AJSLP

Tutorial

Beyond Scores: Using Converging Evidence to Determine Speech and Language Services Eligibility for Dual Language Learners

Anny Castilla-Earls,[a] Lisa Bedore,[b] Raúl Rojas,[c] Leah Fabiano-Smith,[d] Sonja Pruitt-Lord,[e] María Adelaida Restrepo,[f] and Elizabeth Peña[g]

Purpose: Speech-language pathologists have both a professional and ethical responsibility to provide culturally competent services to dual language learners (DLLs). In this tutorial, we recommend that clinicians use a comprehensive assessment of converging evidence to make diagnostic decisions in DLLs in accordance with the American Speech-Language-Hearing Association's Code of Ethics. The content of this tutorial is most appropriate for Spanish–English DLLs between the ages of 4 and 8 years.
Method: We propose a converging evidence approach, in which one single method is not the deciding factor in making diagnostic decisions regarding the dual language and speech production skills of DLLs. Converging evidence refers to the idea that multiple pieces of assessment data must come together and trend in the same direction to make a diagnostic decision. We recommend gathering assessment data using a combination of language experience questionnaires, bilingual language sample analysis using large-scale reference databases, evaluation of learning potential, and standardized testing. These four assessment methods allow clinicians to examine the child in different contexts to determine their strengths and weakness in communication abilities.
Conclusion: We illustrate the converging evidence framework using two case studies to guide the clinician through the diagnostic decision-making process.

School-based speech-language pathologists (SLPs) are required to provide culturally competent services to all children, including those who are dual language learners (DLLs; the American Speech-Language-Hearing Association [ASHA], 2016). The presence of DLLs (children who have learned two languages since birth or who are in the process of learning two languages; Genesee et al., 2004) in U.S. classrooms is increasing (Kena et al., 2016). For example, by 2025, almost 30% of U.S. school children will be Latino (Kena et al., 2016). In addition to Spanish, languages such as Arabic, French, Burmese, Ukrainian, Tigrigna, Afghani, Arabic, and Somali are common in classrooms across the United States. Language diversity is steadily growing with approximately 40,000 refugees admitted into the United States per year (Pew Research Center, n.d.). Because of this language diversity in the United States and the fact that communication disorders are high-incidence disabilities (Kohnert & Medina, 2009), SLPs face a significant clinical challenge: accurate identification of speech and language disorders in children who speak more than one language.

SLPs have openly expressed their struggles with this clinical challenge. Guiberson and Atkins (2012) surveyed SLPs and found that nearly 75% of the respondents indicated that they had a sizable number of DLLs on their caseloads. SLPs reported that the lack of appropriate assessment tools was a major challenge. According to Hammer et al. (2004), SLPs have some confidence when

[a]Department of Communication Sciences and Disorders, University of Houston, TX
[b]Temple University, Philadelphia, PA
[c]The University of Texas at Dallas
[d]The University of Arizona, Tucson
[e]San Diego State University, CA
[f]Arizona State University, Tempe
[g]University of California, Irvine
Correspondence to Anny Castilla-Earls: annycastilla@uh.edu

Editor-in-Chief: Julie Barkmeier-Kraemer
Editor: Mark Guiberson
Received June 3, 2019
Revision received November 7, 2020
Accepted February 11, 2020
https://doi.org/10.1044/2020_AJSLP-19-00179
Publisher Note: This article is part of the Forum: Innovations in Clinical Practice for Dual Language Learners, Part 2.

Disclosure: *Lisa Bedore and Elizabeth Peña are co-authors of the BESA and receive royalties from its sale.*

assessing DLLs whose primary language is English but lack confidence when assessing DLLs who are more proficient in their other language.

SLPs have both a professional and ethical responsibility to provide culturally competent services to DLLs. Through ASHA's Code of Ethics and its 2019 Position Statements, clear guidance is provided on how we should approach speech and language assessment in DLLs to align with federal policy on this issue. As applied to bilingual service delivery, ASHA's Principle of Ethics I, Rule C from ASHA's Code of Ethics (https://www.asha.org/Code-of-Ethics/) requires that we do not discriminate in the delivery of professional services, that we have a professional obligation to ensure that appropriate services are made available, and that we engage in lifelong learning to develop knowledge and skills required to provide culturally and linguistically appropriate services. Despite the inherent challenges associated with clinical service delivery for DLLs, we are ethically required to put forth our very best efforts, based on currently available knowledge, in order to provide equitable services to all children.

Consistent with the ASHA (2019) guidelines, we recommend that clinicians use a comprehensive assessment approach to enable them to use a converging evidence framework to make diagnostic decisions in DLLs. Converging evidence refers to the idea that multiple pieces of assessment data must come together and trend in the same direction to make a diagnostic decision. We recommend gathering assessment data using language experience questionnaires, bilingual language and speech sample analysis using large-scale reference databases (RDBs) when available, evaluation of learning potential, and standardized testing. These four assessment methods allow clinicians to examine the child in different contexts to determine their strengths and weakness in communication abilities. In this tutorial, we first present the four approaches to gather assessment data, and second, we develop the convergence evidence framework using two case studies to guide the clinician through the diagnostic decision-making process. Although the assessment principles discussed in this tutorial may be applied to a diversity of DLLs, the content of this tutorial is most appropriate for Spanish–English DLLs between the ages of 4 and 8 years.

Language Concern and Experience Questionnaires

Parents and teachers can provide informative and reliable insights into DLLs' speech and language knowledge that can be used to judge concerns that contribute to clinical decisions. Questionnaires are useful for guiding quantification of concerns about child language development. The information derived from language experience questionnaires provides data on the nature and timing of child language input and output across their languages that helps contextualize diagnostic decisions.

Parent concern has long been identified as a useful tool for identification of language disorders in DLLs (Restrepo, 1998). Parents and teachers are most accurate in judging the languages that they speak—thus most often it is the case that teachers' English judgments and Spanish parent ratings align with testing results. Parents and teachers may also attend to different aspects of child language, with parents focusing on vocabulary knowledge and teachers judging difficulties with form, including grammar (Bedore et al., 2011; Gutiérrez-Clellen & Kreiter, 2003).

Cultural differences likely influence the results of questionnaires across groups. Caregivers may have different expectations about child development, behavior, and appropriate level of support for their children. In the realm of speech sound disorders, Stertzbach and Gildersleeve-Neumann (2005) found that the question "Do *others* find your child difficult to understand?" versus "Do *you* find your child difficult to understand?" may yield more accurate reports of low intelligibility from Hispanic/Latino mothers during the parent report process. Peña and Mendez-Perez (2006) reported that while parents may describe behaviors that a clinician might associate with communication disorders, parents may not attribute concern to said behaviors. As such, it is more informative to collect parent and/or teacher description of behaviors or behavioral comparisons. SLPs can then make decisions regarding concerns on the basis of these descriptions rather than asking about concerns. Thus, descriptions of behaviors such as elicited on the Inventory to Assess Language Knowledge (ITALK) from the Bilingual English–Spanish Assessment (BESA; Peña et al., 2018), reports about communicative behaviors such as the Communicative Development Inventories (Fenson et al., 2006), or comparisons to the development of siblings or peers (Paradis et al., 2010) may all be informative approaches to gathering language experience data. Analyses of measures, such as the ITALK, show, for example, that parent and teacher concern have overall fair sensitivity (ability of a measure to accurately identify children who have language/speech disorders) and specificity (ability of a measure to accurately identify children who have typical speech/language skills) overall in the early school years (e.g., Pratt et al., in press).

