

Research Article

Synthesizing Information From Language Samples and Standardized Tests in School-Age Bilingual Assessment

Kerry Danahy Ebert^a and Giang Pham^b

Purpose: Although language samples and standardized tests are regularly used in assessment, few studies provide clinical guidance on how to synthesize information from these testing tools. This study extends previous work on the relations between tests and language samples to a new population—school-age bilingual speakers with primary language impairment—and considers the clinical implications for bilingual assessment.

Method: Fifty-one bilingual children with primary language impairment completed narrative language samples and standardized language tests in English and Spanish. Children were separated into younger (ages 5;6 [years; months]–8;11) and older (ages 9;0–11;2) groups. Analysis included correlations with age and partial correlations

between language sample measures and test scores in each language.

Results: Within the younger group, positive correlations with large effect sizes indicated convergence between test scores and microstructural language sample measures in both Spanish and English. There were minimal correlations in the older group for either language. Age related to English but not Spanish measures.

Conclusions: Tests and language samples complement each other in assessment. Wordless picture-book narratives may be more appropriate for ages 5–8 than for older children. We discuss clinical implications, including a case example of a bilingual child with primary language impairment, to illustrate how to synthesize information from these tools in assessment.

Clinical language assessment of school-age children serves multiple purposes. The main goals are to identify a disorder, describe a child's language system, plan for treatment, and monitor ongoing progress (Kohnert, 2013). To accomplish these varied goals, a comprehensive assessment includes direct measures of a child's language as well as indirect measures that have been gathered from reviewing existing educational and medical information, interviewing parents and teachers, and systematically observing within structured and unstructured settings. Cheng (1997) termed this comprehensive assessment framework the *RIOT process*: review, interview, observation, and testing. For bilingual children, it is essential to include both first and second languages (L1 and L2) in this assessment framework.

The present article focuses on the testing portion of the RIOT process, which can serve multiple purposes

within a comprehensive language assessment. Two types of testing tools—standardized tests and language samples—are considered for a group of bilingual children with primary language impairment (PLI). We examine the extent to which these two types of assessment tools are related within each language and discuss how our results can contribute to a broader understanding of bilingual PLI. We present a case example to illustrate how the two sources of information can be integrated to address two specific assessment purposes: describing the overall language system and planning for treatment. Before presenting the data, we frame the study by reviewing the uses of standardized tests and language samples in the assessment of school-age bilingual children. We present prior work on the relations between language samples and standardized tests, and provide a brief overview of the bilingual PLI profile.

Tools for Bilingual Language Assessment

A number of possible tools exist for clinicians to use in the language assessment of bilingual children. Within the RIOT framework, interview tools (such as parent questionnaires and ethnographic interviews) and observation techniques are crucial (for a discussion, see De Lamo White

^aRush University, Chicago, IL

^bSan Diego State University, CA

Correspondence to Kerry Danahy Ebert: Kerry_ebert@rush.edu

Editor: Shelley Gray

Associate Editor: Monique Mills

Received January 28, 2016

Revision received May 16, 2016

Accepted July 29, 2016

DOI: 10.1044/2016_LSHSS-16-0007

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

& Jin, 2011). Additional testing options may include dynamic assessment tasks and processing-based tasks (for a review, see Ebert & Kohnert, 2016). Though the focus of the present study is on two specific testing tools (i.e., standardized tests and language samples), it is important to recognize the full range of options available to clinicians.

Standardized Tests

Despite the array of tools available, survey data indicate a persistent and heavy reliance on English standardized tests in the language assessment of bilingual children. In a sample of school-based speech-language pathologists in the United States, Caesar and Kohler (2007) documented greater use of formal (standardized) tests in English, in comparison to less formal measures (including language samples, parent or teacher interview, and observations). In more recent work, Williams and McLeod (2012) documented similar practices in a group of Australian speech-language pathologists: 81.9% of survey respondents either always or usually included English standardized tests when assessing the language skills of bilingual children. Clinicians may feel drawn to standardized tests because they can offer a structured means to probe multiple different language skills (e.g., Paul & Norbury, 2012), and perhaps because test scores may be more easily accepted and interpreted by other educational and health professionals than measures such as language samples and interviews.

Several perils are present in the use of English standardized tests in the assessment of bilingual children, many of which are exacerbated when such tests are used exclusively. First, the vast majority of English tests do not include bilingual children in the normative sample, essentially invalidating the use of test norms for these children (e.g., De Lamo White & Jin, 2011; Heilmann, Rojas, Iglesias, & Miller, 2016; Kohnert, 2013). In addition, because such tests were originally designed to assess children who speak only English and share a relatively homogeneous cultural background, concerns with content and linguistic bias arise (De Lamo White & Jin, 2011). Last, standardized tests also have more general weaknesses that apply across populations, including the monolingual English-speaking population for which they were designed. These tests assess language in a highly decontextualized manner and are frequently considered suboptimal for the development of specific treatment targets (Ebert & Scott, 2014; Paul & Norbury, 2012).

Although imperfect, English standardized tests may have some value in the assessment of bilingual children. Most research in this area has considered the diagnostic accuracy of English tests or their ability to discriminate between children who do and do not have language disorders. A recent meta-analysis (Dollaghan & Horner, 2011) found that several tools, including standardized testing of English morphosyntax, have promise for contributing to accurate diagnoses. More recent evidence (Gillam, Peña, Bedore, Bohman, & Mendez-Perez, 2013) has suggested that English testing may be particularly valuable in ruling out the presence of a language disorder in children who are sequentially

bilingual (though it remains inadequate for ruling in, or confirming the presence of, a disorder).

For Spanish–English bilingual children, some of the pitfalls in the use of standardized tests have been addressed with the development of tests that are norm referenced on bilingual populations. Several tests have been published for Spanish–English speakers that include normative samples of school-age bilingual populations in the United States, such as the Bilingual English Spanish Assessment (Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2014), the Spanish version of the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-S; Wiig, Semel, & Secord, 2006), and the bilingual version of the Expressive and Receptive One-Word Picture Vocabulary Tests (Brownell, 2001a, 2001b). Such tests now allow clinicians to compare Spanish–English bilingual children with their peers in the normative sample. However, even the expanded testing options do not perfectly represent all Spanish dialects and regions of the United States. Capturing the heterogeneity of bilingual experiences (e.g., age of exposure to L2) within test norms remains challenging.

No single measure can adequately assess the language of school-age bilingual children (Kohnert, 2013). Even for the restricted purpose of identification, the most promising diagnostic accuracy measures must be supplemented by other sources (Dollaghan & Horner, 2011). For a more comprehensive evaluation, standardized tests may retain a place in the assessment of bilingual children (Caesar & Kohler, 2007; Williams & McLeod, 2012), but they should unquestionably be considered in light of other sources of information. Thus, it is important to examine how these tests relate to other assessment tools such as language samples.

Language Samples

Language sample analysis has recently been called the gold standard in language assessment with bilingual children (Heilmann et al., 2016), though it has been used in language assessment for decades. This approach does have a number of benefits, particularly in contrast to standardized tests. Language samples capture contextualized skills that may align with academic language demands (Bedore, Peña, Gillam, & Ho, 2010). Collection of language samples is quick and can be adapted to the child and the setting; for bilingual children, language samples can be collected in a home language via collaboration with an interpreter or trained bilingual paraprofessional (for guidelines, see Langdon & Saenz, 2015).

