were selected from a larger experimental item pool based on items that most discriminated SLI; thus, the subtests are not parallel across languages. The semantics subtest measures semantic breadth and depth through item types including categories, functions, characteristic properties, similarities and differences, analogies, associations, and definitions. A combination of receptive and expressive items is included. The Spanish morphosyntax test includes items relating to articles, present progressive, direct object clitics, and subjunctives, whereas the English test includes possessive –*s,*

third-person singular, regular past tense, plural nouns, present/past auxiliary + progressive –ing, copula, negatives, and passives. Classification accuracy for the English and Spanish subtests of the BESA is 92% sensitivity and 86% specificity, respectively, for 4-year-olds, 89% and 85% for 5-year-olds, and 96% and 92% for 6-year-olds (Peña et al., 2014). Preliminary analyses of classification accuracy of the BESA-ME indicate 100% sensitivity and 86.8% specificity for second graders and 100% sensitivity and 95% specificity for fourth graders. By Plante and Vance's (1994) standards, all are above the 80% "fair" level for diagnostic accuracy.

TNL (Gillam & Pearson, 2004) and TNL-S Experimental Version (Gillam et al., in development)

The TNL and the TNL-S experimental version were administered to all participants to assess narrative discourse in English and Spanish, respectively. The order of administration was counterbalanced, and the tests were separately scored. Both versions of the TNL measure story comprehension and narration skills in children ages 5;0 to 11;11. The Narrative Comprehension (NC) subtest requires children to answer questions about three stories they just heard. In the Oral Narration (ON) subtest, children are asked to retell one story using no visual prompt and to tell two stories, one given a sequence of five pictures and another given a single picture. Hispanic children represented 12% of the norming sample for the English TNL (Gillam & Pearson, 2004). The TNL has also been validated for use with bilingual children resulting in 77% sensitivity and 87% specificity (Gillam, Peña, Bedore, Bohman, & Mendez-Perez, 2013). Validity studies for the TNL-S are currently underway. Preliminary data indicate an alpha level of .888 for the NC subtest and .931 for the ON subtest (Gillam et al., in development). Point-by-point scoring reliability for the NC and ON subtests indicated 95% and 94% reliability, respectively, in English and 97% and 96% reliability, respectively, in Spanish. Reliability was completed using the written test protocols for the NC subtest and the audio-based transcriptions for the ON subtest. Results from a previous validation study that considered the consistency and accuracy of scores across varying proficiency levels, comfort levels, and presentation methods (written transcript and audio-recorded samples) indicated that examiners of varying levels of Spanish proficiency and experience level can reliably and efficiently score an assessment in Spanish when provided with specific scoring procedures (Perme, 2014).

Classification Procedure

We determined group classification by both the screening and the confirmatory testing phases of this study. Eligibility for the SLI and TD groups was derived on the basis of five indicators: (a) participants' BESOS screener scores from Phase 1, (b) TNL in English or Spanish, (c) morphosyntax subtest of the BESA or BESA-ME, (d) semantics subtest of BESA or BESA-ME, and (e) parent and teacher ratings of participants' language performance on the ITALK.

Participants were identified with SLI if they met at least four of the following five criteria: (a) scored below 1 *SD* on age norms in both languages on the BESA or BESA-ME semantics subtest, (b) scored below 1 *SD* below age norms in both languages on the BESA or BESA-ME morphosyntax subtest, (c) scored below 1 *SD* on the BESOS screener in both languages, (d) scored below 1 *SD* on the TNL in both English and Spanish, and (e) scored below an average of 4.25 (of 5) in both languages on the ITALK. We identified children with TD language skills if they scored above −1 *SD* of the mean on two or more of these measures in at least one language.