Language experience questionnaires provide a mechanism for quantifying language input and output, cumulative patterns of exposure, patterns of interactants, and types of language experiences. Early in development, there is a strong match between the amount of experience that children have in each language and their performance (Bedore et al., 2016; Place & Hoff, 2011; Unsworth et al., 2014). As children are first exposed to their second language, it is the opportunities to use the language that contributes most to their gains (Bohman et al., 2010). Gains in semantics or word knowledge are associated with opportunities to hear the language, whereas gains in morphosyntactic performance are associated with opportunities to both hear and use the language. As children progress to early school grades (first and third grade), cumulative experience comes to take on a more important role than current experience, which is most important early in development (Bedore et al., 2016). Given that English is the predominant language of schooling, it is safe to assume that the large

majority of children have regular access to English. In contrast, Spanish is more sensitive to continued use and varies as a function of both current and cumulative exposure. These factors point to the utility of questionnaires in helping SLPs start to understand if children have acquired as much knowledge of their languages as might be expected given their experience.

One recent clinical application of this work is to study how performance in English might systematically vary in regard to language experience (Bedore et al., 2018). Children's English performance on elicited productions as part of the BESA database (Peña et al., 2018) was analyzed relative to children's experience. Children with and without language disorders who hear and use primarily English could be reliably differentiated on the basis of many of the same clinical markers of language disorders in English (e.g., regular past, third-person singular present) as their English monolingual peers. Typically developing children and children with language disorders who still hear and use predominantly Spanish tended to overlap in their accuracy of production of these forms (Bedore et al., 2018). These findings suggest that continued exploration of the relationship between language experience and language performance will provide a way to develop benchmarks that can inform clinical practice relative to what we know about dual language acquisition.

Bilingual Speech and Language Sample Analysis

Bilingual Speech Sample Analysis

Bilingual speech samples are a useful assessment tool for the identification of speech sound disorders in bilingual children. Both single-word and connected speech samples should be recorded in all languages of the child. Word lists from standardized tests, such as the BESA, can be used to collect single-word speech samples, and narrative or play tasks (often used in the language component of the evaluation) can serve as the connected speech samples. Single-word samples should be phonetically transcribed, whereas connected speech samples can serve as a means to judge intelligibility. Phonetic transcriptions are then analyzed for accuracy, error frequency and type, and complexity of the speech sound system. Fabiano-Smith (2019) outlined a series of criterion-referenced measures that can be used to aid in identification of speech sound disorders in bilingual children: (a) percent occurrence of phonological error patterns, (b) Percent Consonants Correct–Revised (PCC-R; Shriberg & Kwiatkowski, 1982), (c) percent intelligibility, (d) phonetic inventory complexity, and (e) stimulability. Criteria that distinguish typical from disordered abilities on these measures have been established in research studies examining both monolingual English-speaking children and bilingual Spanish- and English-speaking children (see Fabiano-Smith, 2019, for more details and cutoff scores). Importantly, these measures test a child's speech abilities at different levels of the sound system and provide not only quantitative data on speech sound performance (i.e., PCC-R, percent occurrence of phonological patterns) but also qualitative data on a child's strengths and weaknesses (i.e., phonetic inventory complexity, type of phonological error patterns used). Looking across both languages, at the type and frequency of phonological errors, allows the SLP to not only diagnose a speech sound disorder but also determine the nature of that disorder (e.g., articulation vs. phonological disorder or combined).

Once speech samples are collected, transcribed, and analyzed across both languages, the typical or disordered phonological status of the child can be observed. Typical skills in one language indicate typical development overall in the bilingual child, as low performance in the other language is most likely due to low proficiency and can be improved with exposure and use. A high rate of errors, across differing levels of the sound system and across both (or all) languages, indicates speech sound disorder in bilingual children.

Bilingual Language Sample Analysis

Language sample analysis (LSA) is recognized as a useful tool for the clinical assessment of child language development and disorders (Paul & Norbury, 2012). Importantly, as noted by Miller et al. (2019), language samples are representative of the manner in which speakers communicate in daily situations by documenting "oral language in a way that captures the speaker's typical and functional language use" (p. 1). When there is a robust database regarding performance on language measures, clinicians are able to make reliable judgments about children's language production using language samples. Bilingual LSA via narrative elicitation is a recommended approach for the clinical assessment of DLLs (Ebert & Pham, 2017; Peña et al., 2006; Rojas & Iglesias, 2009). There are growing databases on bilingual narrative production, such as the Bilingual Spanish and English Story Retell RDBs in Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2019), that clinicians can use to inform decision making (Ebert & Scott, 2014; Goldstein, 2012; Peña et al., 2006; Rojas & Iglesias, 2009).

Narrative LSA offers multiple advantages for assessing DLLs (Goldstein, 2006; Heilmann et al., 2016; Muñoz et al., 2003; Rojas & Iglesias, 2013; Rojas et al., 2016; Thompson et al., 2004). It uses real and functional communication contexts, including academic contexts, and the same context can be used to elicit a sample of Spanish and English. It simultaneously indexes multiple expressive language domains, such as vocabulary and morphosyntax, with preschool- and school-age children. Using LSA, clinicians can incorporate least biased assessment principles. For example, clinicians can prescreen a picture storybook for cultural appropriateness prior to using it to elicit a narrative language sample from a child. Large-scale cross-sectional and longitudinal studies with over 1,200 DLL children have described LSA measures from narrative retells produced by DLL children in English and Spanish (e.g., Castilla-Earls et al., 2019; Miller et al., 2006; Rojas et al., 2019).

In addition, LSA can be used with other measures validated in research studies. For example, Restrepo (1998) demonstrated that the use of a grammaticality index (percentage of grammatical utterances) obtained through LSA in combination with parent report achieved excellent diagnostic accuracy (sensitivity 91.3% and specificity 100%) with DLLs. Other measures of accuracy, productivity, and diversity of tense marking using story retells have shown promise in their ability to discriminate between typically developing DLLs and children at risk of language disorders (Potapova et al., 2018) and to be of use in language growth progress monitoring (Gusewski & Rojas, 2017).