One of the benefits of using language samples in assessment is that a number of measures can be derived from a single sample. The language sample can provide information on multiple language dimensions at micro- and macrostructural levels. At the microstructural level, Rojas and Iglesias (2009) recommended three specific measures for use with Spanish–English bilingual speakers: mean length of utterance in words (MLUW), number of different words (NDW), and words per minute (WPM). MLUW indexes syntactic skill, NDW measures lexical diversity, and WPM indicates overall verbal fluency (Rojas & Iglesias, 2009). In

addition to these three measures, the percentage of grammatical utterances (Gram) may be a useful measure for bilingual children (Bedore et al., 2010; Simon-Cerejido & Gutiérrez-Clellen, 2007). This measure indexes grammatical development and may be particularly sensitive to language impairment in developing bilingual speakers.

Language samples can also be analyzed at the macrostructural level to examine story structure and organization, and several scoring systems have been developed. For narratives, one prominent measure of macrostructure is the Narrative Scoring Scheme (NSS; Heilmann, Miller, Nockerts, & Dunaway, 2010), which has been widely used to analyze English and Spanish language samples from bilingual children (Heilmann et al., 2016). The NSS incorporates seven aspects of narrative macrostructure: introduction, character development, mental states, referencing, conflict resolution, cohesion, and conclusion. Each aspect is rated on a 5-point ordinal scale, and ratings from the seven aspects can be summed to form a total score.

Though language samples are a recommended component of assessment for bilingual children, they too are imperfect tools when used in isolation. For example, language samples do not directly assess receptive language skills. In addition, although a child's performance on language sample measures can be compared to values in a normative database (e.g., Miller & Iglesias, 2012), the databases are not technically equivalent to the psychometric norms that accompany tests (Condouris, Meyer, & Tager-Flusberg, 2003). As a final consideration, the type of language sample that is collected influences the information that can be gained from it. It is important to match the task (e.g., narrative, conversational, or expository) to the child's ability level and the assessment purpose in order to get accurate information about language skills (Nippold, Mansfield, Billow, & Tomblin, 2008).

Relations Between Standardized Tests and Language Sample Measures

Thus far we have discussed language samples and standardized tests separately. It is important to understand how these two types of tools relate to each other, particularly for bilingual children, because we have advocated for the use of multiple tools in assessment. If language sample measures correlate strongly with standardized-test scores, it provides evidence for convergent validity of the tools. Convergent validity refers to the extent to which tools that purport to assess the same construct actually provide similar results (for a discussion, see Greenslade, Plante, & Vance, 2009). However, very high correlations would suggest that the tools provide nearly identical information; if this were the case, clinicians might save time by eliminating one tool or the other when conducting assessments. On the converse, weak or even negative associations between test scores and language sample measures would indicate that the tools do not measure the same abilities, or that they have poor convergent validity.

Ebert and Scott (2014) recently explored this question within a sample of school-age children (ages 6;0 [years; months]–12;8) referred for clinical language assessment. The sample was diverse in terms of language ability (approximately 30% had language skills within the average range and the remaining 70% were judged to have a language disorder) and in terms of language background (23% spoke African American English and another 23% had some exposure to another language but spoke primarily English). Because an overall goal of the study was to examine relations between English standardized tests and English language samples within an ecologically valid sample of diverse learners, the sample was not separated by language ability or language background. However, given the wide age range, the sample was divided into younger (ages 6;0–8;11) and older (ages 9;0–12;8) groups. Correlations between 11 English standardized-test scores and five microstructural measures from narrative language samples in English were calculated. For the younger group, 18 significant correlations between language sample measures and test scores were found. Significant correlations had coefficients in the range of $r = .36-.67$; Ebert and Scott interpreted these values as moderate correlations and suggested that the two types of measures may assess related but not identical aspects of language skill in younger children. For the older group, only four significant correlations were found, indicating that the associations between narrative language sample measures and standardized-test scores were weaker in this group.

In an additional analysis with the same group of children, Ebert and Mikolajczyk (2016) focused on macrostructural language sample measures and found no significant correlations between total NSS scores and five standardized-language-test scores. In contrast to the positive associations between microstructural measures of language samples and test scores found at least within the younger age group (Ebert & Scott, 2014), macrostructural measures may not be related to standardized tests that do not specifically measure skills at the discourse level (Ebert & Mikolajczyk, 2016).

These studies provide an important first step in considering relations between language sample measures and standardized-test scores for English-speaking children. There has been limited work exploring these relations within groups of bilingual children. In a group of Spanish-English bilingual kindergarteners of varying language ability ($n = 170$), Bedore et al. (2010) considered correlations between language sample measures in both L1 and L2 and a single composite language score gathered from the Bilingual English Spanish Assessment. That score was related to three measures from the English language samples (MLUW, NDW, and Gram; range of $r = .29-.46$) and two measures from the Spanish language samples (MLUW and Gram; range of $r = .18-.29$). Modest, positive associations showed convergence between some language sample measures (particularly MLUW and Gram) and the standardized-language-test score.

The three studies reviewed thus far (Bedore et al., 2010; Ebert & Mikolajczyk, 2016; Ebert & Scott, 2014)

included children with a range of language abilities and did not separate children with and without language disorders in their correlation analyses. A few studies have considered relations between language samples and standardized tests within groups of monolingual children with identified language disorders such as autism or PLI (e.g., Bishop & Donlan, 2005; Condouris et al., 2003). Correlations between language sample measures and standardized tests have generally been positive and small to moderate in magnitude, although results have varied by study. The variability may be explained by cross-study differences in specific language-sample measures, collection procedures, and participant characteristics (for a more complete review of this literature, see Ebert & Scott, 2014). Perhaps most important for the present study, the existing literature provides little in the way of performance expectations for bilingual children with language disorders, and even less on how to translate expectations into clinical practice.

Brief Overview of the PLI Profile for School-Age Bilingual Children

There has been a surge of studies involving school-age bilingual children with PLI. Although a comprehensive review of the literature is beyond the scope of this article, we highlight three characteristics of this clinical population that have implications for assessment. First, bilingual children with PLI, by definition, show low performance in both of their languages (Kohnert, 2013), and this must be captured in assessment. Low performance can be measured in standardized-test scores when bilingual children are adequately represented in the normative sample (e.g., Peña et al., 2014).

Second, different narrative skills may develop at different rates. Squires et al. (2014) examined narratives from Spanish–English bilingual children with and without PLI in kindergarten and again in first grade. At each grade level, bilingual children with PLI scored lower than their peers with typical development in each language on measures of narrative micro- and macrostructure. Within the 1-year time interval, the PLI group showed increases in macrostructure scores in English and Spanish, whereas microstructure scores remained the same. When controlling for English input and output, macrostructure scores were related across Spanish and English, whereas microstructure scores were not. These findings suggest that both micro- and macrostructural measures may be sensitive to the presence of language impairment in bilingual children. In addition, bilingual children with PLI may show greater improvements in macrostructure, in part due to the cross-language transfer of skills at the discourse level.

The third characteristic of the PLI profile in bilingual children is a possible increased risk for L1 loss. Restrepo and Kruth (2000) first suggested this possibility, on the basis of a case study of a bilingual child with PLI who showed a decrease in MLUW and an increase in grammatical errors in L1 over time. Using a subset of children from the present study, Ebert, Pham, and Kohnert (2014) found positive

associations between age and English on four measures of lexical knowledge and processing. Correlations between age and Spanish measures were not significantly different from zero, suggesting a plateau of skills in Spanish for bilingual children with PLI. Loss of L1 skills has implications for treatment planning, and we return to these characteristics of the bilingual PLI profile in our Discussion.

Study Purpose and Questions

The purpose of this study is twofold. First, we aim to extend prior work considering the relations between standardized tests and language samples (e.g., Ebert & Scott, 2014) to a group of Spanish–English bilingual children with an identified language disorder. We consider relations within both languages in order to have a comprehensive picture of how each type of tool can contribute to an overall language assessment. Second, we aim to consider the clinical implications of these results for practice with school-age bilingual children. We discuss how our group findings can inform the profile of PLI in school-age bilingual children, and we focus on the potential contribution of each tool type to meet different purposes within an assessment. Together, these aims are intended to provide guidance for clinicians who conduct language assessments with school-age bilingual children.