Experimental Manipulations

EOWPVT-3 (Brownell, 2000a) and EOWPVT-3: SBE (Brownell, 2001)

The EOWPVT-3 and the EOWPVT-3: SBE are individually administered, norm-referenced tests of single-word expressive vocabulary. Each test was administered once to the participants. The tests consist of 170 items presented in developmental sequence. The same items are included in both test editions, which allows for a direct comparison of items in each language. Although 16 of these are typically not administered in the Spanish-Bilingual Edition, we modified administration procedures and administered all items in Spanish to compare performance at the item level across both languages. According to the test authors, these items were excluded on the basis of their item analyses. For the current study, the English and SBE versions of the EOWPVT-3 were administered as English-only and Spanish-only versions, respectively. If the child missed an item or did not respond in the target language, the examiner prompted the child to respond in the nontarget language. For example, if the child gave a response in English during the SBE administration, the response was recorded in English and then elicited in Spanish. Both responses were recorded verbatim. Test administration in each language was discontinued if the child did not obtain a basal. After the ceiling of six consecutive incorrect items was achieved, 14 additional items were administered to ensure a ceiling was obtained across all scoring methods. Basal and ceiling rules provided in the test manuals were followed when computing raw scores. Specifically, responses to additional items above the ceiling were not included in the calculation of raw scores. Six different standard scores were recorded using three methods.

Scoring

Monolingual English

Scoring of responses adhered to the instructions in the test manual (Brownell, 2000a). Instructions and test questions were administered in English, and only responses in English were accepted as correct. Standard scores were derived using English norms.

Monolingual Spanish

Scoring of the Spanish-Bilingual Edition (Brownell, 2001) was similar to the monolingual version given that only Spanish responses were accepted as correct. The 16 do-not-administer items were not included in this score. Standard scores were derived using norms published in the SBE manual.

Within-Test Conceptual Scores (English Norm Comparison and SBE Norm Comparison)

Scoring involved conceptually analyzing each item within each single-language administration, similar to the standardized testing procedures outlined in the EOWPVT-3: SBE manual (Brownell, 2001). Responses in either language during administration were accepted as correct. Two standard scores were derived, one using the norms published in the SBE manual (within-test conceptual Spanish) and another using the English edition manual (within-test conceptual English). The 16 do-not-administer items were included in the within-test conceptual English standard score but not in the within-test conceptual Spanish score, consistent with the standardized procedures.

Across-Test Conceptual Scores (English Norm Comparison and SBE Norm Comparison)

Scoring involved analyzing each item across both administrations. Correct responses in either language were accepted as correct. Two standard scores were derived, one using the norms published in the SBE manual and another using the English edition manual. The 16 do-not-administer items were included in the across-test conceptual English standard score but not in the across-test conceptual Spanish score, consistent with the standardized procedures.

Raw total scores for each scoring method were derived using Visual Basic for Applications macro programming language in Microsoft Excel. Ten percent of the scores in each language were computed manually and compared with the automatically coded results. The code was revised until we reached a threshold of 100% accuracy.

Reliability

To estimate interexaminer reliability for these tests, 20% of all of the measures were independently rescored by a second research assistant, and the point-by-point percent agreement for each test was as follows: 95.4% for the EOWPVT-3, 93.8% for the EOWPVT-3: SBE, 97% for the BESA, 95.8% for the BESA-ME, and 100% for the UNIT.

Results

Comparisons by Ability, Language, and Scoring Method

The first set of analyses addressed whether there were differences in obtained scores by language ability and

by the language of testing (mean scores are displayed in Table 2). We compared the six different standard scores derived from the three different scoring methods using a mixed analysis of variance. The between-subjects factor was ability (SLI and TD). The within-subject factors were language of testing (Spanish and English) and scoring method (single-language, within-test, across-test). For the standard score comparisons, Mauchly's test of sphericity yielded an alpha level less than .05, indicating that we did not meet sphericity assumptions. Therefore, we report the multivariate Wilks's λ results, which do not require sphericity because they are based on difference scores that are computed by comparing scores from each level of the within-subject factor.

There were significant main effects of ability, $F(1, 245) = 31.960$, $p < .001$, $\eta_p^2 = .115$; scoring method, $F(2, 244) = 268.152$, $p < .001$, $\eta_p^2 = .687$; and language, $F(1, 245) = 326.887$, $p < .001$, $\eta_p^2 = .572$. Pairwise comparisons indicated children with TD language scored higher across all scoring methods ($M = 89.39$) compared with children with SLI ($M = 77.50$), $p < .001$. For scoring method, holding language and ability constant, children scored lowest on single-language score ($M = 77.23$), followed by within-test conceptual score ($M = 80.66$), and highest on across-test conceptual score ($M = 92.45$), $p < .001$. Children scored higher in Spanish ($M = 95.74$) than in English ($M = 71.16$), $p < .001$.