A series of papers and chapters provide detailed descriptions of software-assisted, narrative LSA with regard to methodological considerations and its administration with DLLs (Gutiérrez-Clellen et al., 2001; Heilmann et al., 2008, 2010, 2016; Rojas & Iglesias, 2009, 2019). We emphasize the application of the Bilingual Spanish and English Story Retell RDBs from the SALT 20 (Miller & Iglesias, 2019; RDBs also accessible in previous versions of SALT). The Bilingual Spanish and English Story Retell RDBs consist of 4,667 narrative retell language samples elicited from DLL children with typical development in kindergarten through Grade 3, ranging in age from 5;0 to 9;9 (years; months), providing a robust comparison set that can be leveraged to inform the clinical assessment of Spanish–English DLLs.

The recommended protocol for using the large-scale RDBs is to elicit narrative retell language samples using predefined scripts. Predefined scripts and audio samples told by a professional storyteller are available online in English and Spanish for a series of wordless picture storybooks (Mayer, 1969, 1973, 1974): https://www.saltsoftware.com/resources/elicaids/protocols. The clinician uses a predefined script and a storybook to narrate that story for the child in a target language (Spanish or English) and then asks the child to retell the same storybook in the same target language. The narrative should be audio-recorded for transcription. The narrative retell protocol offers a number of advantages as it uses wordless picture storybooks in order to rule out the influence of printed text on children's narrative language production, and it minimizes the demands of memory recall. In addition, the examiner's narrative model provides children with an extended model of the task, and it provides a degree of control over narrative content and its expected duration, ranging from 4 to 5 min, based on the narrative retells of more than 4,500 DLL children with typical development.

The narrative retells are then orthographically transcribed using transcription recommendations (for more details, see Rojas & Iglesias, 2019). Once the narrative retells have been transcribed, they are ready to be analyzed with SALT. Although multiple LSA measures can be analyzed using the Bilingual Spanish and English Story Retell RDBs, two are specifically discussed in the converging evidence case study examples presented later in this tutorial: mean length of utterance in words (MLUw; gross measure of morphosyntax) and number of different words (NDW; measure of lexical diversity). In addition, the proportion of grammatical utterances (%G) can also be calculated using SALT; Restrepo (1998) suggested a cutoff of 20% or more ungrammatical utterances in Spanish in order to identify DLLs with language disorders. In her study, the students were Spanish dominant. It is important, therefore, to take into account the strongest language of the child. In addition, the type and frequency of errors (e.g., overregularization errors vs. omissions) could also provide additional information.

To inform clinical decisions, the user conducts a database comparison. The user should select (a) one or more matching criteria (storybook, age, grade, and/or gender) to restrict the comparison set of database records, (b) a method to equate samples by length (entire transcript or same number of total words, complete and intelligible utterances, or amount of elapsed time), and (c) a standard deviation interval or cutoff (default = 1.0 *SD,* which is modifiable). Performance reports can be generated in SALT 18, which profile the performance of the child in question relative to the comparison set of each RDB, specifying areas of strength and weakness (transcript length, intelligibility, narrative structure, syntax/morphology, semantics, verbal facility, and errors) in each language while also providing specific examples in context from the child's narrative retell production. Interpretation of the results will be illustrated across the two converging evidence case study examples presented later in this tutorial.

In summary, bilingual narrative LSA with the application of large-scale RDBs (Bilingual Spanish and English Story Retell RDBs in SALT) facilitates a powerful, data-driven approach to the assessment of the expressive language development and performance of DLL children in English and Spanish. Database-referenced bilingual narrative LSA can be an important tool to make clinical decisions using a converging evidence framework.

Evaluation of Learning Potential

Two evidence-based approaches that focus on evaluation of children's learning potential are dynamic assessment (DA) and response to intervention (RTI). The advantage of these approaches is that the focus is on children's learning ability rather than on their performance. When children have not had the opportunity or exposure to learn specific language skills, standardized tests may underestimate language ability. For children who are learning language in dual-language environments or who are in the early stages of second language learning, standardized measures focused on English will often be uninterpretable. Low scores may be due to language disorders, cultural differences, or lack of experience. Examination of learning potential through incorporation of teaching in DA or RTI may be informative for clinical decision making, especially for children from low socioeconomic status or dual language backgrounds.

There are a number of features that are shared by DA and RTI. They focus on learning and provision of optimum instruction, and they can serve for identification for special

education eligibility and/or development of intervention goals. What distinguishes these general approaches, however, is a focus on process (i.e., DA) and product (i.e., RTI). These different foci manifest in distinct methods of implementation and evaluation of outcomes. The purpose of DA is to provide learning support and observe child strategies in response to that support. Identification of children's emerging learning strategies and need for development of strategies is at the core of DA, regardless of the assessment content. RTI is a framework for identifying children with emerging difficulties so that timely differentiated and preventive support and intervention can be provided according to children's individual needs. As a conceptual framework for early identification and prevention (Fuchs & Deshler, 2007), RTI is a paradigmatic model and not an assessment procedure.

DA

In general, DA employs a test–teach–retest approach. The test and retest phases of DA focus on a set of language targets that may pose difficulty for the child being tested. Targets selected should be developmentally appropriate and within the child's hypothesized learning ability. The pretest is used to establish a baseline for learning. The posttest (typically the same measure as the pretest or a parallel measure) is used to show the amount of gain or learning from pretest to posttest.

The teaching phase of DA incorporates principles of mediated learning experience to support the child's learning with the overarching goal of self-regulation (Lidz & Peña, 2009). Principles of mediated learning include intentionality, mediation of meaning, transcendence, and competence (Haywood & Lidz, 2007; Peña et al., 2001). Intentionality focuses on the goals for learning and helps the child to understand and articulate the goal. For example, in working on narratives, the clinician might state telling better and more complete stories as a goal. Mediation of meaning helps the child to understand the value or relevance of the stated goal. Here, the clinician would explain why storytelling is important and how stories are a part of everyday communication and learning. Transcendence is used to support hypothetical thinking as related to the stated goal. Questions that help the child to think hypothetically about stories at a metalinguistic level are examples of transcendence. For example, stories are often told in a given order, what would happen if the story was told out of order? This type of questioning is thought to help children develop metacognitive skills. Finally, competence or planning supports a plan for helping the child apply the learned strategies in other contexts. A session might close with reviewing stories and their components and working together to develop a method for remembering all the parts of a story.

A critical aspect of DA is the systematic observation of the child's responsivity to mediated learning experience. Examiners observe the child's deployment of cognitive strategies during the intervention portion of the DA and note the level of support that is required for the child to demonstrate success on a given task. The examiner can note emerging skills and ability to use new strategies during the intervention phase of DA.