To address our first aim, we consider the following research questions:

1. How well do English standardized-test scores relate to English language sample measures within a sample of school-age bilingual children with PLI?
2. How well do Spanish standardized-test scores relate to Spanish language sample measures within a sample of school-age bilingual children with PLI?
3. Do relations between these two types of tools differ according to age?

On the basis of prior work, we anticipate that these tools will provide similar, but not identical, information about the language skills of the participants. Moreover, measures that assess similar aspects of language should provide more convergent results than others. For example, Bedore et al. (2010) indicated that grammaticality and sentence length (Gram and MLUW) may relate most closely to a composite test score. NSS scores, which reflect narrative macrostructure, may have weak associations with test scores, because these tests do not directly assess macrostructure (Ebert & Mikolajczyk, 2016). In terms of Question 2, there is neither an a priori rationale nor prior empirical work to suggest that associations between the two types of tools would be stronger in Spanish than in English (or vice versa). Prior work does suggest that age will play a role, however (Question 3): Ebert and Scott (2014) found substantially stronger correlations between standardized tests and language sample measures in 6- to 8-year-old children than in children aged 9 and above.

To address our second aim, we consider the following questions:

1. How can results from standardized-test scores and language sample measures inform the broader clinical understanding of PLI in school-age bilingual children?
2. How can results from standardized-test scores and language sample measures be integrated within an individual assessment?

On the basis of previous work with school-age bilingual children with PLI (Ebert, Pham, & Kohnert, 2014; Squires et al., 2014), we expect to find minimal increases in L1 skill across most areas, with the possible exception of macrostructural measures. We will illustrate the use of language sample analysis as a complement to test results for an individual child (Question 5). This descriptive exercise, designed to provide an in-depth example for a clinical audience, is presented toward the end of the article.

Method

Participants

A total of 51 school-age Spanish–English bilingual children with PLI participated in this study. The children were recruited through a metropolitan public school district in the upper Midwestern United States to participate in a treatment study (Ebert, Kohnert, Pham, Rentmeester Disher, & Payesteh, 2014). Recruitment followed district procedures; all children who met study criteria were identified by the central district administrative office, and children attending schools that were able to host the treatment study were invited to participate.

All data reported in this study were collected at initial study intake, prior to the delivery of any treatment (for an analysis of other measures at this initial time point, see also Ebert, Pham, & Kohnert, 2014). For all participants, Spanish was used in the home either all or most of the time, according to parent report, and English was used almost exclusively in the school. Thus, these children can be classified as sequentially bilingual. Children ranged in age from 5;6 to 11;2, with a mean age of 8;5. There were nine girls and 42 boys.

Participants qualified to receive school-based speech-language services for language disorder and also had parent-reported concerns about language development. In addition, study testing confirmed that they had hearing within normal limits; nonverbal intelligence scores within the average range; and no history of hearing loss, autism, head injury, cerebral palsy, seizures, or other developmental concerns (for additional details, see Ebert, Pham, & Kohnert, 2014). Study testing also confirmed that all participants showed language skills below developmental expectations in both Spanish and English. Standard scores on the Core Language composite score of the Clinical Evaluation of Language Fundamentals–Fourth Edition in English (CELF-E, Semel, Wiig, & Secord, 2003) ranged from 40 to 69, with a mean of 50; scores on the Core Language composite of the same

test in Spanish (CELF-S) ranged from 45 to 87, with a mean of 62. It is important to note that standard scores are not comparable across languages, because the CELF-E is normed on monolingual English-speaking children, whereas the CELF-S is normed on Spanish–English bilingual children.

Due to the wide range of ages represented in the sample, the sample was divided into two age groups for analyses. The younger group ($n = 28$) consisted of children ages 5;6–8;11, and the older group ($n = 23$) consisted of children ages 9;0–11;2. The specific cut-point of 9 years was selected on the basis of prior work (Ebert & Scott, 2014). Table 1 reports participant scores on study language measures as a function of age group for both English and Spanish measures.

Procedure

All assessment procedures were administered in a quiet location in the participant's school, during either an after-school or summer-school program. Assessments were completed in two to four sessions of 75–90 min each; this time includes additional pretreatment assessment measures that are not reported here (see Ebert, Kohnert, et al., 2014). Assessment tools were administered by trained research assistants fluent in the target language. Research assistants were either certified speech-language pathologists or students in the speech-language-hearing sciences. The order of assessment for the two languages was counterbalanced across participants.

Assessment Measures

The present study considers two types of measures: standardized tests and narrative language samples. The specific measures within each category are described below.

Standardized Tests

Three standardized tests were given in each language. First, all children completed the four subtests composing the Core Language composite score of the CELF-E and the CELF-S. Three of these subtests are consistently administered for children of all ages: Concepts & Following Directions, Recalling Sentences, and Formulated Sentences. Concepts & Following Directions is a measure of conceptual knowledge and verbal memory, in which children listen to single- and multistep instructions and point to the corresponding pictures in the order they are named. Recalling Sentences is a measure of verbal memory and grammatical knowledge, in which children repeat sentences of increasing length and complexity. Formulated Sentences is a measure of expressive grammar and vocabulary, in which children are presented with pictures and asked to generate a sentence for each. The fourth subtest of the Core Language composite varies by age: Children under 9 years old complete the Word Structure subtest and children age 9 and above complete the Word Classes subtest. Because of this variation, scores from this final subtest are not considered here. Test-retest reliability for the CELF-E and CELF-S subtests and

Table 1. Participant performance on language measures in English and Spanish.

Measure	English						Spanish					
	M		SD		Range		M		SD		Range	
	Younger	Older	Younger	Older	Younger	Older	Younger	Older	Younger	Older	Younger	Older
CFD	11.5	35.4	7.1	5.6	2–30	23–46	15.0	25.7	8.5	6.3	5–34	12–36
RS	9.9	25.9	8.2	7.8	0–35	7–45	10.4	13.5	9.9	8.5	0–44	1–38
FS	7.2	22.0	7.5	3.4	0–24	17–30	8.6	13.2	7.9	6.5	0–25	5–26
EOW	36.2	51.0	13.5	12.5	2–68	31–80	32.1	30.6	10.7	14.0	17–57	2–55
ROW	52.4	71.1	12.5	11.5	24–75	49–98	44.9	52.5	15.2	17.8	20–68	20–81
MLUW	5.3	6.1	1.2	0.7	2.7–7.2	4.5–7.3	4.9	5.6	1.1	0.5	2.2–7.6	4.9–7.4
NDW	66.4	74.1	24.9	19.8	21–107	39–106	65.9	64.5	22.8	15.9	22–105	33–90
WPM	79.4	89.0	25.9	18.5	27.3–121.5	52.4–121.6	73.0	80.4	19.2	20.7	32.2–106.1	28.4–107.5
Gram	40.2	60.3	20.2	14.4	0.0–86.0	37.5–86.0	52.4	55.2	16.6	22.8	15.4–87.5	18.2–90.5
NSS	15.3	19.9	5.4	4.8	7–25	8–27	16.0	17.5	4.5	3.6	7–26	10–23
TNW	217.5	234.5	96.1	75.6	49–402	109–350	209.8	203.8	86.1	60.9	50–387	92–305
C-units	40.5	38.0	4.5	10.7	17–84	23–62	42.6	36.6	15.6	10.8	14–74	17–52

Note. Raw scores are reported for tests (CFD, RS, FS, EOW, ROW). The younger group comprised 28 children ages 5;6 (years;months) to 8;11, and the older group comprised 23 children ages 9;0 to 11;2. Scores for EOW and ROW in Spanish include only children who completed the test entirely in Spanish (vs. bilingually): $N = 37$ for EOW ($n = 18$ for younger, $n = 19$ for older) and $N = 36$ for ROW ($n = 17$ for younger, $n = 19$ for older). CFD = Clinical Evaluation of Language Fundamentals (CELF) Concepts and Following Directions subtest; RS = CELF Recalling Sentences subtest; FS = CELF Formulated Sentences subtest; EOW = Expressive One-Word Picture Vocabulary Test; ROW = Receptive One-Word Picture Vocabulary Test; MLUW = mean length of utterance in words; NDW = number of different words; WPM = words per minute; Gram = percentage of grammatical utterances; NSS = Narrative Scoring Scheme total score; TNW = total number of words; C-units = total number of modified C-units.

ages considered here is reported as correlation coefficients, which range from .79 to .94.

