Top Lang Disorders
Vol. 40, No. 2, pp. 182–201

Language Sample Analysis With Bilingual Children

Translating Research to Practice

Kerry Danahy Ebert

Language sample analysis (LSA) has been called the "gold standard" for clinical language assessment with children learning more than one language. The research literature discussing this clinical tool with bilingual children has grown substantially in recent years. This article reviews and synthesizes the literature on LSA in order to provide guidance for clinicians seeking to utilize this tool with bilingual children. The focus is on oral narrative language samples, reflecting the currently available literature. The article reviews procedural considerations in eliciting and coding narratives with bilingual children and considers the evidence of effectiveness for different assessment purposes such as the identification of language disorders and the documentation of dual-language growth over time. Research findings are translated to clinical scenarios. Finally, gaps in the literature are identified. **Key words:** *assessment*, *bilingual*, *children*, *language disorder*, *language sample analysis*, *narrative*

LANGUAGE SAMPLE ANALYSIS (LSA) is one of the most valuable resources in the language clinician's toolkit. It is a flexible tool that provides rich information about the language skills of the child undergoing assessment. Given current demographic patterns, the chances that the child undergoing language assessment is bilingual continue to grow both within and outside the United States. For example, more than half of the population of Europe can hold a conversation in two or more languages (European Commission, 2012), and 21% of the United States' population older than 5 years speaks a language other than English at home (U.S. Census Bureau, 2015). Fortunately, a substantial literature discussing and utilizing LSA with bilingual children exists to guide the clinician working with this population. The purpose of this review is to examine and synthesize this literature.

To date, the literature on LSA with bilingual children is overwhelmingly focused on oral narratives. There has been minimal study of other genres (cf. Ooi & Wong, 2012; Spoelman & Bol, 2012) or modalities (cf. Hsin & Snow, 2017), and as a result the literature review and clinical recommendations in this article consider oral narratives exclusively. Gaps in the literature (such as those related to genre and modality) are discussed at the conclusion of the article.

This review begins with a foundational section establishing concepts and definitions important to the study of bilingualism. The bulk of the review is devoted to answering *why*, *how*, and *what* in relation to LSA with bilingual individuals. That is, I discuss the advantages of LSA for bilingual children (*why* a clinician should use it with this population); describe procedural considerations in collecting, coding, and analyzing language samples with bilingual children (*how* to use

Author Affiliation: *Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis.*

The author has indicated that she has no financial and no nonfinancial relationships to disclose.

Corresponding Author: *Kerry Danahy Ebert, PhD, CCC-SLP, Department of Speech-Language-Hearing Sciences, University of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr SE, Minneapolis, MN 55455 (kebert@umn.edu).*

DOI: 10.1097/TLD.0000000000000209

LSA with this group); and synthesize the literature on using LSA to identify language disorders and to characterize language skills in bilingual children (*what* can be done with LSA). The final section presents two hypothetical case studies in order to illustrate specific clinical actions that could be taken on the basis of the current literature.

BILINGUALISM: CONDITIONS, TERMS, AND VARIABILITY

Within this article, I use the term *bilingual* to mean an individual with significant input in two or more languages during childhood (Kohnert, 2010), and I include studies that meet this definition even if another term (e.g., dual-language learner) was used in the original study. This definition of bilingualism is intentionally broad and usage-based rather than proficiency-based, as is appropriate when working with children with language disorders (Kohnert, 2010). However, it does mean that a diverse group of individuals are considered bilingual and such diversity can complicate clinical language assessment. Too often, assessment of bilingual individuals results in "apples to oranges" comparisons, as the subject's prior language learning experiences or the methods of language assessment do not truly match the comparative data. To characterize variability and reduce inappropriate comparisons, it is important to define some of the key parameters that define bilingual language learning contexts.

One of the most important parameters of bilingual learning contexts is age of acquisition. Historically, there has been a distinction made between *simultaneous* and *sequential* bilingualism. Simultaneous bilingual individuals receive exposure to two languages virtually from birth, but often this is functionally defined as prior to 2 years of age. Sequential bilinguals learn only one language from birth (the L1) and acquire a second (the L2) sometime later in childhood. Of course, the age of exposure to a second language can vary continuously, as can the current and cumulative levels of input and output proportions in each language. Thus, age of initial acquisition, current input and output levels, and historical input and output levels are all relevant factors in shaping an individual's current proficiency in each language.

An individual's relative proficiency in the L1 versus the L2 is often described using the term *dominance*, and this is another key parameter in the study of bilingual language assessment. Individuals with relatively equivalent proficiency in the L1 and the L2 are described as *balanced*, whereas those with greater relative proficiency in one language are described as dominant in that language. However, bilingual individuals (like monolingual individuals) develop proficiency via language use, and as a consequence dominance is related to specific contexts of use (Kohnert, 2010). In other words, a bilingual could be dominant in one language for some contexts and topics but dominant in another for other contexts and topics.

Finally, the societal language environment shapes the language skills of bilingual individuals. In some sociolinguistic contexts, learning a second language tends to occur while the first language is maintained or continues to grow; these environments have been labeled *additive*. In other contexts—termed *subtractive*—learning a second language tends to occur at the expense of development of the L1. For example, in most areas of the United States, educational (and ultimately vocational and social) opportunities overwhelmingly favor the L2, English. Loss of L1 skills is documented among children raised in the United States (e.g., Guiberson, Barrett, Jancosek, & Itano, 2001; Uccelli & Paéz, 2007), and children often shift to dominance in the L2 sometime during the elementary years (e.g., Lucero, 2018; Pham, 2016; also see Kohnert, 2010). The context may be similar for children of immigrants to other countries, such as Turkish–Dutch bilinguals in the Netherlands or Arabic–Swedish bilinguals in Sweden, but there also may be key differences in educational or social structures that influence bilingual development.

Research on language sampling with bilingual children has spanned a variety of these contexts. Investigations have been conducted with different languages, in different educational and social environments, and with children exposed to L2 at varied ages of acquisition. All of these variables are relevant in determining whether research results will generalize to a given group of bilingual children, and it is important for clinicians seeking guidance on LSA with bilingual children to compare the populations they serve with the participants in the literature in terms of these variables. It is therefore important for research studies to provide adequate description of bilingual participants. It also is important for clinicians to try to characterize the linguistic experience of their bilingual clients when conducting assessment. This means gathering information on age of first exposure to the L2, language usage patterns within the home and school environments, and the child's own preferences (i.e., L1 vs. L2) in different contexts. Information on a child's linguistic experiences can be collected via parent questionnaire; at least one such questionnaire is currently freely available to clinicians (the Alberta Language Environment Questionnaire; Paradis, 2010; also see Gatt, O'Toole, & Haman, 2015, for discussion of parent questionnaires on bilingual development).

ADVANTAGES OF LSA WITH BILINGUAL CHILDREN

Language sample analysis has been called the gold standard of language assessment for bilingual children (Heilmann, Rojas, Iglesias, & Miller, 2016). As with other groups of children, LSA offers a flexible, ecologically valid, and potentially efficient tool to use with developing bilingual individuals (Ebert & Pham, 2017; Fiestas & Peña, 2004; Kapantzoglou, Fergadiotis, & Restrepo, 2017; Rojas & Iglesias, 2009). Several of these advantages are particularly important for this population. Flexibility and adaptability are critical for a group that speaks a wide variety of languages and whose varied learning experiences may not prepare them for the rigid context of a standardized test. Ecological validity is particularly important because other prominent assessment methods are frequently *not* ecologically valid for this group (e.g., De Lamo White & Jin, 2011). Efficiency—which here refers to the number of different aspects of language that can be assessed in a single assessment task (e.g., Ebert & Scott, 2014; Heilmann et al., 2008)—is highlighted when there are two languages to be assessed and limited time in which to do so. Finally, LSA may highlight different language skills than other language assessments both for monolingual speakers (Ebert & Scott, 2014) and bilingual speakers (Ebert & Pham, 2017), providing a complementary perspective in assessment.