There were significant Scoring Method × Language, $F(2, 244)$, $p < .001$, and Scoring Method × Ability interactions, $F(2, 244)$, $p < .001$. Post hoc analyses of the significant interactions were conducted using Scheffé's test for multiple comparisons. Results indicated that scores by language were differentially moderated by scoring method. Across-test conceptual scoring was consistently higher, as indicated by the significant main effect. Within-test conceptual scoring was higher for Spanish compared with single-language testing ($p < .01$), whereas there was no difference between the two approaches when scoring in English. For children with SLI, there were smaller differences between the within- and across-test conceptual scoring methods compared with the TD children, who had higher scores in the across-test scoring condition ($p < .05$).

Discriminant Analyses

Given that there were significant differences between children with SLI and children with TD language skills across scoring methods, we were interested in the degree to which each method correctly classified them by ability. We conducted six discriminant function analyses to examine classification accuracy for each scoring method.

To address our question about which method of scoring—single-language or conceptual scoring—on the English EOWPVT-3 (Brownell, 2000a) and EOWPVT-3: SBE (Brownell, 2001) results in higher diagnostic accuracy of SLI in bilingual children, six exploratory discriminant analyses were conducted. For each scoring method, we used discriminant function analysis to compare children

Table 2. Group means by three scoring approaches.

Scoring method		SLI (*n* = 38)		TD (*n* = 209)	
		M	*SD*	*M*	*SD*
Single-language	Spanish	81.63	18.26	95.23	18.58
	English	60.87	11.37	71.18	15.83
Within-test	Conceptual Spanish	88.76	16.92	99.48	17.37
	Conceptual English	62.11	11.26	72.29	15.62
Across-test	Conceptual Spanish	96.74	16.31	112.58	16.15
	Conceptual English	74.89	10.52	85.60	12.18

Note. *M* = mean (based on 100); *SD* = standard deviation (based on 15); SLI = specific language impaired; TD = typically developing.

with and without SLI and classification accuracy. The results are shown in Table 3. Generally, each of the scoring methods was statistically significant (*p* < .001). Overall classification ranged from 57.1% to 68.8%. The highest sensitivity resulted from single-language scoring in English at 81.6%; however, specificity was 52.6% (likelihood ratio [LR] + = 1.72, LR− = 0.35). The highest specificity resulted from across-test scoring in Spanish at 69.4% while sensitivity was 65.8% (LR+ = 2.15, LR− = 0.49). The best overall classification resulted from across-test scoring in Spanish at 68.8% (LR+ = 2.15, LR− = 0.49), and the lowest overall classification resulted from single-language scoring in English at 57.1% (LR+ = 1.72, LR− = 0.35). Results indicated that conceptual scoring, whether within or across tests, resulted in more accurate overall classification accuracy than single-language scoring when derived from either language norms.

Discussion

With the ever-increasing diversity in schools and communities, SLPs have a critical responsibility to accurately identify children who do and do not qualify for services. Currently, there is limited guidance regarding the diagnostic assessment of bilingual children, particularly on how to use assessment information that considers both languages, as do conceptual scores. Although conceptual scoring may be a valuable clinical tool for reducing the bias associated with single-word vocabulary measures, a clear method for obtaining a conceptual score has not been defined. This study used different scoring procedures to compare the discriminant validity of an expressive vocabulary measure in differentiating between bilingual children with and without SLI.

Performance Differences

Consistent with research demonstrating children with SLI exhibit slower vocabulary growth, results from the present study demonstrate that TD children consistently scored higher than children with language impairment. This was represented across all scoring methods. Across-test conceptual scoring resulted in higher scores

than within-test conceptual scoring, suggesting that bilingual children do not always readily switch between languages. These results support the language suppression hypothesis, which stipulates that language switch costs arise as a result of switching to a language that was inhibited on a previous trial (Green, 1998).