A number of studies demonstrate the utility of DA for identifying children with language disorders from culturally and linguistically diverse backgrounds including African American, Latino, Native American, and Deaf populations. DA using a test–teach–retest approach has been developed for a number of language tasks including novel word learning (Kapantzoglou et al., 2012), vocabulary learning (Peña et al., 2001), categorization (Mann et al., 2015; Ukrainetz et al., 2000), and oral narratives (Henderson et al., 2018; Kramer et al., 2009; Peña et al., 2006, 2007). These studies in general show good classification accuracy of language disorders based on modifiability observations by themselves or in combination with gain or the posttest score. Table 1 describes the studies that have used DA and the measures they use for modifiability (see Table 1). The Kapantzoglou et al. (2012) study did not reach 80% correct sensitivity and specificity, but it too demonstrated that the best classification was derived from observation of modifiability plus children's gain scores.

Two recent studies using a test–retest approach focused on the utility of DA for DLLs when conducted in English (Peña et al., 2014; Petersen et al., 2017). While findings demonstrated minimal pre- to posttest gain, children with language disorder could be reliably differentiated based on observations of modifiability during the intervention portion of the DA. Modifiability observations and posttest scores together yielded classification rates close to or above 90% sensitivity and 90% specificity. Being able to reliably differentiate DLL children with and without language disorder could provide SLPs with a method to gain valid diagnostic information for clinical decision making.

DA can be used with bilingual children with speech sound disorders or suspected speech sound disorders via stimulability testing. When children do not demonstrate the ability to produce a speech sound in isolation but are then provided maximum auditory, visual, and tactile feedback through clinician modeling, children may modify and produce the sound in imitation. This outcome provides diagnostic information for the clinician, in both English and Spanish, that perhaps the target sound is emerging but not yet mastered. Children who are unable to produce the target sound given this precise model may present with a more severe impairment or perhaps have not yet reached the developmental level appropriate for the target sound. Numerous studies have examined stimulability testing in monolingual English-speaking children (e.g., Glaspey & Stoel-Gammon, 2007), and standardized tests of stimulability are available for monolingual English speakers (e.g., Glaspey, 2018); however, studies on stimulability that include bilingual children in their subject samples are not available to date. Therefore, current best practices for evaluation of speech sound disorders in bilinguals adapt methods of DA developed for and validated on monolingual English-speaking children.

A challenge for clinicians attempting to use DA clinically is that it takes time to implement (Arias & Friberg,

Table 1. Summary of dynamic assessment observation scales.

Modifiability subscale components	1. LSC	2. MS	3. MLO	4. MS PEARL	5. MI
Attention[a,c,e]	X		X	X	X
Disruption[d,e]					
Discrimination	X				
Comparative behavior	X				
Planning[a]	X		X		
Problem-solving[c]					
Self-regulation[a]	X		X		
Metacognition[c]					
Motivation	X		X		
Transfer	X	X			X
Examiner effort[b]		X		X	X
Number of prompts[d]					
Easy to teach[e]					
Responsivity[b]		X	X		X
Response to feedback[c,e]					
Anxiety			X		
Tolerance to frustration			X		X
Task orientation			X		
Nonverbal self-reward[c]			X	X	
Confidence[d]					
Verbal mediation			X		
Flexibility			X		
Compliance			X		
Task completion				X	
Potential for learning					X
Kapantzoglou et al. (2012)	X	X			
Peña et al. (2001)	X	X			
Mann et al. (2015)			X		
Ukrainetz et al. (2000)	X	X			
Kramer et al. (2009)		X[f]			
Henderson et al. (2018)				X	
Peña et al. (2006)		X[f]			
Peña et al. (2007)			X		
Petersen et al. (2017)					X
Peña et al. (2014)			X		

[a]LSC = Learning Strategies Checklist (Lidz, 1991). [b]MS = Modifiability Scale (Lidz, 1987). [c]MLO = Mediated Learning Observation (Peña et al., 2007). [d]PEARL MS = Modifiability Scale from the Predictive Early Assessment of Reading and Language (Petersen & Spencer, 2014). [e]MI = Modifiability Index (Petersen et al., 2017). [f]These studies scored effort and responsivity from the LSC, but not transfer.

2017). Even some of the shorter DAs take about 20 min for the intervention phase in addition to pre- and posttesting. Two studies have examined whether or not these can be shortened while retaining their diagnostic accuracy. Peña et al. (2007) reanalyzed data from Peña et al. (2006). In this study, modifiability scores at Time 1 and Time 2 were compared for children with and without language disorders. While there were significant differences between children with and without a language disorder, there were no effects for session. Subscores from the modifiability observations (metacognition and flexibility) classified children with 93% sensitivity and 92% specificity. Because the observation scores were highly consistent between the two interventions and because these observations alone classified children at such high levels, it may be possible to make a clinical decision on the basis of just one intervention session. Petersen et al. (2017) examined whether it was possible to make accurate decisions at various time points during the intervention. In their study, the pretest intervention and posttest were all completed in one 25-min session. Since they were able to obtain highly reliable observations from the first 10-min intervention, they speculated that it would be possible to reduce intervention time for the purpose of making a diagnostic decision. These results suggest that it is possible to complete a valid and reliable DA within a relatively short period of time.

Other identified barriers to the implementation of DA are lack of training as well as unfamiliarity with the procedures (Arias & Friberg, 2017). Training in DA is not a routine part of speech-language pathology training, so clinicians often seek this training elsewhere. For those seeking to implement DA, many of the above-cited studies include scripts and rating scales as part of the publication that clinicians can utilize.

RTI

RTI is a curricular approach. It provides specialized, individualized services or more focused small group

instruction that could potentially help underperforming children to accelerate their language acquisition, prepare them for reading instruction, and prevent academic difficulties resulting from language differences. From a functional perspective, any child who does not understand or produce language as expected could receive additional instruction to help them perform at the expected levels for the grade and age. RTI is such a tool. Unfortunately, studies employing RTI methods for phonological intervention are not yet available in the literature; therefore, validated methods developed for monolingual English-speaking children are often used as a model for bilingual intervention for children with speech sound disorders.

RTI includes the use of multiple tiers of instruction and intervention. Students who need more support transition to more intense arrangements of intervention, which are intensified by adjusting the intervention duration, content, frequency of intervention, and the expertise of the interventionist. Educators or SLPs other than classroom teachers assist in the delivery of targeted and intensive interventions, and tiered placement is determined irrespective of special education classification (Troia, 2005). RTI assumes appropriate high-quality classroom instruction in Tier 1, the use of evidence-based practices in all tiers, and valid progress-monitoring measures to determine adequate progress (Whittaker & Batsche, n.d.). If the child, despite more intensive intervention, continues not to make adequate progress and to score below peers, then the child is at high risk of presenting with speech and language disorder and/or other disability and should be referred for diagnostic testing. For example, a child with story comprehension and production difficulties receives regular classroom instruction in telling stories and answering questions about the story in Tier 1. In Tier 2, the child may receive small group instruction a couple of times a week with programs, such as Story Champs (Spencer & Petersen, 2012), that are structured and systematic in teaching story skills. In Tier 3, the child may receive more frequent and smaller group instruction or may have a specific area of difficulty.