There also are advantages of LSA that are unique to bilingual children. The first is the relative ease of collecting information on two different languages. There is a clear consensus in the literature that clinical language assessment of a bilingual child should collect information about both languages (Bedore & Peña, 2008; De Lamo White & Jin, 2011; Ebert & Kohnert, 2016; Ebert & Pham, 2017). However, the difficulties clinicians face in assessing minority home languages are also well documented (e.g., Arias & Friberg, 2017; De Lamo White & Jin, 2011). Language sample analysis may offer one way to collect and analyze information about a language the clinician does not speak if the clinician is able to access an interpreter or other resources (see Langdon & Saenz, 2016, for detailed information on the integration of interpreters into assessment).

Another possible advantage of LSA for bilingual children is the potential for cross-language comparisons. Although clinicians should exercise caution in making global judgments of dominance based on direct comparisons of LSA measures (e.g., determining that a child universally prefers English based on a longer English MLU [Mean Length of Utterance]), collecting language samples in two languages can contribute to characterizing a child's relative strengths and weaknesses

in both the L1 and the L2. This possibility is described in further detail later.

PROCEDURAL CONSIDERATIONS IN LSA WITH BILINGUAL CHILDREN

Clinicians seeking to utilize LSA are faced with a series of decisions in how to collect and analyze the samples. The unique characteristics of bilingual children influence many of these decisions. This section is devoted to procedural decision-making in the use of LSA for clinical purposes, including choices in the elicitation of a narrative language sample, management of code-switching, and selecting appropriate measures to extract from the sample.

Narrative collection and coding

First, there are several materials and procedures that are currently supported in the literature for collecting oral narratives from bilingual children. Within the United States, the majority of investigations have utilized the *Frog* series by Mercer Mayer (e.g., Bedore, Peña, Gillam, & Ho, 2010; Gutiérrez-Clellen, Simon-Cereijido, & Wagner, 2008; Heilmann et al., 2016; Pham, 2016). These books present a series of pictures that support the telling (generation) or retelling of a narrative. For clinicians working with Spanish–English bilingual populations in the United States, one advantage of the *Frog* series is the existence of support via the Systematic Analysis of Language Transcripts (SALT) software (Miller & Iglesias, 2012). Clinicians can access sample scripts in both Spanish and English to facilitate elicitation of several *Frog* stories (Miller, Andriacchi, & Nockerts, 2015). Comparison databases also are available for Spanish–English bilingual children living in Texas and California, aged 5;0–9;9 and completing a story retell task ($N = 4{,}667$) and aged 5;0–9;7 and completing a story generation task ($N = 475$). In addition, a number of studies to be discussed in this review utilize *Frog* stories with bilingual children, providing clinicians with additional data for comparison.

Other tools exist for narrative collection and analysis with bilingual children. A European initiative resulted in a tool specifically for this purpose, the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012). The MAIN is a freely available tool designed for children aged 3–10 years; per its developers, it has been tested with more than 500 children and 15 different languages. The MAIN utilizes six-picture illustrated sequences to support generation or retelling of four different stories, and the tool includes comprehension questions and a protocol for scoring macrostructure elements. A number of recent studies have utilized the MAIN (e.g., Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016; Kapalková, Polišenská, Marková, & Fenton, 2016; Tsimpli, Peristeri, & Andreou, 2016), providing potential comparison data for clinicians with matching samples.

Finally, the Edmonton Narrative Norms Instrument, or ENNI, also has been utilized in LSA research with bilingual children (Cleave, Girolametto, Chen, & Johnson, 2010; Govindarajan & Paradis, 2019). To date, these investigations have been limited to assessment of the L2 (English) of children born in Canada who speak a variety of L1s; the clinical utility of the tool for bilingual children would be facilitated by investigations in other languages as well.

In summary, there are several narrative elicitation tools that are supported by data from bilingual children. Although it is possible to collect narrative language samples using other materials (e.g., videos) or no formal materials at all (e.g., personal narratives), the structure provided by a consistent set of picture stimuli may facilitate transcription (Heilmann et al., 2008), as well as comparisons with other children.

Even after selecting a means for collecting a language sample, clinicians working with bilingual children face additional choices. It is considered best practice to collect samples in different languages on different days, sticking to one language of assessment during a single session whenever possible (Gagarina et al.,

2012; Rojas et al., 2016; Squires et al., 2014). In addition, children may find it pragmatically unwelcome to repeat the same story more than once and therefore it is better to use distinct stories for each language (Gutiérrez-Clellen, Simon-Cereijido, & Leone, 2009).

Clinicians also may be interested in making some comparisons across languages, which means that they need comparable (but not identical) stories for each language. Fortunately, at least two of the tools described here (i.e., MAIN and the *Frog* stories) provide the opportunity to collect parallel but distinct stories in each language. Both tools have multiple stories. Those used with the MAIN were designed to be parallel and have been used for cross-language comparisons in research (Kapalková et al., 2016). Finally, although the *Frog* stories were not originally created to be equivalent to each other, research suggests that several LSA measures (including MLU-w [MLU in words], number of total utterances (NTU), and the Narrative Scoring Scheme) should be relatively comparable across different *Frog* stories in Spanish and in English (Heilmann et al., 2016). It should be noted, however, that this same work indicated that a lexical diversity measure (Number of Different Words, or NDW) did differ significantly across *Frog* stories (Heilmann et al., 2016).

Once language samples are collected, there are additional considerations in the coding and analysis phases. Although it is outside the scope of this work to review coding schemes across languages, it is important to note that LSA coding schemes are language-specific. For example, segmentation rules should be altered for languages that allow subject deletion (Heilmann et al., 2008; Pham, 2016), serial verbs (Pham, 2016; To, Stokes, Cheung, & T'sou, 2010), or conjunction without an explicit connective (To et al., 2010). For bilingual children, characteristics of the L1 may influence the L2 and therefore coding may need to be modified for the L2 as well. For example, when language samples are collected from Spanish–English bilingual children, the segmentation rules that are used for Spanish also should be used in English, even though they differ from English segmentation rules for monolingual speakers (Miller & Iglesias, 2012).

Code-switching

Code-switching is a unique feature of bilingual discourse that affects LSA. Code-switching can be defined as the use of more than one language in discourse (Gutiérrez-Clellen et al., 2009), meaning that the "nontarget" language appears in the language sample. Clinicians seeking to utilize LSA with bilingual children may be concerned about how to manage code-switching. It is important to first note that the available literature indicates that code-switching is not a particularly common occurrence in children's discourse, at least not within the context of language assessment (Ebert & Mammolito, 2015; Greene, Peña, & Bedore, 2013; Gutiérrez-Clellen et al., 2009; Iluz-Cohen & Walters, 2012). For example, Gutiérrez-Clellen et al. (2009) examined oral narrative and conversational language samples of Spanish–English bilingual children with and without language disorder, aged 5 to 6 years, and documented that an average of 6.5% of utterances across samples contained code-switching. They did, however, note that the number of children who code-switched was higher for conversational samples than for narrative samples, suggesting that code-switching may be less frequent in more constrained tasks (Gutiérrez-Clellen et al., 2009). In an analysis of 68 narrative language samples from school-aged Spanish–English bilingual children with a language disorder, Ebert and Mammolito (2015) found that 74% of English samples and 39% of Spanish samples contained no code-switching at all.