The second standardized test was a measure of expressive vocabulary, in which children were asked to name pictures in English (Expressive One-Word Picture Vocabulary Test [EOW-E]; Brownell, 2000a) or in Spanish (Expressive One-Word Picture Vocabulary Test–Spanish-Bilingual Edition [EOW-S]; Brownell, 2001a). The third test was a measure of receptive vocabulary, in which children were asked to point to pictures corresponding to a vocabulary word spoken in English (Receptive One-Word Picture Vocabulary Test [ROW-E]; Brownell, 2000b) or in Spanish (Receptive One-Word Picture Vocabulary Test–Spanish-Bilingual Edition [ROW-S]; Brownell, 2001b). Both Spanish vocabulary measures were administered only in Spanish, rather than bilingually as specified in the test instructions. In other words, credit on the EOW-S was given only for items named correctly in Spanish; on the ROW-S, all words were presented in Spanish only. This procedure was consistent with other language assessments, which were conducted in only one language at a time. However, the EOW-S and ROW-S were inadvertently administered bilingually to some participants; scores from these bilingual administrations are not included here. As a result, the number of participants with scores is lower for these measures ($n = 37$ for EOW-S; $n = 36$ for ROW-S). Test–retest reliability coefficients for the ROW and EOW tests range from .91 to .97.

Language Samples

The second type of measure was a narrative language sample, collected in English and in Spanish on separate days. Children were asked to tell their own story to the wordless

picture book *Frog, Where Are You?* (Mayer, 1969). They were provided with a brief introduction to the task in the target language, including explicit instructions to look through the book before beginning the story. Additional minimal, open-ended prompts were provided in the target language as needed (e.g., “Tell me more,” “¿Qué más?”). Examiners were instructed to ignore occasional words in the nontarget language, but to prompt the child if he or she used multiple words or phrases in the nontarget language (e.g., “In English, please,” “Dímelo en español”).

Language samples were audio-recorded and later transcribed by trained study staff fluent in the language of story administration. Samples were segmented into modified communication units (C-units) according to SALT guidelines for samples collected from bilingual children (Miller & Iglesias, 2012). Codes were then added for mazes, root words, and unintelligible segments, again following SALT guidelines. The second author, a Spanish–English bilingual speech-language pathologist, independently relistened to all language samples to verify transcription accuracy and modified C-unit segmentation and SALT coding; this is an established method of ensuring transcription and coding accuracy (Heilmann et al., 2008).

The coded transcripts were then analyzed using SALT software (Miller & Iglesias, 2012). For the present study, five language sample measures were calculated: MLUW, NDW, WPM, Gram, and total NSS score. MLUW, NDW, and WPM were calculated using the standard measures function of SALT. The total number of words and the total number of utterances (C-units) in each sample were also calculated, in order to fully characterize the language samples; these values are included in Table 1.

To calculate Gram, trained research assistants judged each utterance to be grammatical or ungrammatical. Utterances with at least one grammatical error were judged to be ungrammatical (e.g., “the kid jump out the window to get his dog,” “and he go floor,” “y un rana se quedó abajo”). Research assistants were instructed to ignore semantic information, including whether or not the utterance corresponded to the story, in making these judgments. To obtain interrater reliability for Gram coding, a total of 12 transcripts in each language were coded by an independent judge. Line-by-line agreement was 92.4% for English and 94.2% for Spanish.

NSS scores were assigned to each transcript by trained research assistants. Assistants were trained using the materials published by Heilmann et al. (2010) and Miller and Iglesias (2012), as well as an internal rubric developed specifically for the *Frog, Where Are You?* story. Each sample was assigned a rating of 1 to 5 for each of the seven components of the NSS. The scores were summed to create a total NSS score (maximum possible score = 35). A total of 12 stories in each language were independently rescored by a second judge to obtain interrater reliability. Krippendorff’s alpha (Krippendorff, 2004) was calculated as the reliability metric (Heilmann et al., 2010). Values were $\alpha = .902$ for English and $\alpha = .923$ for Spanish, well above the suggested minimum of value of .800 for reliable data (Krippendorff, 2004).

Analyses

Before conducting analyses, we examined skewness and kurtosis values for each dependent variable to verify that there were no violations of the normality assumption. We then examined the correlations between age and each of our dependent variables. Even after dividing our sample into two age groups, there was evidence that age might continue to influence our relations of interest (i.e., between tests and language sample measures; see Tables 2 and 3 for correlations between age and dependent variables). Therefore, we used partial-correlation analyses with age removed

to examine the relations between raw scores from standardized tests and language sample measures in each language. Separate analyses were conducted for Spanish and for English, as well as for younger and older age groups. In each language, raw scores from five tests (three subtests of CELF plus EOW and ROW) as well as four language-sample measures (MLUW, NDW, WPM, and Gram) were entered into the correlation. Raw scores were preferred to standard scores for these analyses because they eliminate reference to the normative samples (which differ between the Spanish and English tests). Raw test scores also parallel the language sample measures in that neither use scores that are transformed or adjusted for age. We interpreted the effect size of the correlation coefficient following Cohen’s (1988) guidelines: $r = .50$ is a large effect, $r = .30$ is a medium effect, and $r = .10$ is a small effect.

Results

English Assessment Tools

The results of the partial-correlation analyses for English test scores and language sample measures appear in Table 2. For the younger group, a total of nine correlations reached statistical significance. Gram correlated with four of five test scores, with coefficients ranging from .46 (CELF-E Recalling Sentences) to .54 (CELF-E Concepts & Following Directions). MLUW correlated with two of five test scores: CELF-E Concepts & Following Directions ($r = .53$) and CELF-E Recalling Sentences ($r = .43$). The remaining three language sample measures correlated with one standardized test each: NDW with EOW-E ($r = .51$), WPM with CELF-E Concepts & Following Directions ($r = .53$), and NSS with EOW-E ($r = .55$). The effect sizes for the significant correlations were medium to large. For the older group, only one correlation reached statistical significance: MLUW with ROW-E ($r = .45$). Remaining correlations were small to medium-size but not statistically different from zero in this sample. There were no significant negative correlations in either age group.

Table 2. Correlations with age and partial correlations (age removed) between English standardized tests and English language sample measures by age group.

Standardized test (raw score)	Language sample measure											
	Age		MLUW		NDW		WPM		Gram		NSS	
	Younger	Older	Younger	Older	Younger	Older	Younger	Older	Younger	Older	Younger	Older
Age	—	—	.59**	.17	.46*	.11	.34	.41	.60**	.09	.70**	-.05
CFD	.60**	-.01	.53**	.17	.35	-.05	.53**	-.16	.54**	.15	.19	.13
RS	.72**	.47*	.43*	.14	.28	.25	.38	.37	.46*	.08	.26	.36
FS	.65**	.16	.20	.14	-.13	.08	.10	.14	.51**	.10	-.14	.19
EOW	.54**	.58**	.27	.11	.51**	.34	.37	.08	.48*	.30	.55**	.28
ROW	.74**	.26	-.08	.45*	.23	.11	-.08	.13	.25	.25	.25	.29

Note. The top row and first two columns display bivariate correlations between chronological age and all English dependent measures. Remaining rows display partial correlations between English language sample measures and raw scores from English standardized tests. Abbreviations are as in Table 1.