Speech production models address how message conceptualization, language formulation, and articulatory output occur incrementally (De Bot, 2004; Levelt, 1993, 1996). To complete a word naming task, a semantic concept is retrieved, the phonological and lexical word components are formulated, and the word is articulated to match the language context. Since a bilingual's two languages share conceptual characteristics, both languages are potentially activated during the semantic system's incremental lexical selection process (Isurin, Winford, & De Bot, 2009; Kormos, 2013, 2014; Patterson & Pearson, 2012). In a test situation, the test taker may face challenges in suppressing the response from the language in which the lexical item is more strongly activated. Meuter and Allport (1999) found that language switching costs were consistently larger (responses were slower) when switching to the dominant language from the weaker language than vice versa. They suggested that naming in the weaker language requires active inhibition or suppression of the stronger competitor language and that this inhibition persists into the subsequent switch trials in the form of "negative priming" of the dominant lexicon as a whole. It is possible that lower scores for both children with SLI and children with TD language skills in the within-test conceptual scoring condition are reflective of these switching costs.

Despite these findings, bilingual children with SLI demonstrated a significantly smaller gap in scores between the within- and across-test scoring conditions when compared with TD children, as evidenced by the interaction between scoring method and ability. For children with typical development, testing in one language at a time and then deriving a conceptual score allowed them to achieve higher scores. This provides evidence that they experienced more of the switch costs associated with within-test language switching. This was less the case for children with SLI, who showed a smaller difference in scores between the two methods. Children with SLI very often have smaller

Table 3. Discriminant functions comparing three different scoring approaches.

Discriminant function		Cutoff score	Overall classification	Sensitivity	Specificity	LR+ [95% CI]	LR− [95% CI]	Wilks's λ	χ^2	Canonical correlation squared	*p*
Single-language	Spanish	88.43	65.2	57.9	66.6	1.73 [1.24, 2.41]	0.63 [.43, .93]	0.93	16.71	0.07	< .001
	English	66.02	57.1	81.6	52.6	1.72 [1.40, 2.12]	0.35 [.18, .69]	0.94	14.27	0.06	< .001
Within-test	Conceptual Spanish	94.12	66.0	65.8	66.0	1.94 [1.44, 2.61]	0.52 [.33, .81]	0.96	12.01	0.05	.001
	Conceptual English	67.20	59.1	76.3	56.0	1.73 [1.37, 2.19]	0.42 [.24, .76]	0.94	14.27	0.06	< .001
Across-test	Conceptual Spanish	104.66	68.8	65.8	69.4	2.15 [1.58, 2.92]	0.49 [.31, .77]	0.88	29.00	0.11	< .001
	Conceptual English	80.25	63.6	68.4	62.7	1.83 [1.39, 2.42]	0.50 [.31, .81]	0.91	24.49	0.10	< .001

Note. LR = likelihood ratio; CI = confidence interval.

vocabularies, and while for them the one-language-at-a-time approach may have been optimal, they nonetheless demonstrated more restricted vocabulary knowledge compared with their typical peers under all conditions.