For a clinician concerned with assessing the target language, the most common approach is to eliminate utterances with code-switching from the analysis (e.g., Bedore et al., 2010; Gutiérrez-Clellen, Restrepo, & Simon-Cereijido, 2006; Uccelli & Paéz, 2007). It is possible, though, that code-switching may provide clinically relevant information. It has been proposed that children with

language disorders may exhibit code-switching more frequently than typically developing peers. However, the available literature conflicts on this point (Ebert & Mammolito, 2015; Greene et al., 2013; Gutiérrez-Clellen et al., 2009; Iluz-Cohen & Walters, 2012). Moreover, the frequency of code-switching is clearly influenced by contextual variables such as the setting and characteristics of the examiner (Iluz-Cohen & Walters, 2012). As is seen in the following section, there are likely more reliable LSA measures for identifying children with language disorders. Code-switching is likely to be related to a child's language dominance (Ebert & Mammolito, 2015; Gutiérrez-Clellen et al., 2009), and frequent code-switching may be a sign that the child is relatively uncomfortable speaking the language used for assessment. In addition, most speakers follow grammatical rules when code-switching; for example, switching languages between a root verb and a bound morpheme is not acceptable (see Gutiérrez-Clellen et al., 2009, for a list of these rules). Children who violate these grammatical rules when code-switching could lack grammatical knowledge in one language, prompting an atypical switch to the other language. However, more research is needed to provide clear guidance on how code-switching can be used in clinical contexts.

Measures of interest

Finally, clinicians will need to determine the most useful measures to extract from their language samples. In general, the LSA measures used with bilingual children are comparable with those used with monolingual children, and the reader is referred elsewhere for a detailed review of possible measures and their interpretation (for relevant discussions, see Eisenberg, this issue; Guiberson, this issue; Lundine, this issue; also see Bedore et al., 2010; Miller et al., 2015). Research on LSA with bilingual children has spanned both microstructure and macrostructure and includes measures of sentence length, grammaticality, fluency, lexical diversity, and productivity. Table 1 provides a brief summary of the most commonly used LSA measures in the literature on bilingual children, and the sections that follow provide an in-depth discussion of what these measures may show in terms of bilingual language development and disorders.

CAN LSA MEASURES BE USED TO IDENTIFY LANGUAGE DISORDERS?

One of the most important purposes of clinical language assessment is to determine whether or not a language disorder is present. Given the problems that accompany using other assessment methods—most notably, norm-referenced tests—to identify language disorders in bilingual children (De Lamo White & Jin, 2011), the viability of LSA for this purpose is of key interest in the literature. The majority of literature in this area is focused on primary language disorders rather than secondary language disorders (i.e., those stemming from autism spectrum disorder, intellectual disability, or sensory loss). In this review, I follow the recent international consensus recommendation (Bishop, Snowling, Thompson, Greenhalgh, & CATALISE Consortium, 2017) and adopt the term *developmental language disorder* (*DLD*) to refer to primary DLD.

This section is devoted to reviewing the assessment and analysis methods and measures that optimize LSA to identify DLD in bilingual children. I first consider studies that have addressed a related but important question: whether bilingual children look comparable with their monolingual peers on LSA measures. I then proceed to review studies that have compared bilingual children with language disorders with their typically developing bilingual peers, with special consideration of studies that have extended analyses to measures of diagnostic accuracy such as sensitivity and specificity.

Monolinguals versus bilinguals

Studies comparing bilingual children with monolingual children on LSA measures

Table 1. Summary of common language sample analysis measures in studies of bilingual children

Domain	Area	Name or Abbreviation (s)	Description	Example Studies Using Measure With Bilinguals
Microstructure	Sentence length and complexity	MLU-w	Mean Length of Utterance in words	Altman et al., 2016; Rezzonico et al., 2015; Verhoeven et al., 2011
		SI, clausal density	Subordination Index, calculated as the average number of clauses per utterance	Cleave et al., 2010; Lucero, 2018; Tsimpli et al., 2016
	Lexical diversity	NDW, TDW	Number of Different Words in the language sample	Jacobson & Walden, 2013; Pham, 2016
		D, VocD	Statistic representing average type–token ratio; calculated by a computer program using random sampling of increasingly longer sections of the language sample	Jacobson & Walden, 2013; Kapantzoglou et al., 2017
	Productivity	NTU, TNU	Number of Total Utterances in the language sample	Verhoeven et al., 2011
		TNW, NTW	Total Number of Words in the language sample	Heilmann et al., 2008; Uccelli & Paéz, 2007
	Fluency	WPM	Words produced Per Minute	Rojas & Iglesias, 2013
	Grammaticality	% Gram	Percentage of utterances in the sample that are grammatical	Ebert & Pham, 2017; Simon-Cereijido & Gutiérrez-Clellen, 2007
		NEU	Average Number of grammatical Errors per Utterance	Restrepo, 1998
		Errors	Counts of specific grammatical errors, such as gender or number errors, omission of bound morphemes, substitution or omission of prepositions; percentage accuracy on specific forms such as finite verbs	Gutiérrez-Clellen et al., 2008; Iluz-Cohen & Walters, 2012
Macrostructure	Story grammar	SG Count	Number of story grammar elements included, such as goals, attempts, outcomes	Altman et al., 2016
		Rating scales	Specific story grammar rating scales are included on the MAIN and ENNI	Boerma et al., 2016; Govindarajan & Paradis, 2019
	Internal states	IS count; mental state verbs	Count of different words related to characters' internal states or mental states	Boerma et al., 2016; Tsimpli et al., 2016
	Mixed scale	NSS	Narrative Scoring Scheme; rating scale for seven macrostructural elements including both story grammar and mental states	Heilmann et al., 2016; Lucero, 2018

Note. ENNI = Edmonton Narrative Norms Instrument; MAIN = Multilingual Assessment Instrument for Narratives.

address the question of whether reduced exposure to a specific language creates a disadvantage on this assessment tool. This question can be seen as a prerequisite to considering the diagnostic accuracy of LSA, because it determines how closely matched a comparison sample must be to appropriately evaluate a specific bilingual child. If bilingual children score below monolingual peers on LSA measures, then it is critical that the language exposure of the child undergoing assessment closely matches the linguistic experiences of the comparison group. Studies that address this question have spanned an array of languages, though they have universally considered only the L2 of the bilingual children (Boerma et al., 2016; Bonifacci, Barbieri, Tomassini, & Roch, 2018; Cleave et al., 2010; Gutiérrez-Clellen et al., 2008; Rezzonico et al., 2015; Tsimpli et al., 2016; Verhoeven, Steenge, & van Balkom, 2011). There is a clear consensus in this literature that bilingual children do not score below their monolingual peers on macrostructural measures, such as story grammar scores (Boerma et al., 2016; Cleave et al., 2010; Rezzonico et al., 2015) or references to internal states (Tsimpli et al., 2016). The statistical equivalence of bilingual and monolingual children on macrostructural measures has been demonstrated within groups of typically developing children and within groups of children with DLD (Cleave et al., 2010). These studies have spanned a wide age range, including 4-, 5-, and 6-year-old children as well as 9-year-olds. They also have spanned L2s including English, Greek, and Dutch, although the dominance and linguistic experiences of participants have not been well reported. Past language experience does influence storytelling styles (Fiestas & Peña, 2004). However, it appears that macrostructural variables can be scored within these different storytelling styles, reducing the bias of macrostructural variables when evaluating developing bilingual children.

The picture is far less clear for microstructural measures. Although some studies (e.g., Cleave et al., 2010; Spoelman & Bol, 2012) have found that bilingual children perform comparably with monolingual peers on microstructural LSA measures, others have found lower scores on at least some microstructural features for bilingual groups (Rezzonico et al., 2015; Tsimpli et al., 2016; Verhoeven et al., 2011). For example, Rezzonico et al. (2015) administered a story retelling task in English to forty 4-year-old children distributed across four groups (monolingual children with and without DLD; bilingual children with and without DLD). Measures of sentence length and lexical diversity did not distinguish between the monolingual and bilingual groups, but bilingual groups scored lower on a grammaticality measure (the percentage of verbs produced in the correct grammatical form). Notably, the bilingual children in this study heard and spoke the societal language (English) more than the minority language even within the home setting; despite relatively low exposure to the minority home language, the bilingual children were not comparable with the monolinguals on the grammaticality measure.

Other studies with different bilingual populations support this finding. Gutiérrez-Clellen et al. (2008) also found lower scores for 4- to 6-year-old Spanish–English bilinguals on a grammaticality measure, the use of finite verb marking. The children in this study had relatively balanced proficiency in their two languages. Finally, Verhoeven et al. (2011) found that 7- to 9- year-old bilingual children had shorter utterances, lower percentages of grammatical utterances, and higher rates of a specific grammatical error (omission of an agreement marker) in narratives told in their L2, Dutch.