RTI with multiple tiers has several advantages over the traditional general–special education dichotomy, the greatest being that supplemental intervention is delivered to those that need it, not just those with the appropriate diagnosis. In an RTI model, time is used providing additional instructional support for the child and not determining the causes of the delay. Despite the success of RTI for early reading intervention, language intervention has been largely neglected and is an area of great need for DLLs, particularly those living in poverty. If the goal is to ensure all children receive the necessary instruction and support to succeed in school, then more systematic language intervention should be considered for children with language differences and those at risk for low academic achievement.

In the traditional system, children who receive language support experience no intermediate step, such as Tier 2 intervention. There is no strategy for eliminating environmental confounds to language delays and no way to prevent speech and language disorders. Students go straight from classroom instruction to special education. In the ideal RTI model, children who do not respond well to the more intensive and focused stimulation or can only make gains in this high-intensity system would eventually qualify for special education. However, those who need a "push" will not be diagnosed with disabilities and progress to a more successful level in their education. Although the model seems promising and has been used successfully in some contexts (Spencer et al., 2015), it also presents with concerns, especially for DLLs.

One concern for children who are DLLs is that the instructional methods may not be culturally and linguistically appropriate or that the progress monitoring measures are not valid for them. A child who is learning English and receives intensive English reading instruction may make gains in English but may still lag behind in decoding or reading comprehension because they are still learning English. Moreover, the gains in the native language may be protracted due to the focus of the intervention and education being provided only in English with no native language support and therefore complicating the diagnostic process (Restrepo et al., 2010, 2013)

A second concern is how to differentiate those with speech and language disorders from those who are typical learners, when all the children benefit from the intervention. In a pilot study with preschool DLLs, children with language disorders made significant improvements with intervention, and changes were comparable to children with typical development (Restrepo et al., 2018). The results indicated that the children responded well to the intervention, but unless we have a way to rate or quantify the effort to effect change, it is difficult to differentiate who is having difficulty and who is not. Therefore, tools such as those used in DA, described above, to assess effort and responsiveness may be useful in the process of differentiating difference from disorder.

The danger with learning potential in this context of RTI is the assumption that the barriers and biases are removed in the process. However, learning occurs in cultural and linguistic contexts; the teaching styles and expectations differ across cultures and so do children's way of learning. Furthermore, the language of instruction impacts the learning rate and the depth of knowledge (Restrepo et al., 2010, 2013). Also, although learning potential can solve some biases in assessment, it is not free of biases. RTI can provide converging evidence in the assessment process by helping the clinician determine learning potential and ability in the cultural context. It provides progress monitoring data and assessment data that can be part of multiple methods for identifying speech and language disorders. If a child is not making adequate progress, is still scoring below the peers, and there are no other cultural, linguistic, or external factors that explain underperformance, then the child may need special education services (Whittaker & Burns, n.d.).

A final concern with RTI is the misuse of the process in delaying timely diagnosis and specialized intervention. The recommendation from the National Center for

Learning Disabilities and the Individuals with Disabilities Education Act is a timely diagnosis. It is critical that RTI not be used to delay such process and intervention (*Eligibility for Special Education Under a Specific Learning Disability Classification*, 2019; Whittaker & Batsche, n.d.). However, instructional response can be a valid component of a comprehensive assessment process as early as possible. Early identification and intervention are most important to ensure that the children receive appropriate and timely services and to minimize any negative impact on their academic achievement.

Standardized Testing

In clinical practice, standardized testing seems to be considered a trustworthy measure of child language ability. Standardized testing is perceived as comprehensive, easy to administer, and required in many cases to qualify a child for speech and language services within the school system (e.g., Fulcher-Rood et al., 2018, 2019; Huang et al., 1997; Pavelko et al., 2016). The results from standardized testing are often thought to be crucial to make decisions about service eligibility. For example, using a case review approach, Fulcher-Rood et al. (2019) found that, although SLPs administer a variety of assessment tools (e.g., parent report, language samples) during the assessment process, results from standardized testing drive diagnostic decisions in 98% of the cases (Fulcher-Rood et al., 2018). Despite this clinical preference for standardized testing, research supporting the use of standardized testing for the accurate identification of language disorders is minimal (Dollaghan, 2004; Dollaghan & Horner, 2011). In addition, there is a lot of variation in the quality and accuracy of the various diagnostic tests available for the assessment of speech and language abilities in DLLs (McCauley & Swisher, 1994; Plante & Vance, 1994). Furthermore, scores derived from standardized tests are susceptible to both random measurement error inherent to standardized testing and human error administration (Boyd et al., 2013; Rhoades & Madaus, 2003). Therefore, scores from standardized testing must be interpreted with caution. This is particularly important for DLLs who show variation in their patterns of language dominance and ability in both of their languages. To guide clinicians in their diagnostic process, we provide four principles of standardized test interpretation: (a) use standardized tests that are appropriate for DLLs; (b) when selecting a standardized test, review the evidence of diagnostic accuracy provided in the manual and its quality; (c) use empirically derived cutoff scores for the identification of speech and language disorders; and (d) interpret standard scores using confidence intervals.

Use Standardized Tests That Are Appropriate for DLLs

Current research evidence suggests that using standardized tests that target only one of the languages of the DLLs is not a valid approach to identify DLLs with language and speech disorders (Barragan, et al., 2018; Bedore et al., 2018; Lazewnik et al., 2018; Paradis & Crago, 2000; Paradis et al., 2008). On one hand, English-only testing for Spanish–English DLLs has been shown to overidentify typically developing children as presenting with language disorder because English language skills in these children are emergent and not fully developed (Bedore et al., 2018; Paradis & Crago, 2000; Paradis et al., 2008). On the other hand, Spanish-only testing does not seem to be an appropriate approach either, even when children are primarily Spanish-speaking DLLs. For example, Barragan et al. (2018) found that, when the Clinical Evaluation of Language Fundamentals–Fourth Edition in Spanish was used in a population sample of Spanish-speaking DLLs, it identified over half of the DLLs in the study as children with language disorders. Similarly, Lazewnik et al. (2018) found that using the Clinical Evaluation of Language Fundamentals Preschool–Second Edition (CELF Preschool-2) in Spanish only to assess Spanish–English DLLs resulted in a pattern of overidentification of language disorders. Standardized tests that evaluate speech sound disorders and include DLLs in the normative sample are few, are limited to Spanish and English languages, and are restricted to preschool-aged children (e.g., the BESA). This absence of standardized phonological assessment tools for bilingual children with suspected speech sound disorders can lead to both over- and underidentification of disorder (Fabiano-Smith, 2019). Therefore, the evidence suggests that testing only one language of the DLLs, even when the language is the dominant language, is not an appropriate approach for accurate identification of DLLs with language disorders. Furthermore, in Barragan et al., this group presented multiple risk factors, including low parental education, low income, and English-only education that focused on English instruction and no content instruction that seemed to affect performance.