Classification Accuracy

The scoring procedures included single-language testing in English and Spanish, within-test conceptual scoring, and across-test conceptual scoring compared with monolingual English and bilingual norms. The highest classification accuracy was derived using an across-test conceptual scoring method, whereas single-language testing yielded the lowest classification accuracy, suggesting that across-test conceptual scoring is the most accurate method of assessing a bilingual child's Spanish and English vocabulary. Despite using empirically derived cutoff scores, however, classification accuracy was "unacceptable" (below 80%) throughout all scoring methods (Plante & Vance, 1994). Sensitivity and specificity ranged from 56% to 79%, which is consistent with the range (68%–77%) found in a study looking at classification accuracy of four monolingual English vocabulary tests (Gray et al., 1999). These findings provide convincing evidence that SLPs should not use vocabulary tests to identify language impairment. This is consistent with the stated purposes of the EOWPVT-3 and EOWPVT-3: SBE, which exclude using the tests for the diagnosis of language impairment (Brownell, 2000a, 2001), and with previous literature (Gray et al., 1999). These results show that conceptual scoring does not make up for these limitations. Reliance on a single test is insufficient for accurate diagnosis and should be considered in addition to parental and teacher report, clinical judgment, and in combination with other tests (Bishop & McDonald, 2009). There is also a need to select tests for which there is evidence of accurate identification. For example, Plante and Vance (1994) found that the Structured Photographic Expressive Language Test–II (Werner & Kresheck, 1983) provided 90% sensitivity and specificity, whereas the Test of Auditory Comprehension of Language–Revised (Carrow-Woolfolk, 1985) provided less than 80% sensitivity and specificity in the identification accuracy of children with SLI. When used in combination, these tests were less accurate than the use of the most accurate test alone. Within the clinical setting, consequences for misidentified children can have serious implications (Plante & Vance, 1994). An unidentified child with SLI misses the available opportunities for academic and communication support that can promote success in school. On the other hand, a child who is misidentified with SLI will likewise miss the opportunities and resources for academic success. Both can result in children whose emotional and developmental well-being is compromised.

The current results suggest that when the purpose of testing is descriptive in nature (i.e., describing a child's vocabulary knowledge), clinicians should assess bilingual vocabulary in each language and then derive conceptual scores to obtain a more accurate representation of a child's lexicon. However, it is important to note that by standards set forth by Plante and Vance (1994), none of the methods investigated achieved "fair" (at least 80%) much less "good" levels (at least 90%) of diagnostic accuracy. These findings are consistent with those of monolingual studies evaluating the diagnostic utility of single-word receptive and expressive tests (Gray et al., 1999). Thus, here, clinicians are cautioned against using expressive single-word vocabulary tests for SLI diagnoses.

Although our results support across-test conceptual scoring of expressive language, it is recognized that clinicians' evaluation time is limited. A third, less time-consuming manner of obtaining a conceptual score could potentially provide scores comparable to the across-test conceptual score. Specifically, children may first be tested in their more dominant language and then asked to return only to incorrect responses in their other language at the end of the administration. According to the EOWPVT-3: SBE manual, language dominance is determined by parent and teacher report or by administering 10 questions regarding language dominance. This procedure does not entail switching back and forth between languages, and children are required to respond to fewer items within one administration, reducing fatigue and administration time. The Bilingual Verbal Ability Tests (Muñoz-Sandoval, Cummins, Alvarado, & Ruef, 1998) use this method for obtaining conceptual scores where students are allowed to return to items they miss and respond in their other language. This approach was not directly assessed through the present study and thus merits further research.

Limitations

This study had a limited sample size, with only 38 participants with SLI in total. As such, there is some degree of uncertainty associated with the values found in this study. A larger sample would likely serve to increase confidence in our findings. Furthermore, in the current study, we analyzed conceptual scoring of vocabulary using a confrontation naming task. Results may differ with different language tasks, such as narrative elicitation or other naming tasks (e.g., category generation). It should also be noted that the EOWPVT-3 and the EOWPVT-3: SBE primarily elicit nouns, suggesting that these results may not be generalizable to other word classes.

Future Research

Future research should test whether dominance influences language suppression given that the language suppression hypothesis implies that the stronger a language is, the more strongly it will be suppressed when it is the non-target language. Therefore, switching to the stronger language should incur a cost. Although all children were asked to actively switch to the other language if they responded incorrectly, in the current study, only a small subset was able to do so successfully at least once. Specifically, 18.6% switched to Spanish during the English administration and

38.6% switched to English during the Spanish administration. The higher proportion of students switching to English compared with Spanish could be due, in part, to the context in which the children were tested, given that the majority was exposed to English more at school than in their home settings. Most often, children translated an incorrect answer in the target language or could not respond. Systematically exploring switch costs for children with and without SLI may better illustrate the circumstances under which bilinguals can and cannot successfully utilize both languages in an online task.

Acknowledgments

References

Alt, M. (2011). Phonological working memory impairments in children with specific language impairment: Where does the problem lie? *Journal of Communication Disorders, 44*(2), 173–185.

Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research, 47*(2), 407–420.

Alt, M., & Spaulding, T. (2011). The effect of time on word learning: An examination of decay of the memory trace and vocal rehearsal in children with and without specific language impairment. *Journal of Communication Disorders, 44*(6), 640–654.

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*(2), 84–97.

Bedore, L. M., Peña, E. D., Garcia, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Language, Speech, and Hearing Services in Schools, 36*(3), 188–200.

Bedore, L. M., Peña, E. D., Griffin, Z. M., & Hixon, J. G. (2016). Effects of Age of English Exposure, Current Input/Output, and grade on bilingual language performance. *Journal of Child Language, 43*(03), 687–706.

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*(2), 133–146.

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 859–873.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*(04), 525–531.

Bishop, D. V., & McDonald, D. (2009). Identifying language impairment in children: Combining language test scores with parental report. *International Journal of Language & Communication Disorders, 44*(5), 600–615.

Bracken, B. A., & McCallum, T. S. (1998). *Universal Nonverbal Intelligence Test*. Itasca, IL: Riverside Publishing.

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

Brownell, R. (2000b). *Receptive One-Word Picture Vocabulary Test*. Novato, CA: Academic Therapy Publications.

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

Campbell, J. M., Bell, S. K., & Keith, L. K. (2001). Concurrent validity of the Peabody Picture Vocabulary Test-as an intelligence and achievement screener for low SES African American children. *Assessment, 8*(1), 85–94.

Carrow-Woolfolk, E. (1985). *Test of Auditory Comprehension of Language–Revised*. Allen, TX: DLM Teaching Resources.

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish–English bilinguals from 22 to 30 months: Implications for assessment. *Journal of Speech, Language, and Hearing Research, 56*(5), 1637–1649.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50*(4), 491–511.

De Bot, K. (2004). The multilingual lexicon: Modelling selection and control. *International Journal of Multilingualism, 1*(1), 17–32.

Dollaghan, C. (1998). Spoken word recognition in children with and without specific language impairment. *Applied Psycholinguistics, 19*(02), 193–207.

Dunn, D. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition*. San Antonio, TX: Pearson.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., . . . Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development,* i–185.

Gillam, R. B., & Pearson, N. A. (2004). *Test of Narrative Language: Examiner's Manual*. Austin, TX: Pro-Ed.

Gillam, R. B., Peña, E. D., Bedore, L. M., Bohman, T. M., & Mendez-Perez, A. (2013). Identification of specific language impairment in bilingual children: I. Assessment in English. *Journal of Speech, Language, and Hearing Research, 56*(6), 1813–1823.

Gillam, R. B., Peña, E. D., Bedore, L. M., & Pearson, N. (in development). *Test of Narrative Language, Spanish Adaptation.*

Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost–benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 640.

Gray, S. (2004). Word learning by preschoolers with specific language impairment predictors and poor learners. *Journal of Speech, Language, and Hearing Research, 47*(5), 1117–1132.

Gray, S. (2005). Word learning by preschoolers with specific language impairment effect of phonological or semantic cues. *Journal of Speech, Language, and Hearing Research, 48*(6), 1452–1467.

Gray, S. (2006). The relationship between phonological memory, receptive vocabulary, and fast mapping in young children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 49*(5), 955–969.

Gray, S., & Brinkley, S. (2011). Fast mapping and word learning by preschoolers with specific language impairment in a supported learning context: Effect of encoding cues, phonotactic

probability, and object familiarity. *Journal of Speech, Language, and Hearing Research, 54*(3), 870–884.

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

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(02), 67–81.

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

Gutiérrez–Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics, 24*(02), 267–288.

Gutiérrez-Clellen, V. F., Restrepo, M. A., & Simón-Cereijido, G. (2006). Evaluating the discriminant accuracy of a grammatical measure with Spanish-speaking children. *Journal of Speech, Language, and Hearing Research, 49*(6), 1209–1223.

Hemsley, G., Holm, A., & Dodd, B. (2010). Patterns in diversity: Lexical learning in Samoan-English bilingual children. *International Journal of Speech-Language Pathology, 12*(4), 362–374.