One plausible hypothesis is that the difference across studies is due to differing exposure to the L2 for the bilingual participants. That is, perhaps studies that have found no difference between monolinguals and bilinguals on microstructural LSA measures have included bilinguals with higher levels of L2 exposure (who therefore look more like children who only speak the L2). However, this is not the case. Rezzonico et al. (2015) and Cleave et al. (2010) studied very similar

populations: 4-year-old children in Canada with varied L1s who were relatively dominant in the L2, English (in both studies, children's output in the L2 at home was around 10%). Yet, these two studies came to opposite conclusions. The group studied by Gutierrez-Clellen et al. (2008) was more balanced, with approximately 50% input in each language at both home and school; this study reached a similar conclusion to Rezzonico et al. (2015).

Thus, although Cleave et al. (2010) conclude that, "language sample measures, such as narratives, may be a less biased way to assess the language skills of dual language learners" (p. 519), it appears premature to conclude that bilingual children can be compared with monolingual peers on microstructural language sample measures. However, LSA could still be a valuable tool for the identification of language disorders when children are compared with peers with similar language experiences.

Bilinguals with typical language development versus bilinguals with language disorder

A substantial body of literature has done just this: compared groups of bilingual children with language disorders (typically, DLD) with their typically developing bilingual peers to determine whether significant group differences exist. This is an important question, but it does not fully determine whether LSA measures can distinguish between typical development and language disorder in bilingual populations. It is necessary to conduct discriminant accuracy analyses that examine the ability to classify each individual case into the correct category (i.e., typical development or language disorder). Ideally, studies should report group sensitivity and specificity as well as individual likelihood ratios (Dollaghan, 2007) before making conclusions about a procedure's ability to identify a disorder. However, relatively few studies of LSA in bilingual children have conducted these types of analyses.

The following sections review literature comparing bilinguals with DLD with typically developing bilingual peers on various LSA measures, with special consideration given to studies that have considered classification accuracy. Because of the quantity of available literature in this category, the review is further subdivided into sections on microstructure and macrostructure.

Microstructure

At the microstructural level, investigations of the ability of LSA to identify language disorders within bilingual populations have considered a range of measures, including sentence length, lexical diversity, productivity, and grammaticality. Perhaps the most common dependent variable is sentence length as measured by MLU or a variant (such as average MLU of the longest three to five utterances). Results from such investigations have varied, ranging from studies that find no group differences between children with and without DLD (Govindarajan & Paradis, 2019; Verhoeven et al., 2011), those that do find significant differences between groups in sentence length (Altman, Armon-Lotem, Fichman, & Walters, 2016; Iluz-Cohen & Walters, 2012; Kapantzoglou et al., 2017; Rezzonico et al., 2015), and those that have found not only a group-level difference but also that MLU contributes to diagnostic models detecting DLD (Lazewnik et al., 2019; Ooi & Wong, 2012; Restrepo, 1998). For example, Lazewnik et al. (2019) recently examined a host of different assessment tools in a group of 30 Spanish–English bilingual children with and without DLD, aged 4;0–5;11. Narrative retells were collected in both languages, and MLU-w was extracted only from the better language. This variable alone had adequate sensitivity (85.7%) but poor specificity (57.1%). However, it could be combined with a norm-referenced test score to create a model that accurately classified participants with 100% sensitivity and 92.9% specificity (Lazewnik et al., 2019). Although such results indicate that sentence length may identify bilingual children with DLD under some conditions, a more common finding is that MLU is not the best LSA measure for

identifying DLD in bilingual children; that is, MLU may distinguish between groups, but other LSA measures are more powerful predictors (Altman et al., 2016; Govindarajan & Paradis, 2019; Iluz-Cohen & Walters, 2012; Kapantzoglou et al., 2017; Simon-Cereijido & Gutiérrez-Clellen, 2007). With the exception of one study (Simon-Cereijido & Gutiérrez-Clellen, 2007), all of these studies included children in the 5- to 7-year-old age range. Both Lazewnik et al. (2019) and Ooi and Wong (2012), who found that MLU helped identify bilinguals with DLD, included slightly younger children; it is possible that MLU is most useful for identifying DLD in bilinguals younger than 5 years.

Measures of lexical diversity also have yielded mixed results. The majority of studies that have considered lexical diversity have utilized NDW without a correction for length of sample (cf. Ebert & Scott, 2014; also see Jacobson & Walden, 2013, for a discussion of lexical diversity measures). Several studies have found group differences on this measure in languages such as Hebrew (Altman et al., 2016; Iluz-Cohen & Walters, 2012), English (Rezzonico et al., 2015), and Greek (Tsimpli et al., 2016). However, another study (Jacobson & Walden, 2013) found no group differences between Spanish–English bilinguals with and without DLD on NDW. In addition, no studies have found that NDW contributes to accurate classification.

Two recent studies have considered D as a measure of lexical diversity in Spanish–English bilingual samples from the United States (Jacobson & Walden, 2013; Kapantzoglou et al., 2017), and an additional study examined D in Chinese–English bilingual children in Malaysia (Ooi & Wong, 2012). Results from these studies conflict; Jacobson and Walden (2013) as well as Ooi and Wong (2012) found no group differences on D, whereas Kapantzoglou et al. (2017) found not only group differences but also a significant role for D in a classification model that demonstrated 90% sensitivity and 85% specificity. One notable difference between these studies is that Kapantzoglou et al. (2017) collected language samples only in the L1, whereas Ooi and Wong (2012) considered only the L2. Jacobson and Walden (2013) considered both the L1 and the L2 (and found no group differences in either language); however, their participants were older (at 7–10 years of age) than those included by Kapantzoglou et al. (2017). Overall, lexical diversity measures may be particularly sensitive to such methodological differences and thus do not appear to be completely reliable for distinguishing bilingual children with and without DLD.

Measures of overall productivity, such as Number of Total Words (NTW) and NTU, are unlikely to be the best choice for differentiating bilingual children with DLD from their peers. A few studies have conducted group-level comparisons using measures of overall productivity, with mixed results (cf. Altman et al., 2016; Iluz-Cohen & Walters, 2012; Verhoeven et al., 2011). However, no studies have concluded this is the most promising measure for identification or included it in classification models.

Measures of overall grammaticality and specific grammatical errors appear to be the strongest candidates for reliably differentiating between children with and without DLD in groups of bilinguals. Restrepo (1998) was one of the first to examine grammatical errors in language samples as a means of distinguishing bilingual children with DLD from their typical peers. In a group of 5- to 7-year-old Spanish–English bilingual children, the number of grammatical errors per T-unit from a language sample collected in the L1 (Spanish) demonstrated perfect specificity and 70% sensitivity in identifying children with DLD. In the years since Restrepo's (1998) work, measures of grammaticality and errors have shown the largest differences between bilingual children with and without DLD in multiple studies examining LSA. This conclusion has held true across investigations of the L1-only (Kapantzoglou et al., 2017; Simon-Cereijido & Gutiérrez-Clellen, 2007), the L2-only (Gutiérrez-Clellen et al., 2008;

Rezzonico et al., 2015; Verhoeven et al., 2011), and both languages combined (Altman et al., 2016; Iluz-Cohen & Walters, 2012; Jacobson & Walden, 2013).