In contrast, using a bilingual approach for the assessment of DLLs seems to yield more accurate identification of DLLs with language and speech disorders (Lazewnik et al., 2018). One method to apply a bilingual approach in language assessment is to test both languages of the DLLs and use the score in the best language for identification purposes. This is the approach used in the BESA (Peña et al., 2018). Lazewnik et al. (2018) found that the "best language" approach used in the BESA resulted in acceptable diagnostic accuracy for the identification of DLLs with language disorders. Another method to use a bilingual approach is to combine the standardized results of two omnibus tests, such as the CELF Preschool-2 in Spanish and English, to yield a composite score using the average of both tests. This composite score resulted in acceptable diagnostic accuracy (Lazewnik et al., 2018). Clinicians can utilize the word lists from standardized tests of phonology to evaluate speech sound disorders in DLLs by using the test's word lists in English and Spanish to record speech productions and then analyze those samples using criterion-referenced measures (e.g., phonetic inventory complexity, PCC-R, percent occurrence of phonological patterns) rather than calculating a standard score (Fabiano-Smith, 2019). In conclusion, the evidence suggests that the best approach to correctly identify DLLs with language disorders is to test both languages.

Bilingual versions of tests, such as the Expressive One-Word Picture Vocabulary Test–Fourth Edition (Martin & Brownell, 2013) and the Preschool Language Scales–Fourth Edition (Zimmerman et al., 2012), allow for responses in either language, and standardized scores are based on the total number of correct responses relative to the norming sample. Adaptations have been made to order items based on bilingual performance, but such measures are only as informative as the diagnostic accuracy of the original measure (Anaya et al., 2018). Bedore et al. (2010) found that 61% of a sample of Spanish–English DLLs who had data on semantic and morphosyntactic performance on the BESA had mixed dominance; 33% of the children had stronger semantics performance in English, whereas 25% demonstrated stronger cloze and/or sentence repetition scores in English. All possible patterns of dominance were observed in this group of bilinguals. Paradis et al. (2003) made a similar observation in a group of French–English children with language disorders. Furthermore, language production patterns and language experience can mismatch especially when children are more dominant in one language (Bedore et al., 2012). This is important because it is precisely at these transition points that clinicians need to know which language findings should most inform their decisions. Evidence from the BESA norming sample shows testing in both languages is needed when children have between 30% and 70% current exposure to the two languages.

When Selecting a Standardized Test, Review the Evidence of Diagnostic Accuracy Provided in the Manual and Its Quality

Sensitivity and specificity are two indices of diagnostic accuracy that describe the ability of a test to correctly identify children with language/speech disorders and children with typical language skills. Sensitivity refers to the ability of a diagnostic test to accurately identify children who have language/speech disorders. Specificity refers to the ability of a diagnostic test to accurately identify children who have typical language skills as not presenting with disorders. Sensitivity and specificity are both essential for diagnostic accuracy since the goal is to accurately separate these two groups of children in order to identify those who require language-speech services. As a guideline, diagnostic tests should have at a minimum a sensitivity and specificity of 80% (Plante & Vance, 1994). It is important to note that even diagnostic tools with good diagnostic accuracy are not free from diagnostic error: Although a diagnostic measure with 90% diagnostic accuracy is considered to be good (Plante & Vance, 1994), it misidentifies 10% of children in practice.

Likelihood ratios (LRs) are indicators of the relationship between specificity and sensitivity and express the probability that a test is clinically informative. Positive LR (sensitivity/1 − specificity) indicates the impact of a positive score on the probability of a child to truly have a disorder. Negative LR (1 − sensitivity/specificity) indicates the impact of a negative score on the probability of a child to truly have typical language skills. Tests with a positive LR over 10 are clinically informative to identify a child with a language disorder, whereas tests with a negative LR of .10 are clinically informative to rule out the presence of a language disorder (Dollaghan, 2007).

It is critical that SLPs select diagnostic tools that report diagnostic accuracy. Various diagnostic tests currently offer diagnostic accuracy information in their testing manuals, and there is also a growing body of external research evidence examining the diagnostic accuracy of various standardized tests (e.g., Barragan et al., 2018; Lazewnik et al., 2018). However, it is also important that clinicians evaluate the quality of this evidence. In particular, clinicians should look for information that shows evidence of bias, which poses a threat to the validity of diagnostic accuracy information of a test. For information on how to evaluate the quality of a diagnostic test, see Dollaghan (2004) and Dollaghan and Horner (2011).

In addition, it is fundamental that clinicians recognize that diagnostic accuracy is influenced by variation in the population that is being tested (Cohen et al., 2016). DLLs pose a case of high variability because how and when they learn and use two languages vary; as such, the application of any single measure or procedure for diagnostics is subject to validity threats if the applicability of the measure/procedure relative to the person being tested is not taken into account. For example, the BESA was developed for children in the United States who are acquiring English and Spanish in their community. However, at different levels of experience (i.e., primarily Spanish speaking) or with children who have different academic experiences (e.g., those enrolled in bilingual education), different tests or different norms may prove to be more informative (Barragan et al., 2018; Restrepo & Silverman, 2001).

Use Empirically Derived Cutoff Scores for the Identification of Speech and Language Disorders

Using a cutoff score to guide all eligibility decision making is not a recommended clinical practice (e.g., Spaulding et al., 2006). Research evidence supports the notion that cutoff scores should be particular to a standardized test and its norming sample. For example, tests such as the BESA and the Test of Early Grammatical Impairment (Rice & Wexler, 2001) provide cutoff scores that differ by age group and/or the desired specificity/sensitivity. Importantly, using a measure of deviation from the norm (standard deviation) to determine who is eligible for services fails to consider the presence of a disorder and instead focuses on the severity of the disorder (e.g., children on the lowest extreme of the language distribution of skills instead of those with language deficits; see Spaulding et al., 2006, for a discussion on this). In addition, children with language disorders often score within normal limits when a core language score is used because specific deficits in language (e.g., grammar) are often masked by other language abilities (e.g., vocabulary).

Interpret the Information Using Confidence Intervals

As mentioned previously, standard scores are prone to error (Boyd et al., 2013). Therefore, standard scores

should be interpreted using confidence intervals. Confidence intervals provided in diagnostic tests, such as the Core Language Score of the CELF Preschool-2 (Semel et al., 2004), are often presented at 68%, 90%, and 95%. If a 95% confidence interval is used, a score of 75 on the CELF Preschool-2 should be interpreted as being 95% confident (notice that this is not 100% confident) that the true score ranges between 68 and 82. When the confidence interval is used in conjunction with cutoff scores, higher confidence could be placed in a score for which the whole confidence interval is completely below or above the cutoff score. When a child has a score whose confidence interval includes the cutoff score, it is imperative to look more into other areas of language performance.