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

Table 3. Correlations with age and partial correlations (age removed) between Spanish standardized tests and Spanish language sample measures by age group.

Standardized test (raw score)	Language sample measure											
	Age		MLUW		NDW		WPM		Gram		NSS	
	Younger	Older	Younger	Older	Younger	Older	Younger	Older	Younger	Older	Younger	Older
Age	—	—	.63**	-.17	.46*	.01	.07	.27	.01	-.52**	.66**	-.01
CFD	.40*	-.01	.44*	.39	-.11	-.16	.33	-.08	.21	.26	.58**	.18
RS	.37	-.03	.56**	.38	-.02	.41	.30	.22	.37	.37	.39*	.51*
FS	.43*	.01	.49**	-.02	-.23	.32	.22	.35	.59**	.35	.41*	.20
EOW	.04	-.11	.47	.37	-.04	.22	.57*	.39	.48*	.51*	.29	.45
ROW	.12	-.60**	.68**	.13	.46	-.09	.45	.14	.09	.23	.58*	.09

Note. The top row and first two columns display bivariate correlations between chronological age and all Spanish dependent measures. Remaining rows display partial correlations between Spanish language sample measures and raw scores from Spanish standardized tests. Abbreviations are as in Table 1.

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

Spanish Assessment Tools

Table 3 displays the partial correlations between Spanish test scores and Spanish language sample measures. For the younger group, 11 correlations reached statistical significance. MLUW correlated with four of five tests with medium to large effect sizes: Correlation coefficients ranged from .44 to .68. The NSS score also correlated with four of five tests—all three CELF-S subtests and ROW-S—with correlation coefficients ranging from $r = .39$ to .58. Gram correlated with two of five tests, CELF-S Formulated Sentences and EOW-S ($r = .59$ and $r = .48$). WPM correlated with EOW-S ($r = .57$), again with a large effect size. For the older group, two correlations reached significance: EOW-S correlated with Gram, and CELF-S Recalling Sentences correlated with the NSS; both correlations had the same coefficient ($r = .51$), indicating a large effect. There were no significant negative correlations in either age group.

Correlations With Age

For the younger group, age correlated with four of five language sample measures in English and five of five test scores in English (see Table 2). Correlation coefficients ranged from .46 to .74, suggesting large increases in English skills with age. In Spanish (see Table 3), age related to three of five language sample measures for the younger group. Age was positively related to Spanish NSS, MLUW, and NDW, suggesting increases in Spanish narrative quality, sentence length, and lexical diversity, at least in the younger age range. In contrast to relations with English tests, age related to two of five test scores in Spanish, with correlation coefficients ranging from .40 to .43, suggesting weaker growth or stagnation in Spanish skills in this age range.

In the older group, none of the English language-sample measures related significantly to age. Two of five English test scores related to age. In Spanish, age correlated significantly with one test score (ROW-S) and one language-sample measure (Gram), but both correlations were negative. Children in this age group appear to show decreased Spanish

receptive vocabulary skills and decreased Spanish grammaticality with age.

Discussion

The two main aims of this study were (a) to examine relations between tests and language sample measures within a group of bilingual children with PLI and (b) to consider clinical implications for bilingual assessment. Regarding the first aim, the results were consistent with prior work (e.g., Bedore et al., 2010; Bishop & Donlan, 2005; Ebert & Scott, 2014), which has generally found moderate positive associations between standardized-test results and language sample measures across different populations (e.g., monolingual clinical samples, bilingual children with typical development). The strongest correlations in this study fell near Cohen's (1988) benchmark for large effects ($r = .50$), providing some evidence of convergent validity within a new population (bilingual children with PLI) and in two languages (Spanish and English).

Even the largest correlations in our data, however, do not show complete overlap between the two tools. Language samples and tests appear to be related but not identical for school-age bilingual children. Moreover, there were several examples of divergence or nonsignificant relations between tools. Our findings indicate that standardized tests and language samples are not interchangeable (i.e., one cannot simply replace the other in assessment), because they provide different types of information.

Relations Within English Versus Spanish

Because clinicians should consider both languages when assessing bilingual children (Kohnert, 2013), we considered relations between the two types of assessment tools in Spanish and in English. A comparison of the results for our Questions 1 and 2 revealed many similarities. In both English and Spanish, the number of significant correlations was similar (i.e., a total of 10 correlations in English and

13 in Spanish). All significant correlations were positive, upholding the general conclusion of convergence between language samples and standardized tests for both English and Spanish.

Differences between the two languages were more subtle. For example, although measures of sentence length and grammaticality (i.e., MLUW and Gram) emerged in analyses in both languages, Gram had more significant correlations with test scores in English, whereas MLUW had more significant correlations with test scores in Spanish. Perhaps the most striking difference between English and Spanish was the relation between test scores and NSS. Consistent with previous work (Ebert & Mikolajczyk, 2016), correlations between NSS and test scores in English were minimal. In contrast, there were numerous relations between NSS and test scores in Spanish, at least for the younger group. We return to this cross-linguistic difference later when we address the second aim of the study.

Relations by Age

We considered the role of age by dividing our sample into two age groups. We hypothesized that language-sample measures and standardized tests would be more closely related in younger children (ages 5;6–8;11) than in older children (ages 9;0–11;2). The correlation results robustly support this hypothesis. Summing across languages, there were 20 significant correlations within the younger group and just three in the older group. Although the older group was slightly smaller than the younger group ($n = 23$ vs. $n = 28$), it is unlikely that sample size can explain these divergent results. Inspection of Tables 2 and 3 clearly indicates that correlation coefficients are larger in the younger group.

One explanation for the differing results across age groups is the nature of the language sample task. Language samples collected using retells of wordless picture books (such as *Frog, Where Are You?*) appear sensitive to growth in both Spanish and English for 5- to 8-year-old children, on the basis of a large sample of children with typical language development (Rojas & Iglesias, 2013) and a smaller sample of children with PLI (Squires et al., 2014). There has been little investigation to date of the suitability of picture-book storytelling tasks for bilingual children who are older than 8 years; the impact of procedural differences such as telling versus retelling the stories is also unexplored for bilingual children. For monolingual children, there is evidence that more complex tasks such as retelling fables (Nippold et al., 2015) and explaining a task (Nippold et al., 2008) may result in more complex language in older children. It is possible that one of these more complex tasks would have yielded language sample measures more convergent with standardized tests within the older group of bilingual children in the current study. This hypothesis should be tested in future studies.

Our findings support the use of wordless picture books to elicit narrative language samples from bilingual children in the 5- to 8-year-old age range. For older children, it may be necessary to use a more complex task or to use more

complex measures to assess language development accurately. For both groups, our findings support the clinical recommendation of using multiple measures within both languages to fully capture bilingual children's abilities. For clinicians, however, the most crucial question may be how to use multiple tools effectively. We devote the final section of this article to clinical implications, including an illustration of the complementary uses of language samples and standardized-test scores.

Integrating Information

The second aim of this study is to highlight the clinical implications of the group results. We first consider how the results can contribute to understanding the PLI profile in school-age bilingual children and then present an individual case example as a clinical illustration.