In addition to the grammatical errors per T-unit variable explored by Restrepo (1998), several additional grammatical measures have been shown to identify DLD in bilingual children. Language-specific grammatical errors, such as gender errors in Hebrew (Altman et al., 2016), omission of agreement markers in Dutch (Verhoeven et al., 2011), and finite verb errors in English (Gutiérrez-Clellen et al., 2008), have shown promise in differentiating children with DLD at the group level; however, it must be noted that these markers may be language-specific. For example, the omission of past tense markers—a measure that shows promise in English—did not differentiate children with DLD from typical bilinguals in Dutch (Verhoeven et al., 2011). There may be more "universal" grammatical measures that differentiate bilingual children with DLD, such as an overall measure of verb accuracy (Rezzonico et al., 2015) or omission errors (Jacobson & Walden, 2013). The overall percentage of ungrammatical utterances in a sample has shown particular promise: it differentiated groups of bilingual children with and without DLD in Dutch (Verhoeven et al., 2011) and also contributed to accurate classification of individual children in two studies of Spanish–English bilinguals (Kapantzoglou et al., 2017; Simon-Cereijido & Gutiérrez-Clellen, 2007). It also may be relatively straightforward for a clinician to code reliably (e.g., Ebert & Pham, 2017) and less dependent on elicitation technique than other microstructural measures (Kapantzoglou et al., 2017). Nonetheless, additional work will be helpful in establishing the conditions (e.g., ages, languages, and elicitation conditions) under which this measure reliably identifies DLD in bilinguals.

In summary, microstructural measures and their ability to identify DLD in bilingual samples have been of significant interest in the literature. Sentence length and lexical diversity have inconsistently differentiated bilingual children with DLD from their peers, whereas grammatical measures appear to consistently identify DLD. However, the previous section comparing monolinguals with bilinguals with typical language injects a note of caution into these conclusions: specific grammatical measures also appear to be the most sensitive to language experience (Bonifacci et al., 2018; Gutiérrez-Clellen et al., 2008; Rezzonico et al., 2015; Verhoeven et al., 2011). This means that clinicians should be careful to compare a bilingual child only with others with very similar language learning experiences before using grammatical measures to identify a language disorder.

Macrostructure

In the area of macrostructure, a number of studies have found no difference between bilingual groups with typical language development and those with DLD on measures of story grammar (Altman et al., 2016; Iluz-Cohen & Walters, 2012; Tsimpli et al., 2016). Using story retells in a sample of Hebrew–English preschool children, Altman et al. (2016) found no differences between children with and without DLD on macrostructure as measured by the proportion of story grammar events (goals, attempts, and outcomes) reproduced in the story. They concluded that "narrative microstructure is the core of the problem for [DLD] in bilingual as well as in monolingual children" (Altman et al., 2016, p. 185).

This conclusion may be premature in light of a recent surge in investigations of narrative macrostructure in bilingual children with and without DLD. Many of these investigations have demonstrated significant group differences (Boerma et al., 2016; Govindarajan & Paradis, 2019; Rezzonico et al., 2015; Squires et al., 2014). For example, Boerma et al. (2016) used the MAIN (Gagarina et al., 2012) to examine macrostructural skills in Dutch, the L2 of a group of 66 preschool bilingual children with and without DLD. The children's L1s varied. Across both the comprehension and production macrostructural variables on the MAIN, bilingual children

with DLD scored below their TD peers with large effect sizes (Boerma et al., 2016). Furthermore, using a classification model with three macrostructural measures—the number of internal state terms produced by the child plus two story grammar comprehension scores—researchers achieved 79% sensitivity and 88% specificity within the bilingual sample. Paradis, Schneider, & Duncan (2013) also found that the story grammar score for a narrative told in the L2 (English) could contribute to the identification of DLD in a sample of 5-year-old bilinguals with varied L1s, although other variables had better predictive power in this work. These results suggest that macrostructural variables could indeed be diagnostically useful.

The causes of conflicting results in the area of narrative macrostructure are not clear but could relate to differences across study samples in the parameters of bilingualism discussed at the beginning of this review. For example, Govindarajan and Paradis (2019) attribute differences across studies to varying levels of L2 exposure in participants, combined with differences in task difficulty and bilingual children's ability to transfer some linguistic skills more easily than others. More specifically, they suggest that because bilingual children transfer narrative macrostructure skills to an L2 more quickly than other skills, macrostructure may be a good identifier of DLD in the early stages of L2 learning (Govindarajan & Paradis, 2019). As the capability of macrostructure to identify DLD in bilingual children is explored, it also is important to recall the nearly consistent conclusion that bilingual children are more comparable with monolingual children in this area than on microstructural indices (Boerma et al., 2016; Cleave et al., 2010; Rezzonico et al., 2015; Tsimpli et al., 2016). If the circumstances under which this variable reliably differentiates children with DLD could be established, it could be a less biased assessment measure.

Narrative language samples also have been used as a vehicle for conducting dynamic assessment, another potentially less biased assessment for bilingual children (De Lamo White & Jin, 2011). Dynamic assessment of macrostructure and microstructure from narrative language samples has shown promise for identifying bilingual children with DLD (e.g., Peña, Gillam, & Bedore, 2014). However, an in-depth review of dynamic assessment is outside the scope of this review and the reader is referred elsewhere (e.g., Peña et al., 2014) for more information.

USING LSA TO CHARACTERIZE LANGUAGE IN BILINGUAL CHILDREN

The identification of a disorder is not the only purpose of a clinical language assessment. Several additional assessment purposes may be subsumed under the broad goal of characterizing a child's language skills. For example, clinicians may wish either to describe specific strengths and weaknesses or to document growth over time. These questions lead to within-child comparisons, as opposed to the between-child comparisons discussed in the last section. Important within-child questions for bilingual children include comparisons across languages and over time.

Cross-linguistic comparisons

For both monolingual and bilingual children, LSA is an excellent tool for documenting ecologically valid language weaknesses and using them to develop targets for intervention (Costanza-Smith, 2010; Ebert & Pham, 2017; Rojas & Iglesias, 2009). To do this, clinicians may compare LSA measures with each other—say, to determine whether grammaticality is a relative weakness in comparison with lexical diversity for an individual child—and then examine the language sample for specific errors to target (in this example, specific grammatical errors). For bilingual children, clinicians also may wish to make comparisons across languages. Clinicians may be interested in utilizing LSA to determine overall dominance or to make more nuanced comparisons (such as whether sentence length or grammaticality is better in the L1 or the L2).

To make such determinations, it is important to establish that cross-linguistic comparisons are valid. There are certainly examples in the literature of direct cross-linguistic comparisons at the group level (Iluz-Cohen & Walters, 2012; Kapalková et al., 2016) and the individual level (Ebert & Pham, 2017). Clinicians can take several precautions to increase the chances of making valid comparisons across languages. First, sampling procedures should be as closely matched as possible across languages. Kapalková et al. (2016) utilized the MAIN, which contains multiple comparable stories, and Ebert and Pham (2017) utilized two of the *Frog* series books. In contrast, Iluz-Cohen and Walters (2012) elicited narratives with wordless versions of three different folktales and concluded that some significant cross-linguistic differences were an artifact of children providing longer samples for one of the three stories. In other words, differences across the stories they used to elicit samples made it difficult to compare samples across languages.

Once stories are collected, some measures may provide more valid cross-linguistic comparisons. Clinicians may wish to consider macrostructural measures such as MAIN scores, counts of story grammar events, or Narrative Scoring Scheme scores (again, provided the stories used for elicitation are comparable; see Table 1 for more information on measures). Several recent studies have compared the L1 with the L2 on macrostructural measures (Altman et al., 2016; Iluz-Cohen & Walters, 2012; Squires et al., 2014).

Microstructural measures may vary in their utility for cross-linguistic comparison. MLU is a classic measure of global language development that has frequently been used for cross-linguistic comparisons, but there is near-universal agreement that it should be calculated in words rather than morphemes in order to do so (Gutiérrez-Clellen, Restrepo, Bedore, Peña, & Anderson, 2000; Rojas & Iglesias, 2009). This recommendation stems from cross-linguistic differences in morphological systems.

Beyond MLU, Rojas and Iglesias (2009) recommend NDW and words per minute (WPM) as good measures to collect in both languages. These are calculated by the SALT software and raw scores can be compared with a database mean. Ebert and Pham (2017) provide a case example of how such comparisons can be conducted for an individual child, utilizing the comparison data from SALT. This case study examines z scores across measures and languages to find relative areas of weakness, followed by a qualitative analysis of specific errors to develop therapy goals.