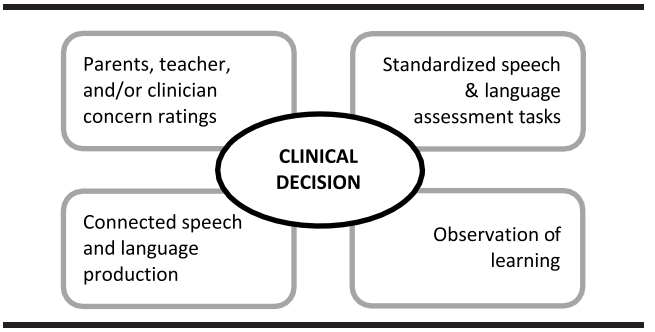
In regard to speech sound disorders, phonological skills are assessed in both, or all, languages of the child as part of a comprehensive pediatric bilingual assessment battery. The BESA provides phonology subtests in both English and Spanish and has strong psychometric properties. We are limited in our options, however, for testing languages other than English and Spanish and in our ability to derive scores for children younger than age 4;0 and older than age 6;11, as the current normative sample of the BESA is limited to children in this age range. To supplement standardized testing, or in the absence of a psychometrically sound standardized test in a specific language, informal measures allow us to arrive at an accurate diagnosis. The most current, evidence-based recommendation for the evaluation of speech in bilingual children is the use of multiple informal measures, in both languages, for the diagnosis of speech sound disorders.

Converging Evidence

Converging evidence is the standard approach used in research studies (Castilla-Earls et al., 2020; Gutierrez-Clellen & Simon-Cereijido, 2007; Peña et al., 2015). In the converging evidence framework, no single method provides the answer to the complex assessment of DLLs. In contrast, a variety of assessment data are used to make a clinical decision. We recommend collecting data from language experience questionnaires, language and speech samples, standardized testing, and measures of learning potential. Using converging evidence, the clinician uses a best practice approach to determine whether there is a language/speech sound disorder. To do so, the clinician weighs the different points of evidence based on multiple measures to make a decision in the context language and educational experiences (see Figure 1).

To use a converging evidence approach, a clinician first administers and collects a variety of assessment data. Second, the clinician weights all the available data equally because all the data suggested in this tutorial are valid for the identification of language and speech disorders in DLLs. To illustrate, although standardized testing is one of the components in the converging evidence approach, it does not by itself dictate the final diagnostic decision. Instead, standardized testing may be one of the pieces of available data, just as language experience questionnaires may comprise another one of the pieces of available data. It is important to note that it is possible to use a convergent evidence approach without a standardized test. In many cases, standardized tests are not appropriate for DLLs. It is also possible to reach a diagnosis of language/speech disorder, even when a standardized testing yields a score that could be considered within normal limits if the other evidence suggests a disorder. Finally, a clinician reaches converging evidence to make a diagnostic decision when the majority of the evidence suggests either a language disorder or typical language skills. Case Studies 1 and 2 show the process of making a clinical decision using a converging evidence approach.

Figure 1. Converging evidence framework.



Case Study 1

Our first case study is a girl who was age 6;1 at the time of the assessment. She was a participant in a research study investigating grammatical markers of language disorders in bilingual children. At the time of data collection, she was not receiving speech-language services. We collected information from parents and teachers using language experience questionnaires. In addition, we collected narrative retell language samples in both Spanish and English using the "Frog" books. The language sample was analyzed using SALT 18 (Miller & Iglesias, 2019). Last, we administered the BESA Morphosyntax (Peña et al., 2018) and the CELF Preschool-2 test in Spanish (Semel et al., 2009) and English (Semel et al., 2004). Results from the assessment are summarized in Table 2.

Following our converging evidence approach for clinical decision making, we first weigh all information available. In this case, we have information from three areas: language experience questionnaires, narrative retell language samples, and standardized testing. From the language experience questionnaires, we have evidence that the parents were concerned about the language skills of their daughter and that the teacher was somewhat concerned about the child's language. These concerns are often the initial reason why an evaluation takes place in the school system. Therefore, this information is crucial to establish that there is a potential impact in the communicative environment both at home and at school.

Table 2. Summary of assessment results for Case Study 1.

Language experience questionnaires	Standardized testing
• Parents reported that they were worried about their child's language. In the PRSLP (Restrepo, 1998), 19 questions were marked as concerns. • Teacher reported that the child was very low in English and was somewhat concerned about the child's language. **Language sample analysis** • MLUw-Spanish = 5.16 (–1.75 *SD*) • NDW-Spanish = 75 (1.02 *SD*) • PGU-Spanish = 67.57% • MLUw-English = 5.68 (–0.76 *SD*) • NDW-English = 36 (–1.1 *SD*) • PGU-English = 21.05%	• BESA Morphosyntax Spanish: 83 (CZ ss 4; SR ss 10) • BESA Morphosyntax English: 58 (CZ ss 3; SR ss 2) • CELF Preschool-2 Spanish Core Language: 76 (EP ss 6; RO ss 9; CB ss 2) • CELF Preschool-2 English Core Language: 50 (SS ss 3; WS ss 1; EV ss 1)

Note. PRSLP = parent report of speech and language problems; BESA = Bilingual English–Spanish Assessment; CZ = Cloze Task; ss = scaled score; SR = Sentence Repetition; CELF Preschool-2 = Clinical Evaluation of Language Fundamentals Preschool–Second Edition; EP = Estructura de Palabras; RO = Recordando Oraciones; CB = Conceptos Basicos; SS = Sentence Structure; WS = Word Structure; EV = Expressive Vocabulary; MLUw = mean length of utterance in words; NDW = number of different words; PGU = percentage of grammatical utterances.

Language sample information revealed that the child's strongest language in a functional environment was Spanish. She was more grammatical in Spanish and had a higher NDW in Spanish than in English, indicating a larger and more diverse vocabulary in Spanish. In Spanish, her strongest language, she produced utterances that were shorter in length in comparison to her peers (−1.75 *SD*s from the comparison group) and more ungrammatical than utterances from typically developing children at age 6;1. Her percentage of grammatical utterances in her strongest language was below the cutoff score of 80% that has been used for the identification of language disorders in bilingual children (Restrepo, 1998). Therefore, the evidence from the retell language sample points in the direction of the presence of a language disorder.