Bilingual PLI Profile

Consistent with prior studies that have measured performance in the two languages of bilingual children with PLI (e.g., Ebert, Pham, & Kohnert, 2014), our findings highlight the issue of L1 loss. We found numerous correlations between age and English skills, all of them positive in nature, indicating increases in English with age. In contrast, there were relatively fewer correlations between age and Spanish, showing that performance on many Spanish measures did not increase with age. Furthermore, two correlations between age and Spanish in the older group were negative (see Table 3), suggesting decreases in some Spanish skills for children ages 9–11.

Sequentially bilingual children in the United States are at risk for L1 loss due to lower social status of the non-English language and limited opportunities for input and practice (Pearson, 2007). Bilingual children with language disorders may be at an even greater risk for L1 loss, because they encounter the same social factors as their bilingual peers with typical development and struggle with additional difficulties in language learning (Ebert, Pham, & Kohnert, 2014; Restrepo & Kruth, 2000). Treatment planning for this population needs to include systematic support of the first language in order to increase communication between parents and children and promote children's overall well-being (Kohnert, Yim, Nett, Kan, & Duran, 2005).

Both bilingual and cross-linguistic approaches to treatment can be used to provide this L1 support alongside L2 learning opportunities (for discussion, see Kohnert, 2013; Kohnert & Derr, 2012). A bilingual treatment approach highlights overlapping features between the L1 and L2 to promote the transfer of skills across languages. A cross-linguistic treatment approach focuses on features that are specific to each language (i.e., where the L1 and L2 do not overlap); these features must be targeted separately. Bilingual and cross-linguistic approaches can be used in conjunction with each other to address children's overall communication needs. We include examples of each approach in our case example later.

Next we compare performance on the language-sample measures to a recent longitudinal study of bilingual children with PLI who completed similar narrative tasks in Spanish and English (Squires et al., 2014). That study found that bilingual children with PLI showed growth from kindergarten to first grade on macro- but not microstructural aspects in both languages. Consistent with those results, we found that the largest correlations with age were found with our macrostructural measure, NSS, in both languages in the younger group. Unlike the results of Squires et al., however, our data show that age was related to increases in microstructural skills as well, at least within the younger group. Differences between studies could stem from the measurement of microstructure. In the present study, we examined general measures of lexical and grammatical productivity (i.e., NDW and MLUW), while Squires et al. used specific linguistic features in their scoring system (e.g., mental verbs, elaborated noun phrase). An implication for assessment may be that general measures of microstructure may capture age-based differences, whereas fine-grain measures may help to inform treatment planning through identifying specific vocabulary and grammatical targets.

We now return to the cross-linguistic difference in the association between test scores and NSS scores, which were highly related in Spanish but unrelated in English. This cross-linguistic difference may reflect distinct contexts for learning each language. Because study participants received school instruction primarily in English, they had experience with academic or decontextualized language in English (including the skill of taking standardized tests). Zero correlations between test scores and NSS scores in English may reflect a dissociation between contextualized and decontextualized skills for this language. Children seemed to be developing both types of language skills in English (as shown in positive correlations with age), and these skills may not necessarily overlap. In contrast, participants spoke Spanish as the main home language and thus had exposure to contextualized language skills in this language. However, they presumably had less experience with decontextualized language in Spanish than in English, given the absence of academic instruction in Spanish. This pattern is reflected in the positive correlations between age and Spanish NSS and sparse correlations between age and Spanish test scores. For Spanish, it appears that children who performed better on narrative macrostructure also had more Spanish language skills in general, as reflected in higher test scores. This finding underscores the importance of including measures of both contextualized and decontextualized language skills in the assessment process for school-age children. Furthermore, the inclusion of contextualized language measures (such as narrative macrostructure) may be particularly helpful in capturing skills in the home language of bilingual children.

Case Example

We selected an individual child from the younger group for a more in-depth analysis. Child A was a boy, age 8;9. He achieved a standard score of 91 on a nonverbal

intelligence test, indicating skills within the average range in this area. In contrast, scores on language testing in both Spanish and English fell well below the average range. Table 4 summarizes his standardized-test scores and language sample measures, including comparisons to the normative data for tests and to the Bilingual Unique Story database in SALT (Miller & Iglesias, 2012).

Test results support the broad conclusion that Child A demonstrates a language disorder. For example, he performed more than 2 *SDs* below the mean in both languages on the omnibus language measure, the CELF Core Language composite. As we have discussed, the Spanish and English tests were based on different normative samples, limiting the ability to make direct comparisons of test scores across languages. In contrast, the language sample measures can be directly compared across languages because the SALT database is consistent across languages and includes only bilingual children. Compared with his same-age bilingual peers, Child A appears to be within 1 *SD* of the mean for the majority of language sample measures (see Table 4), highlighting a relative strength in his contextualized language skills. However, it is important to note that comparison samples from databases are not psychometrically equivalent to normative samples of standardized tests (Condouris et al., 2003) and that the geographic origins of the SALT bilingual databases (which were collected in Texas and California) do not match Child A's background (in the upper Midwest).

Language sample analysis can also provide a qualitative depiction of Child A's language skills in context. This analysis can also facilitate the formulation of treatment recommendations. Table 5 provides an in-depth analysis of Child A's strengths and weaknesses across the language domains of vocabulary, grammar, and narrative macrostructure in English and in Spanish, derived from his narrative language samples (see the Appendix for the corresponding raw transcripts).

In the vocabulary domain, standardized tests (i.e., EOW and ROW scores) suggest that Child A has a severe weakness in Spanish and a moderate deficit in English (see Table 4). Quantitative measures from the language samples are consistent with the test scores, as NDW is notably higher in English than in Spanish. Table 5 then provides a qualitative analysis of vocabulary, which reinforces the same conclusions. Child A used specific labels such as *moose* and *owl* in his English language sample, where he used more ambiguous terms in Spanish, such as *algo* (something). For this child, intervention might focus on building vocabulary, particularly in Spanish. Using a bilingual treatment approach, clinicians can select vocabulary targets that overlap in form and meaning between the languages (i.e., cognates, such as *elephant/elefante*; see Kelley & Kohnert, 2012). It will also be important to assess Child A's vocabulary needs on the basis of his educational and social context (using all the components of the RIOT framework) before setting specific goals. For example, the educational expectations in most states are now based on the Common Core State Standards (National Governors Association Center for Best Practices and Council of Chief State School Officers,

Table 4. Case example: Test scores and language sample measures for Child A.

Type	Measure	English		Spanish	
		Score	SDs from mean	Score	SDs from mean
Standardized test	CELF CFD	2	-2.67	8	-0.67
	CELF WS	2	-2.67	1	-3.00
	CELF RS	2	-2.67	3	-2.33
	CELF FS	1	-3.00	6	-1.33
	CELF Core	46	-3.60	65	-2.33
	ROW	72	-1.87	55	-3.00
	EOW	65	-2.33	55	-3.00
	Language sample measure	MLUW	6.2	-0.98	5.0
	NDW	85	0.40	65	-1.00
	WPM	107.6	0.45	73.0	-0.83
	Gram (%)	50.0		55.6	
	NSS	22	-0.21	21	-0.58

Note. Tests are reported as scaled scores for subtests (CFD, WS, RS, FS) and as standard scores for composites and full tests (CELF Core, ROW, EOW). Scaled and standard scores are used here to facilitate interpretation for readers, although differences in normative samples are important (see discussion in text). *SD from mean* = the number of *SDs* between Child A's score and the mean, on the basis of test norms (for standardized tests) or comparison to children within 6 months' chronological age in the SALT Bilingual Unique Story database.

Abbreviations are as in Table 1, with the following additions: CELF WS = Clinical Evaluation of Language Fundamentals (CELF) Word Structure subtest; CELF Core = CELF Core Language composite score.