However, the availability of comparison data for these measures is currently limited to Spanish–English bilinguals. Using raw scores, NDW may not be an optimal measure for cross-linguistic comparison, given its dependence on sample length and evidence that it differs across stories that are otherwise comparable (Heilmann et al., 2016). Similarly, measures that are highly language-specific or context-dependent are unlikely to provide valid cross-linguistic comparisons using raw scores. Measures of overall productivity, for example, are both context- and language-dependent. Similarly, counts of specific grammatical errors rely on the morphosyntax of the language. Global indices of grammaticality (such as percent grammatical utterances) have greater potential for cross-linguistic comparison (e.g., Ebert & Pham, 2017), but additional work is needed to examine the conditions under which these comparisons are valid.

Cross-linguistic transfer

The literature comparing LSA measures across languages (but within children) also may be useful for cases in which direct assessment of both languages is not possible. If close cross-linguistic relationships have been demonstrated for a particular domain, clinicians may be able to directly assess one language and then make inferences about the other for that domain (Kapalková et al., 2016). In contrast, if cross-linguistic relationships are weak, then cross-language inferences are contraindicated. Cross-linguistic transfer is perhaps even more important for treatment planning; if a skill can transfer

across languages, a clinician can utilize cross-linguistic transfer to facilitate its growth in a weaker language.

Narrative macrostructure may be the strongest candidate for cross-linguistic transfer (Fiestas & Peña, 2004; Iluz-Cohen & Walters, 2012; Squires et al., 2014; Uccelli & Paéz, 2007), both in typically developing children and in those with DLD. Macrostructural skills are hypothesized to be more dependent on underlying cognitive–linguistic abilities and less dependent on language-specific experience and knowledge (see Iluz-Cohen & Walters, 2012, for discussion). However, results in this literature have been mixed, with some studies showing that macrostructural skills are related across languages (Lucero, 2018) or may actually transfer across languages (Bonifacci et al., 2018; Squires et al., 2014), and other studies finding cross-linguistic differences (Kapalková et al., 2016) or nonsignificant associations (Altman et al., 2016). The typological relationship between the L1 and L2 of the participants in these studies may play a role in the conflicting results; both Squires et al. (2014) and Lucero (2018) studied Spanish–English bilinguals and at least some of the bilinguals in Bonifacci et al.'s (2018) study spoke closely related languages (e.g., Spanish and Italian; the L1s of participants varied in this work), whereas the participants in the two studies that did not find cross-linguistic macrostructural relationships spoke more distantly related language pairs (e.g., Slovak–English in Kapalková et al., 2016; Hebrew–English in Altman et al., 2016).

Microstructural skills are argued to draw more heavily on language-specific experience, reducing the amount of cross-linguistic transfer. However, there are several studies that show significant cross-linguistic relationships on lexical diversity measures among typically developing bilingual children (Kohnert, Kan, & Conboy, 2010; Lucero, 2018; Pham, 2016). Although the number of words a child knows in each language is clearly dependent on experience or exposure, lexical diversity measures from narratives may tap into an underlying ability to use language in storytelling (Pham, 2016). This explanation would predict positive cross-linguistic relations on measures such as NDW, and the potential to infer that lexical diversity—that is, the ability to use an array of words in storytelling—may be a general weakness if a child scores poorly in one language. It should be noted, however, that the majority of studies on cross-linguistic lexical diversity relations have considered children with typical language development. In a small group ($N = 12$) of 5- to 6-year-old Hebrew–English children with DLD, Altman et al. (2016) found positive but nonsignificant ($r = .47$) cross-linguistic relations on NDW from narrative retells. Additional evidence from children with DLD would strengthen the case for relying on cross-linguistic transfer of lexical diversity skills in clinical situations.

Within-language growth

Finally, the literature supports the use of LSA to document language growth over time, and this has been demonstrated specifically for bilingual children (Lucero, 2018; Pham, 2016; Rojas & Iglesias, 2013; Squires et al., 2014). Characterizing language growth during treatment is clinically important. Moreover, tools such as norm-referenced tests may be ill-suited to this purpose, particularly for bilingual children. Language sample analysis can show individual growth in two languages in sentence length (Lucero, 2018; Pham, 2016; Rojas & Iglesias, 2013), lexical diversity (Pham, 2016; Rojas & Iglesias, 2013), verbal fluency (Rojas & Iglesias, 2013), and macrostructure (Squires et al., 2014). Many of the caveats noted in previous sections, such as using comparable stories and conditions for eliciting samples across time, apply to this type of within-child comparison as well.

The existence of studies tracking bilingual language growth over time using LSA also provides important comparison data for clinicians. For example, the large-scale, longitudinal analysis conducted by Rojas and Iglesias (2013) documented the language growth of more than 1,700 Spanish–English bilingual

children from kindergarten through the second grade. The prototypical growth trajectories generated by this project show phenomena such as a deceleration of L1 growth in the first grade, overall rapid growth in the L2 but loss of L2 skill during summer vacations, and differences in growth trajectories across different LSA measures.

Children who speak other languages or who are learning in a different sociolinguistic environment may demonstrate different patterns of growth than those documented in this work (Rojas & Iglesias, 2013). A few other studies have documented patterns in other samples of Spanish–English bilinguals with and without disorders (Lucero, 2018; Squires et al., 2014; Uccelli & Paéz, 2007). For example, Squires et al. (2014) found that Spanish–English bilinguals with and without DLD improved story macrostructure in both languages between kindergarten and the first grade. However, children with DLD did not improve on microstructural measures in either language within that time frame, and those with typical language development improved only on Spanish microstructural measures. In a sample of typically developing school-aged Vietnamese–English bilinguals who were followed for 4 years, Pham (2016) found that both MLU and NDW grew in English over time. In Vietnamese (the L1), MLU grew but NDW did not.

It is clear that further work will be needed to increase the breadth of the literature on dual-language growth as documented by LSA. At this stage, clinicians should know that growth in each language is not necessarily linear or even positive for bilingual children, even those with typical language development. Slowed growth might be expected in the L2 during the summer months for children whose primary source of L2 input is school (Rojas & Iglesias, 2013), and a lack of growth might be apparent in the L1 for those who speak a minority language without widespread community support (Pham, 2016). A lack of growth might be pronounced for children with DLD (Squires et al., 2014).

CASE STUDIES

To illustrate some of the ways in which the literature on LSA in bilingual children can guide clinical decision-making, this section presents two distinct case studies.

Case 1: Yasemin

Yasemin is a 6-year-old girl in the Netherlands who is entering her third year of formal schooling. Her family speaks exclusively Turkish at home and her schooling is in Dutch. Yasemin has no known history of developmental delays, but she is struggling to acquire Dutch and to keep pace with academic expectations. She is referred for assessment to determine whether a developmental language disorder is present. The clinician speaks only Dutch.

The clinician assessing Yasemin could look first to the work of Boerma et al. (2016), who included 5- and 6-year-old children who spoke Dutch as their L2. About one third of the bilingual participants spoke Turkish as a first language. However, the reported range of exposure to Dutch before the age of 4 years among participants is 20%–67%; the clinician should further probe whether Yasemin had any exposure to Dutch via the community, preschool programs, or other sources before the age of 4 years in order to determine how closely she matches the participants on this parameter. Following Boerma et al. (2016), the clinician could administer and score the MAIN in Dutch (Gagarina et al., 2012). This would require Yasemin to produce her own story and also to answer comprehension questions about both her own story and a model story. The clinician could focus on macrostructural variables by counting the number of internal state words Yasemin produces, the percentage of internal state elements she includes in her story, and the percentage of comprehension questions related to internal states that Yasemin answers correctly. For bilingual children, these three measures correctly classified 85% of children with DLD as well as 85% of children with typical language development in Boerma et al. (2016).