Hollingshead, A. B. (1975). *Four factor index of social status.* Unpublished manuscript, Yale University, New Haven, CT.

Horohov, J. E., & Oetting, J. B. (2004). Effects of input manipulations on the word learning abilities of children with and without specific language impairment. *Applied Psycholinguistics, 25*(01), 43–65.

Iglesias, A. (2001). What test should I use? *Seminars in Speech and Language, 22,* 3–14.

Isurin, L., Winford, D., & De Bot, K. (2009). *Multidisciplinary approaches to code switching* (Vol. 41). Philadelphia, PA: John Benjamins.

Jackson-Maldonado, D., & Bates, E. (1988). *Inventario del Desarrollo de Habilidades Comunicativas.* San Diego, CA: University of California, Center for Research on Language.

Junker, D. A., & Stockman, I. J. (2002). Expressive vocabulary of German-English bilingual toddlers. *American Journal of Speech-Language Pathology, 11*(4), 381–394.

Kambanaros, M., Michaelides, M., & Grohmann, K. K. (2015). Measuring word retrieval deficits in a multilingual child with SLI: Is there a better language? *Journal of Neurolinguistics, 34,* 112–130.

Kiernan, B., & Gray, S. (1998). Word learning in a supported-learning context by preschool children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41*(1), 161–171.

Kormos, J. (2013). Sentence production in a second language. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics.* Hoboken, NJ: Wiley.

Kormos, J. (2014). *Speech production and second language acquisition.* Abingdon, UK: Routledge.

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica, 128*(3), 416–430.

Leonard, L. B. (2014). *Children with specific language impairment.* Cambridge, MA: MIT Press.

Levelt, W. J. (1993). *Speaking: From intention to articulation* (Vol. 1). Cambridge, MA: MIT Press.

Levelt, W. J. (1996). A theory of lexical access in speech production. *Proceedings of the 16th Conference on Computational Linguistics, 1,* 3.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language:

Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*(12), 1507–1515.

Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., & Gillam, R. B. (2015). Utility of a language screening measure for predicting risk for language impairment in bilinguals. *American Journal of Speech-Language Pathology, 24*(3), 426–437.

Marchman, V. A., & Martinez-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *Journal of Speech, Language, and Hearing Research, 45*(5), 983–997.

McCauley, R. J., & Swisher, L. (1984). Use and misuse of norm-referenced test in clinical assessmenta hypothetical case. *Journal of Speech and Hearing Disorders, 49*(4), 338–348.

McGregor, K. K., Newman, R. M., Reilly, R. M., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 45*(5), 998–1014.

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1), 25–40.

Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language, 67*(1), 224–237.

Muñoz-Sandoval, A. F., Cummins, J., Alvarado, C. G., & Ruef, M. L. (1998). *Bilingual Verbal Ability Tests* (BVAT). Itasca, IL: Riverside Publishing.

Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research, 48*(2), 439–458.

Oetting, J. B., Rice, M. L., & Swank, L. K. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research, 38*(2), 434–445.

Ordóñez, C. L., Carlo, M. S., Snow, C. E., & McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology, 94*(4), 719.

Patterson, J. L., & Pearson, B. Z. (2004). Bilingual lexical development: Influences, contexts, and processes. In B. A. Goldstein (Ed.), *Bilingual language development and disorders in Spanish–English speakers* (pp. 77–104). Baltimore, MD: Brookes.

Patterson, J. L., & Pearson, B. Z. (2012). Bilingual lexical development, assessment and intervention. In B. A. Goldstein (Ed.), *Bilingual language development and disorders in Spanish–English speakers* (pp. 113–129). Baltimore, MD: Brookes.

Pearson, B. Z. (2001). Logic and mind in Spanish–English children's narratives. In L. Verhoeven & S. Strömqvist (Eds.), *Narrative development in a multilingual context* (Vol. 23, pp. 373–398). Amsterdam, The Netherlands: John Benjamins.

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

Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*(1), 93–120.