2010). Child A was attending third grade and would therefore be expected to “acquire and use accurately grade-appropriate conversational, general academic, and domain-specific words and phrases” (English Language Arts—Language Standard 3.6). It is thus critical for clinicians to identify and integrate these grade-appropriate words into treatment.

Grammar appears to be an area of relative weakness for Child A, at least expressively. His percent grammaticality figures (50.0% in English and 55.6% in Spanish) are at or

below the means reported by Bedore et al. (2010) for bilingual kindergarteners (i.e., children 3 years younger than Child A). Low test scores on standardized tests that emphasize grammatical skills (such as Word Structure, Recalling Sentences, and Formulated Sentences) converge with this observation. His Spanish MLUW was lower than those of his peers with typical development (2.24 *SDs* below the mean), reflecting reduced sentence productivity in Spanish. Qualitative analysis of both language samples shows that Child A uses mainly simple sentences (i.e., one-clause statements).

Table 5. Strengths and weaknesses for Child A by domain and language on the basis of language sample analysis.

Domain	English		Spanish	
	Strengths	Weaknesses	Strengths	Weaknesses
Vocabulary	<ul style="list-style-type: none"> Specific labels: <i>moose, hamster, owl, bees</i> 	<ul style="list-style-type: none"> Mental states limited to <i>trying</i> Transitions limited to <i>then</i> 	<ul style="list-style-type: none"> Transition <i>finalmente</i> (finally) 	<ul style="list-style-type: none"> Imprecise—e.g., <i>algo se fue arriba</i> (something went up) No mental states Grammaticality: 55.6% Used mainly simple sentences. Errors with gender agreement—e.g., <i>los rocas</i> for <i>las rocas</i> Errors with number agreement—e.g., <i>los abejas estaba cayendo</i> for <i>los abejas estaban cayendo</i>
Grammar		<ul style="list-style-type: none"> Grammaticality: 50.0% Used mainly simple sentences Tense-marking errors—e.g., <i>then the boy and dog wake up</i> Errors with definite/indefinite articles (the/a)—e.g., <i>And the dog too</i> before introducing a dog to the story. 		
Narrative macrostructure	<ul style="list-style-type: none"> General beginning and ending Time established—e.g., <i>in the night</i> Descriptive element: <i>put on his boots fast</i> 	<ul style="list-style-type: none"> Story was mainly a description of actions Unclear referents 	<ul style="list-style-type: none"> Time established—e.g., <i>en la noche</i> (at night) 	<ul style="list-style-type: none"> No beginning or ending No setting Story was mainly a description of actions Unclear referents

Note. Analysis is based on narratives collected in the child's two languages, including microstructures (see also Table 4) and macrostructures.

Common Core standards further indicate that Child A would be expected to “speak in complete sentences ... to provide requested detail or clarification” (English Language Arts—Speaking and Listening Standard 3.6) and to demonstrate command of a variety of English grammatical conventions in writing and speaking (English Language Arts—Language Standard 3.1). These standards can contribute to the development of appropriate goals within this area of weakness for Child A. For example, one treatment goal can focus on expanding sentence structure. This can be targeted using a bilingual approach, given the similarities across languages in complex sentence structure, such as embedded clauses. Advanced conjunctions (e.g., *before*, *while*) can be taught in Spanish and in English, and complex sentences can be practiced in home and school settings (for discussion, see Kohnert, 2013). Child A also demonstrates language-specific grammatical errors including the omission of verb tense in English and errors with number agreement in Spanish (see Table 5 for examples). Because these grammatical features are not shared across languages, a cross-linguistic treatment approach may be most appropriate for these targets (i.e., separate grammatical targets for each language).

As a final matter, discourse skills were not captured in the standardized tests used here. Instead, macrostructural analysis of Child A’s language samples can provide some information about discourse-level skills. In comparison to Child A’s vocabulary and grammatical skills, narrative macrostructure appears to be a relative strength. As shown in Table 5, he included some key narrative elements in both languages, such as marking time. There were more narrative elements in his English language sample (such as a beginning and ending and some key main actions) than in Spanish. The use of more narrative elements in English could reflect his history of school instruction in English, though such a conclusion should be verified with other assessment components, such as teacher interview.

In sum, the language sample analyses converge with test scores in several areas and provide important qualitative information for planning treatment. Consistencies across Spanish and English were identified in the language samples, such as Child A’s inclusion of narrative elements and his reliance on simple sentences. These observations can feed directly into the planning of treatment that uses bilingual and cross-linguistic approaches (Kohnert, 2013; Kohnert & Derr, 2012). Areas that can be treated bilingually include vocabulary (i.e., cognates) and macrostructural aspects of narratives (Kelley & Kohnert, 2012; Squires et al., 2014). Language-specific grammatical errors can be identified from the language samples and are best targeted cross-linguistically.

Conclusions

This article has illustrated areas of convergence between language samples and standardized tests in a sample of school-age bilingual children with PLI. Although the correlation analyses here are limited by a relatively small sample size and by the restricted range of participant abilities,

our results are very similar to previous studies on the convergence of standardized tests and language samples within other populations (e.g., Bedore et al. 2010; Bishop & Donlan, 2005; Ebert & Scott, 2014). Of course, results among these studies are not identical—for example, Ebert and Scott (2014) found a strong relation between English MLU and a standardized test of receptive vocabulary, whereas we did not. More work will be needed to tease apart the variables that influence relations between tests and language samples, given the complexity of the construct of interest (language) and the populations under study (children with differing levels of language ability and exposure). The information provided here can ultimately be used to guide clinicians in conducting thorough and accurate assessments for Spanish–English bilingual children.

Acknowledgments

Data collection for this project was supported by National Institute on Deafness and Other Communication Disorders Grant R21DC010868 (awarded to Kathryn Kohnert), and manuscript preparation was supported by National Institute on Deafness and Other Communication Disorders Grant R03DC013760 (awarded to Kerry Danahy Ebert). Portions of the analysis were presented in an unpublished master’s thesis by Angela Mammolito. We thank Kathryn Ficho and Andrea Morales for assistance with language-sample analysis, as well as the many student research assistants who assisted with data collection for this project. The contributions of Jill Rentmeester Disher and the Minneapolis Public School District were invaluable to this project. We wish finally to thank the participants and their families.

References

- Bedore, L. M., Peña, E. D., Gillam, R. B., & Ho, T.-H. (2010). Language sample measures and language ability in Spanish-English bilingual kindergarteners. *Journal of Communication Disorders, 43*, 498–510.
- Bishop, D., & Donlan, C. (2005). The role of syntax in encoding and recall of pictorial narratives: Evidence from specific language impairment. *British Journal of Developmental Psychology, 23*, 25–46.
- Brownell, R. (2000a). *Expressive One-Word Picture Vocabulary Test*. Novato, CA: Academic Therapy Publications.
- Brownell, R. (2000b). *Receptive One-Word Picture Vocabulary Test*. Novato, CA: Academic Therapy Publications.
- Brownell, R. (2001a). *Expressive One-Word Picture Vocabulary Test—Spanish-Bilingual Edition*. Novato, CA: Academic Therapy Publications.
- Brownell, R. (2001b). *Receptive One-Word Picture Vocabulary Test—Spanish-Bilingual Edition*. Novato, CA: Academic Publications.
- Caesar, L. G., & Kohler, P. D. (2007). The state of school-based bilingual assessment: Actual practice versus recommended guidelines. *Language, Speech, and Hearing Services in Schools, 38*, 190–200.
- Cheng, L.-R. L. (1997). Diversity: Challenges and implications for assessment. *Journal of Children’s Communication Development, 19*, 55–62.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