The clinician also might calculate the percentage of grammatical utterances in Yasemin's narrative language sample, as this was noted to be one of the most promising measures for identifying DLD in the literature reviewed here (though it was not explored in Boerma et al.'s, 2016 study). Verhoeven et al. (2011) found that the percentage of grammatical utterances was much lower for bilingual children who spoke Dutch as the L2 (with Moroccan or Turkish as the L1), although the participants in this study were slightly older than Yasemin. Seven-year-old children with DLD averaged nearly 67% ungrammatical utterances in Dutch in this study, whereas those with typical language development averaged 50%. For one additional comparison on this measure, the clinician could refer to Bedore et al. (2010), who report the percentage of grammatical utterances in Spanish–English bilinguals close to Yasemin's age. The average percentage of ungrammatical utterances in the L2 was 54%, across a group that included children with and without DLD. Taken together, the clinician might expect around 50% of Yasemin's Dutch utterances to be grammatical if she is typically developing and lower if she has DLD.

It is important to note that a comprehensive evaluation for Yasemin would also examine her abilities in her L1, perhaps through a combination of parent report and assessment of Turkish using an interpreter. However, the studies summarized here highlight the ways in which the clinician could use LSA in the L2 to begin the process of determining whether or not Yasemin demonstrates DLD.

Case 2: Fernando

Fernando is a 7-year-old boy who has received school-based speech–language services for the past 2 years. He is learning both Spanish and English; his family uses mostly Spanish at home (around 80%, as his older siblings prefer to use English with each other), and he has been exposed to English in preschool and elementary school. Fernando is completing a reassessment to evaluate his progress in language treatment over the past year and to determine new goals. The clinician speaks both Spanish and English.

Fernando's clinician could begin by collecting narratives in both Spanish and English. Ideally, he or she would collect these samples on separate days. The clinician could choose to use two different *Frog* stories, as these are well supported in research on Spanish–English bilinguals in the United States (Bedore et al., 2010; Ebert & Pham, 2017; Gutiérrez-Clellen et al., 2008; Heilmann et al., 2016; Rojas & Iglesias, 2013). If Fernando lives in a community that is similar to those where the SALT bilingual databases were collected, the clinician may want to utilize the software to determine how far Fernando falls from the mean for his age on measures such as MLU, NDW, WPM, and grammatical errors in both languages. For example, if Fernando fell 2.2 standard deviations below average on his Spanish MLU in the last year's sample but now scores 1.4 standard deviations below average for this measure, the clinician might conclude that Fernando has made significant growth in constructing Spanish sentences.

The clinician also might look at Fernando's scores across languages to help determine relative areas of weakness and his goals for the upcoming year. For example, the clinician could consider the macrostructure of Fernando's stories, taking into account evidence that macrostructure may be cross-linguistically comparable (Ebert & Pham, 2017; Iluz-Cohen & Walters, 2012; Kapalková et al., 2016). The clinician could score the NSS for this purpose or simply count the number of story grammar events in each sample and compare them. To extend this example, perhaps Fernando shows an emerging ability to mention story grammar elements in his Spanish story—including characters, setting, conflict, and resolution—but a limited ability to use these same elements in English stories. The clinician might set a goal to

address macrostructure in English and then work with Fernando to explicitly connect the Spanish story grammar elements he is using to their English counterparts. Of course, it is likely that Fernando would have other targets and that the clinician would use additional data to assess growth and progress. Nonetheless, the literature reviewed here would support the utility of LSA in Fernando's assessment.

SUMMARY AND FUTURE DIRECTIONS

In the last two decades, the literature relevant to bilingual LSA has grown tremendously. As a result, we can begin to provide answers to the questions of *why*, *how,* and *what* that were raised at the beginning of this review. Clinicians have a solid rationale for choosing LSA in the assessment of bilingual children, as well as guidelines for eliciting narrative samples, managing code-switching, and determining which measures are most valid and useful for their population and assessment purposes. Several LSA measures can differentiate between bilingual children with and without language disorders, though there appears to be a tension between the measures that most consistently identify disorder and the degree to which these measures are influenced by prior language experience (e.g., specific grammatical errors not only appear to be highly successful differentiators within homogeneous populations but also discriminate against children with less language-specific experience; macrostructural measures not only appear less likely to disadvantage bilingual children but also to differentiate children with disorders less consistently). In addition to helping identify the presence of language disorders, LSA can be used to compare skills across languages, predict whether skills will transfer across languages, and document language growth over time.

Despite this growing literature, there are a number of gaps in our study of LSA with bilinguals. Perhaps, the most glaring are those related to genre and modality. As noted earlier, the study of LSA in bilinguals is limited almost exclusively to oral narratives. A few investigations have integrated conversational samples (Ooi & Wong, 2012; Spoelman & Bol, 2012). In addition, Hsin and Snow (2017) recently investigated the written persuasive language samples of 41 bilingual children in fourth to sixth grades. Given the evidence that picture-based narratives may not adequately capture language growth in older children (Ebert & Pham, 2017; Ebert & Scott, 2014; Nippold et al., 2015), as well as the importance of written language, additional investigations of bilingual children that extend beyond oral narratives are sorely needed.

In addition, there is an ongoing need to resolve conflicting results in several of the areas reviewed here. Can macrostructural measures consistently identify children with DLD? How closely must linguistic experience be matched in order to make valid comparisons using microstructural measures? One path toward resolving these conflicts in future work is to describe bilingual participants more clearly. Although the age of participants, the languages they speak, and the length of exposure are frequently described in published studies of bilingual individuals, parameters such as the dominant societal language and the contexts of exposure to each language are much less frequently documented. Paying closer attention to these variables can help establish the circumstances under which particular phenomena apply as well as provide better comparison data for clinicians who work with bilingual children.

The diversity of bilingual language learners means that a great deal of additional work would be welcomed to investigate LSA for different languages, ages, genres, purposes, and sociolinguistic contexts. Further contributions will enrich the knowledge base regarding this flexible clinical tool and enhance clinical services for developing bilingual children.

REFERENCES

Altman, C., Armon-Lotem, S., Fichman, S., & Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English–Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics*, *37*, 165–193. doi:10.1017/S0142716415000466

Arias, G., & Friberg, J. (2017). Bilingual language assessment: Contemporary versus recommended practice in American schools. *Language, Speech, and Hearing Services in Schools*, *48*, 1–15. doi:10.1044/2016_LSHSS-15-0090

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *Bilingual Education and Bilingualism*, *11*, 1–29. doi:10.2167/beb392.0

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. doi:10.1016/j.jcomdis.2010.05.002

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T., & CATALISE Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*, 1068–1080. doi:10.1111/jcpp.12721

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *International Journal of Language and Communication Disorders*, *51*, 626–638. doi:10.1111/1460-6984.12234

Bonifacci, P., Barbieri, M., Tomassini, M., & Roch, M. (2018). In few words: Linguistic gap but adequate narrative structure in preschool bilingual children. *Journal of Child Language*, *45*(1), 120–147. doi:10.1017/S0305000917000149

Cleave, P. L., Girolametto, L. E., Chen, X., & Johnson, C. J. (2010). Narrative abilities in monolingual and dual language learning children with specific language impairment. *Journal of Communication Disorders*, *43*, 511–522. doi:10.1016/J.JCOMDIS.2010.05.005

Costanza-Smith, A. (2010). The clinical utility of language samples. *Perspectives on Language Learning and Education*, *17*, 9. doi:10.1044/lle17.1.9

De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language and Communication Disorders*, *46*, 613–627. doi:10.1111/j.1460-6984.2011.00049.x

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore: Paul H. Brookes Publishing.

Ebert, K. D., & Kohnert, K. (2016). Language learning impairment in sequential bilingual children. *Language Teaching*, *49*, 301–338. doi:10.1017/s0261444816000070

Ebert, K. D., & Mammolito, A. P. (2015, June). *Code-switching in the narratives of bilingual children with language impairment*. Poster presented at the Symposium of Research on Child Language Disorders, Madison, WI.