Peña, E. D. (2001). Assessment of semantic knowledge: Use of feedback and clinical interviewing. *Seminars in Speech and Language, 22*(1), 51–64.

Peña, E. D., Bedore, L. M., Gutiérrez-Clellen, V. F., Iglesias, A., & Goldstein, B. A. (2010a). *Bilingual English-Spanish Assessment—Middle Extension, Experimental Version (BESA-ME)* (Unpublished test). University of Texas at Austin, Austin, TX.

Peña, E. D., Bedore, L. M., Gutierrez-Clellen, V. F., Iglesias, A., & Goldstein, B. A. (2010b). *Bilingual English Spanish Oral Screener (BESOS): Pre-K, 1st, and 3rd grade* (Unpublished test). University of Texas at Austin, Austin, TX.

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

Peña, E. D., Bedore, L. M., & Zlatic-Giunta, R. (2002). Category-generation performance of bilingual children: The influence of condition, category, and language. *Journal of Speech, Language, and Hearing Research, 45*(5), 938–947.

Peña, E. D., Gutierrez-Clellen, V., Iglesias, A., Goldstein, B., & Bedore, L. M. (2014). *BESA: Bilingual English-Spanish Assessment Manual*. San Rafael, CA: AR-Clinical Publications.

Perme, A. L. (2014). *Measures of narrative performance in Spanish-speaking children on the Test of Narrative Language–Spanish*. (Unpublished master's thesis). University of Texas at Austin, Austin, TX.

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25*(1), 15–24.

Rescorla, L., Roberts, J., & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech, Language, and Hearing Research, 40*(3), 556–566.

Rice, M. L., Buhr, J. C., & Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Disorders, 55*(1), 33–42.

Rice, M. L., Buhr, J., & Oetting, J. B. (1992). Specific-language-impaired children's quick incidental learning of words: The effect of a pause. *Journal of Speech, Language, and Hearing Research, 35*(5), 1040–1048.

Rice, M. L., Cleave, P. L., & Oetting, J. B. (2000). The use of syntactic cues in lexical acquisition by children with SLI. *Journal of Speech, Language, and Hearing Research, 43*(3), 582–594.

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 37*(1), 106–122.

Rice, M. L., Tomblin, J. B., Hoffman, L., Richman, W. A., & Marquis, J. (2004). Grammatical tense deficits in children with SLI and nonspecific language impairment relationships with nonverbal IQ over time. *Journal of Speech, Language, and Hearing Research, 47*(4), 816–834.

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development, 83*(2), 508–525.

Sheng, L., & McGregor, K. K. (2010). Object and action naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 53*(6), 1704–1719.

Sheng, L., Peña, E. D., Bedore, L. M., & Fiestas, C. E. (2012). Semantic deficits in Spanish–English bilingual children with language impairment. *Journal of Speech, Language, and Hearing Research, 55*(1), 1–15.

Spaulding, T. J. (2010). Investigating mechanisms of suppression in preschool children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 53*(3), 725–738.

Tomblin, J. B., & Zhang, X. (1999). Language patterns and etiology in children with specific language impairment. In H. Tager-Flusberg (Ed.), *Neurodevelopmental disorders: Contributions to a new framework from the cognitive neurosciences* (pp. 361–382). Cambridge, MA: MIT Press.

Umbel, V. M., Pearson, B. Z., Fernández, M. C., & Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development, 63*(4), 1012–1020.

Weismer, S. E., & Hesketh, L. J. (1998). The impact of emphatic stress on novel word learning by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41*(6), 1444–1458.

Werner, E. O. H., & Kresheck, J. D. (1983). *Spanish Structured Photographic Expressive Language Test–II*. DeKalb: Janelle Publications, Inc.

Williams, K. T. (1997). *Expressive Vocabulary Test–Second Edition* (EVT-2). Circle Pine, MN: American Guidance Service.

Yan, S., & Nicoladis, E. (2009). Finding le mot juste: Differences between bilingual and monolingual children's lexical access in comprehension and production. *Bilingualism: Language and Cognition, 12*(03), 323–335.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool Language Scale–Fifth Edition*. San Antonio, TX: The Psychological Corporation.