- Condooris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology, 12*, 349–358.
- De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language & Communication Disorders, 46*, 613–627.
- Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research, 54*, 1077–1088.
- Ebert, K. D., & Kohnert, K. (2016). Language learning impairment in sequential bilingual children. *Language Teaching, 49*, 301–338.
- Ebert, K. D., Kohnert, K., Pham, G., Rentmeester Disher, J. R., & Payesteh, B. (2014). Three treatments for bilingual children with primary language impairment: Examining cross-linguistic and cross-domain effects. *Journal of Speech, Language, and Hearing Research, 57*, 172–186.
- Ebert, K. D., & Mikolajczyk, E. V. (2016). Narrative quality measures in school-age children referred for language assessment. *International Journal of Speech-Language Pathology, 18*, 354–363.
- Ebert, K. D., Pham, G., & Kohnert, K. (2014). Lexical profiles of bilingual children with primary language impairment. *Bilingualism: Language and Cognition, 17*, 766–783.
- Ebert, K. D., & Scott, C. M. (2014). Relationships between narrative language samples and norm-referenced test scores in language assessments of school-age children. *Language, Speech, and Hearing Services in Schools, 45*, 337–350.
- 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*, 1813–1823.
- Greenslade, K. J., Plante, E., & Vance, R. (2009). The diagnostic accuracy and construct validity of the Structured Photographic Expressive Language Test—Preschool: Second Edition. *Language, Speech, and Hearing Services in Schools, 40*, 150–160.
- Heilmann, J., Miller, J. F., Iglesias, A., Fabiano-Smith, L., Nockerts, A., & Andriacchi, K. D. (2008). Narrative transcription accuracy and reliability in two languages. *Topics in Language Disorders, 28*, 178–188.
- Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the Narrative Scoring Scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology, 19*, 154–166.
- Heilmann, J. J., Rojas, R., Iglesias, A., & Miller, J. F. (2016). Clinical impact of wordless picture storybooks on bilingual narrative language production: A comparison of the “Frog” stories. *International Journal of Language & Communication Disorders, 51*, 339–345.
- Kelley, A., & Kohnert, K. (2012). Is there a cognate advantage for typically developing Spanish-speaking English-language learners? *Language, Speech, and Hearing Services in Schools, 43*, 191–204.
- Kohnert, K. (2013). *Language disorders in bilingual children and adults* (2nd ed.). San Diego, CA: Plural.
- Kohnert, K., & Derr, A. (2012). Language intervention with bilingual children. In B. A. Goldstein (Ed.), *Bilingual language development & disorders in Spanish-English speakers* (2nd ed., pp. 337–356). Baltimore, MD: Brookes.
- Kohnert, K., Yim, D., Nett, K., Kan, P. F., & Duran, L. (2005). Intervention with linguistically diverse preschool children: A focus on developing home language(s). *Language, Speech, and Hearing Services in Schools, 36*, 251–263.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.
- Langdon, H. W., & Saenz, T. I. (2015). *Working with interpreters and translators: A guide for speech-language pathologists and audiologists*. San Diego, CA: Plural.
- Mayer, M. (1969). *Frog, where are you?* New York, NY: Dial Press.
- Miller, J. F., & Iglesias, A. (2012). *SALT: Systematic Analysis of Language Transcripts* (Research version) [Computer software]. Middleton, WI: SALT Software.
- National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010). *Common Core State Standards*. Washington, DC: Authors.
- Nippold, M. A., Frantz-Kaspar, M. W., Cramond, P. M., Kirk, C., Hayward-Mayhew, C., & MacKinnon, M. (2015). Critical thinking about fables: Examining language production and comprehension in adolescents. *Journal of Speech, Language, and Hearing Research, 58*, 325–335.
- Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2008). Expository discourse in adolescents with language impairments: Examining syntactic development. *American Journal of Speech-Language Pathology, 17*, 356–366.
- Paul, R., & Norbury, C. F. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (4th ed.). St. Louis, MO: Elsevier Mosby.
- Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics, 28*, 399–410.
- Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2014). *BESA: Bilingual English Spanish Assessment manual*. San Rafael, CA: AR-Clinical Publications.
- Restrepo, M. A., & Kruth, K. (2000). Grammatical characteristics of a Spanish-English bilingual child with specific language impairment. *Communication Disorders Quarterly, 21*, 66–76.
- Rojas, R., & Iglesias, A. (2009, March). Making a case for language sampling: Assessment and intervention with (Spanish-English) second language learners. *The ASHA Leader, 14*(3), 10–13.
- Rojas, R., & Iglesias, A. (2013). The language growth of Spanish-speaking English language learners. *Child Development, 84*, 630–646.
- Semel, E., Wiig, E. H., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals—Fourth Edition*. San Antonio, TX: The Psychological Corporation.
- Simon-Cerejido, G., & Gutiérrez-Clellen, V. F. (2007). Spontaneous language markers of Spanish language impairment. *Applied Psycholinguistics, 28*, 317–339.
- Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gillam, R. B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language & Communication Disorders, 49*, 60–74.
- Wiig, E. H., Semel, E., & Secord, W. A. (2006). *Clinical Evaluation of Language Fundamentals—Fourth Edition, Spanish*. San Antonio, TX: The Psychological Corporation.
- Williams, C. J., & McLeod, S. (2012). Speech-language pathologists’ assessment and intervention practices with multilingual children. *International Journal of Speech-Language Pathology, 14*, 292–305.

Appendix

Raw Language Transcripts for Child A

English

There was a boy that was seeing the frog. (And the) and the dog too. (And the) and the boy went to sleep. And dog. Froggie ran away outside in the night. Then the boy and the dog wake up. And then he see the jar. The frog was gone. He put on his boots fast. And then going out. (They) boy yelled. And the dog was gonna to fall. And the dog fall. Then the boy fall. But the dog lick him. And his keep in yelling. Then the bees there X the dog was look/ing at. (Then the) then the dog was trying to follow the bees. Then the boy was trying to see a little hole. It was like a hamster or rat. Then he find on the tree (a hole) a hole. (And the) and the dog ran trying to play with bees. But the bees fall. It was an owl. And then the bees trying to chase the dog. And then the boy fell. Then the owl flew away. Then he went up to the rock and yeah. Then there was (like a like) a moose or something (that) that was. The moose was (running) running. And the dog was running too. Then the boy and the dog fall in the pool. Then they fell in a water. Then he heard (like) a frog shouting. Then the dog was be quiet. (And) and he check on the back. Then he found his frog (and the) and the frog lady. And he find (the) the babies frog. Then the boy take one home. He can. Then he say bye to the frogs. The end.

Spanish

El niño está (vie) viendo (su) su (rana) rana en la noche. Luego se durmió. Y su rana se fue. Luego el niño se levantó. Luego su rana no está. Y el perro estaba esperando. El niño estaba buscando. El perro se cayó. Luego el niño se cayó. Luego el perro (lo lo hi) lo hizo. El niño (estaba) estaba lo buscando. Y estaba buscando los árboles (y los uh los árboles) y donde están los rocas. Él buscó un chiquito casa. Pero ése era como un ratón. El perro miró los abejas. Se cayó los abejas. El niño buscó X así otra cosa (donde está los) donde está el árbol. Luego se cayó. Luego los abejas estaba cayendo. El niño (estaba) estaba corriendo. Luego (él) él buscó un grande roca. Luego se fue arriba. Y puede ver poquito. Luego algo (se) se fue arriba. Luego corrió. Pero el niño y el perro se cayó en el agua. Luego el niño se cayó. Splash. (Luego eh) finalmente buscó otra. (Y) Y escuchó como un rana. Luego el perro estaba callado. Luego el niño dice shh. Luego miró para atrás. Luego buscó su rana. Y otro rana mujer. Luego tenía bebés de rana. El niño llevó uno. Y dijo bye.