Ebert, K. D., & Pham, G. (2017). Synthesizing information from language samples and standardized tests in school-age bilingual assessment. *Language, Speech, and Hearing Services in Schools*, *48*, 42–55. doi:10.1044/2016_LSHSS-16-0007

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. doi:10.1044/2014_LSHSS-14-0034

European Commission (2012) Special Eurobarometer 386: Europeans and their languages (Executive summary) (PDF). Retrieved November 13, 2019, from Europa web portal: http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_386_en.pdf

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools*, *35*, 155–168. doi:10.1044/0161-1461(2004/016

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balciuniene, I., et al. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, *56*, 1–13.

Gatt, D., O'Toole, C., & Haman, E. (2015). Using parent report to assess early lexical production in children exposed to more than one language. In S. Armon-Lotem, J de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 149–193). Bristol, England: Multilingual Matters.

Govindarajan, K., & Paradis, J. (2019). Narrative abilities of bilingual children with and without developmental language disorder (SLI): Differentiation and the role of age and input factors. *Journal of Communication Disorders*, *77*, 1–16. doi:10.1016/j.jcomdis.2018.10.001

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

Guiberson, M. M., Barrett, K. C., Jancosek, E. G., & Itano, C. Y. (2001). Language maintenance and loss in preschool-age children of Mexican immigrants: Longitudinal study. *Communication Disorders Quarterly*, *28*, 4–17.

Gutiérrez-Clellen, V. F., Restrepo, M. A., Bedore, L., Peña, E., & Anderson, R. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools*, *31*, 88–99. doi:10.1044/0161-1461.3101.88

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

Gutiérrez-Clellen, V. F., Simon-Cereijido, G., & Leone, A. E. (2009). Code-switching in bilingual children with specific language impairment. *International Journal of Bilingualism*, *13*, 91–109. doi:10.1177/1367006909103530

Gutiérrez-Clellen, V. F., Simon-Cereijido, G., & Wagner, C. (2008). Bilingual children with language impairment: A comparison with monolinguals and second language learners. *Applied Psycholinguistics*, *29*, 3–19. doi:10.1017/S0142716408080016

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. doi:10.1097/01.TLD.0000318937.39301.76

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 and Communication Disorders*, *51*, 339–345. doi:10.1111/1460-6984.12201

Hsin, L., & Snow, C. (2017). Social perspective taking: A benefit of bilingualism in academic writing. *Reading and Writing: An Interdisciplinary Journal*, *30*, 1193–1214. doi:10.1007/s11145-016-9718-9

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism*, *15*, 58–74. doi:10.1017/S1366728911000538

Jacobson, P. F., & Walden, P. R. (2013). Lexical diversity and omission errors as predictors of language ability in the narratives of sequential Spanish–English bilinguals: A cross-language comparison. *American Journal of Speech-Language Pathology*, *22*, 554–565. doi:10.1044/1058-0360(2013/11-0055)

Kapalková, S., Polišenská, K., Marková, L., & Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak–English children: A cross-language comparison. *Applied Psycholinguistics*, *37*, 145–164. doi:10.1017/S0142716415000454

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

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

Kohnert, K., Kan, P. F., & Conboy, B. T. (2010). Lexical and grammatical associations in sequential bilingual preschoolers. *Journal of Speech, Language, and Hearing Research*, *53*, 684–698. doi:10.1044/1092-4388(2009/08-0126)

Langdon, H. W., & Saenz, T. I. (2016). *Working with interpreters and translators: A guide for speech-language pathologists and audiologists*. San Diego, CA: Plural Publishing.

Lazewnik, R., Creaghead, N. A., Breit Smith, A., Prendeville, J.-A., Raisor-Becker, L., & Silbert, N. (2019). Identifiers of language impairment for Spanish–English dual language learners. *Language, Speech, and Hearing Services in Schools*, *50*, 126–137.

Lucero, A. (2018). The development of bilingual narrative retelling among Spanish–English dual language learners over two years. *Language, Speech, and Hearing Services in Schools*, *49*, 607–621. doi:10.1044/2018_lshss-17-0152

Miller, J. F., Andriacchi, K., & Nockerts, A. (2015). *Assessing language production using SALT: Clinician's guide* (2nd ed.). Middleton, WI: SALT Software, LLC.

Miller, J. F., & Iglesias, A. (2012). *SALT: Systematic Analysis of Language Transcripts. Software for the analysis of oral language*. Middleton, WI: SALT Software, LLC.

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. doi:10.1044/2015_JSLHR-L-14-0129

Ooi, C. C. W., & Wong, A. M. Y. (2012). Assessing bilingual Chinese–English young children in Malaysia using language sample measures. *International Journal of Speech-Language Pathology*, *14*, 499–508. doi:10.3109/17549507.2012.712159

Paradis, J. (2010). *Alberta language environment questionnaire*. Retrieved from https://d1pbog36rugm0t.cloudfront.net/-/media/arts/departments-institutes-and-centres/linguistics/chesl/documents/aleq-26mar2010.pdf

Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating children with language impairment among English-language learners from diverse first-language backgrounds. *Journal of Speech, Language, and Hearing Research*, *56*, 971–981. doi://10.1044/1092-4388(2012/12-0050)

Peña, E. D., Gillam, R. B., & Bedore, L. M. (2014). Dynamic assessment of narrative ability in English accurately identifies language impairment in English language learners. *Journal of Speech, Language,*

and Hearing Research, *57*, 2208–2220. doi:10.1044/2014_JSLHR-L-13-0151

Pham, G. (2016). Pathways for learning two languages: Lexical and grammatical associations within and across languages in sequential bilingual children. *Bilingualism*, *19*, 928–938. doi:10.1017/s1366728915000899

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1398–1411. doi:10.1044/jslhr.4106.1398

Rezzonico, S., Chen, X., Cleave, P. L., Greenberg, J., Hipfner-Boucher, K., Johnson, C. J., et al. (2015). Oral narratives in monolingual and bilingual preschoolers with SLI. *International Journal of Language and Communication Disorders*, *50*, 830–841. doi:10.1111/1460-6984.12179

Rojas, R., & Iglesias, A. (2009). Making a case for language sampling. *ASHA Leader*, *14*, 10–13. doi:10.1044/leader.ftr1.14032009.10

Rojas, R., & Iglesias, A. (2013). The language growth of Spanish-speaking English language learners. *Child Development*, *84*, 630–646. doi:10.1111/j.1467-8624.2012.01871.x

Rojas, R., Iglesias, A., Bunta, F., Goldstein, B., Goldenberg, C., & Reese, L. (2016). Interlocutor differential effects on the expressive language skills of Spanish-speaking English learners. *International Journal of Speech-Language Pathology*, *18*, 166–177. doi:10.3109/17549507.2015.1081290

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

Spoelman, M., & Bol, G. W. (2012). The use of subject-verb agreement and verb argument structure in monolingual and bilingual children with specific language impairment. *Clinical Linguistics and Phonetics*, *26*, 357–379. doi:10.3109/02699206.2011.637658

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 and Communication Disorders*, *49*, 60–74. doi:10.1111/1460-6984.12044

To, C. K.-S., Stokes, S. F., Cheung, H.-T., & T'sou, B. (2010). Narrative assessment for Cantonese-speaking children. *Journal of Speech, Language, and Hearing Research*, *53*, 648–669. doi:10.1044/1092-4388(2009/08-0039

Tsimpli, I., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics*, *37*, 195–216. doi:10.1017/S0142716415000478

Uccelli, P., & Paéz, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, *38*, 225–236. doi:10.1044/0161-1461(2007/024)

U.S. Census Bureau. (2015). *2009–2013 American community survey: Detailed languages spoken at home and ability to speak English for the population 5 years and over: 2009–2013.* Retrieved November 13, 2019, from https://www.census.gov/data/tables/2013/demo/2009-2013-lang-tables.html

Verhoeven, L., Steenge, J., & van Balkom, H. (2011). Verb morphology as clinical marker of specific language impairment: Evidence from first and second language learners. *Research in Developmental Disabilities*, *32*, 1186–1193. doi:10.1016/j.ridd.2011.01.001