In regard to standardized testing, we have two sources of information: We have data from the BESA (Peña et al., 2018) and data from the CELF Preschool-2 in Spanish (Semel et al., 2009) and English (Semel et al., 2004). Because the BESA was standardized with bilingual children and the girl in this case study was bilingual, the information from the BESA represents the best available data from a standardized assessment tool. The BESA Morphosyntax is reported to have a sensitivity of 88.9 and a specificity of 88.2 for 6-year-olds in the testing manual. In addition, external review of the diagnostic accuracy of the BESA suggested that the BESA had a sensitivity of 93.3 and a specificity of 86.7 (Lazewnik et al., 2019); therefore, we consider that this test was adequate for the identification of language disorders in this case study. The girl obtained a Morphosyntax BESA standard score in Spanish of 83 and in English of 58. The score of 83 in Spanish represented her best language. This information aligned with previous information about naturalistic language that also suggested that Spanish was the strongest language for this child. The cutoff score for identification using the Morphosyntax score in the BESA is 81 for a 6-year-old, and the 95% confidence interval for a 6-year-old in Morphosyntax is 6.57. Therefore, we can be 95% confident that the score of the child is between 76.43 and 89.57. Importantly, this range of scores included the cutoff score; therefore, we interpret this information with caution. Because we have information available from another standardized test, we move to examine the girls' performance on the CELF Preschool-2 in Spanish and English. The CELF Preschool-2 in English was standardized with monolingual English speakers; therefore, the English scores did not truly represent the abilities of the case study in question. However, we see that the scores of English are lower than the scores of Spanish. The CELF Preschool-2 in Spanish was standardized with Spanish speakers in the United States, so the scores from this test are more representative of the language abilities of this child in comparison with the English scores from the BESA. The sensitivity and specificity of this test were reported to be 86 and 89, respectively, in the testing manual when using a cutoff score of 85 (Semel et al., 2009). However, external examination of its diagnostic accuracy was lower. Lazenwiski et al. (2019) reported that the CELF Preschool-2 had a sensitivity of 85.7 and a specificity of 64.3 when Spanish scores were interpreted as suggested in the manual. If we take into account the LRs (positive LR: 2.4; negative LR = 0.22), we see that the CELF Preschool-2 is clinically informative to rule out the presence of a disorder, but not to identify a child with a language disorder (Dollaghan, 2007). So, we conclude that the scores of the CELF Preschool-2 are not clinically informative for identification of a language disorder in this case study. The information from standardized testing is suggestive of a disorder but inconclusive.

In this case study, we conclude that the child has a language disorder, as evidenced by the concerns of parents and teacher and the performance in her retell samples. The evidence from the retell language samples indicated poor language abilities in both Spanish and English, when compared to other age-matched bilingual children. The child produced utterances that were shorter, that were more ungrammatical, and with less varied vocabulary than other children in the same age group.

Case Study 2

Our second case study is a girl who was age 6;4 at the time of the assessment. She was a participant in a research study to determine the utility of a range of diagnostic measures in determining if bilingual children have language disorders. At the time of assessment, she was in kindergarten and was age 6;4. For the overall assessment protocol, children (as well as families and teachers) completed a range of measures in Spanish and English that included parent and teacher interviews, standardized tests that tap into Spanish and English knowledge including the BESA, language samples using the tell–retell protocol comparable to that laid out in the SALT manual, and DA of narratives. Data on some of these measures have been presented in earlier work (Gillam et al., 2013). Here, we selected representative measures in each of the quadrants to illustrate the process by which we gathered converging evidence about the speech and language skills of this child.

Table 3 provides a summary of the measures and outcomes. Parents and teachers completed the Bilingual Input–Output Survey and the ITALK from the BESA (Peña et al., 2018). The Bilingual Input–Output Survey showed that, in combination, home and school use of English and Spanish exceeded 30%, so assessment in both languages for the purpose of interpreting the BESA scores was necessary. The ITALK questions focused on the child's knowledge in the domains of vocabulary knowledge, speech intelligibility, language comprehension, and sentence structure and grammaticality. In this case, teacher ratings were lowest for the grammaticality. Parents reported that sentences were appropriately long but that speech was sometimes hard to understand and that her sentences were sometimes ungrammatical. Taken together, these scores are indicative of concern.

A look across language scores from the BESA, narrative retell LSA following the SALT protocol, and DA of narratives provides converging evidence of language learning difficulties especially with regard to morphosyntax. Combined data from the BESA Syntax and Semantics subtests show best language scores (higher in Spanish for both domains) yield a standard score of 81. This low standard score is driven by the low score on the Morphosyntax subtest. Performance on the Semantics subtest was a strength.

Language production measures based on the narrative retells show that MLUw was higher in English, but MLUw in both languages was well below the mean for age-matched peers, which was consistent with parent and teacher report of some ungrammatical utterances. Grammaticality is quite low—a cutoff of 80% is often recommended as indicated above, and these scores were well below that cutoff. As in the BESA, when compared to children who produce the same MLUw, the child's NDW was above the expected range, reflecting the semantic strengths observed in the BESA. It is always important to ensure that low scores are the result of language disorders, and observation of child learning is an excellent tool in this regard. DA of narratives in English was used for the purposes of this observation. The child told a narrative in English and then was provided with a scripted narrative intervention to determine the extent to which she was able to modify her productions. In this case, some scores were high, indicating

Table 3. Summary of assessment results for Case Study 2.

Parent and teacher ratings	**Standardized testing**											
BIOS	BESA											
• Home 74% Sp/26% En		Phonology		Morphosyntax				Semantics				
• School 61% Sp/39% En		Sp	En	Sp		En		Sp		En		
ITALK	Scaled	2	3	2	5	3	< 1	11	6	8	8	
• Home score 3.8/5	Summed			7		3		17		16		
• School score 3.8/5	Standard	60	**65**	**68**		58		**93**		90		
	Best score	65		68				93				
	Language index	81										
Language sample	**Observation of learning**											
LSA retell (± 2 mos. age match; NTW length match)	Dynamic assessment of narratives											
Spanish	Story components (scored 1–5)						Pre		Post			
• MLUw = 2.77 (–3.49 *SD*)	• Setting (time & place)						1		2			
• NDW = 55 (1.76 *SD*)	• Character information						1		1			
• PGU = 28%	• Temporal order of events						2		2			
English	• Causal relationships						1		1			
• MLUw = 3.14 (–3.6 *SD*)	Story ideas and language (scored 1–5)											
• NDW = 63 (2.43 *SD*)	• Complexity of ideas						2		2			
• PGU = 13%	• Complexity of vocabulary						1		1			
	• Grammatical complexity						4		5			
	• Knowledge of dialogue						2		5			
	• Creativity						1		1			
	Episode structure (scored 1–7)						1		2			

Note. Boldface data indicate language with higher score. BIOS = Bilingual Input–Output Survey; Sp = Spanish; En = English; ITALK = Inventory to Assess Language Knowledge; BESA = Bilingual English–Spanish Assessment; LSA = language sample analysis; NTW = number total words; MLUw = mean length of utterance in words; NDW = number of different words; PGU = percentage of grammatical utterances.

that she was able to employ models of story elements to her advantage. For example, she scored 4 and 5 in the area of grammatical complexity, which reflects her ability to add verbs to her utterances (rather than add morphological elements as was evaluated in the BESA and the LSA). In contrast, she scored in the 1–2 range on her use of character (i.e., character names) and setting information (e.g., using prepositional phrases to indicate location), suggesting that she was not able to respond to instruction in these areas. Her modifiability ratings were good indicating that she was using multiple strategies, even when tasks were difficult.

The other area of note in the parent report was speech intelligibility. The BESA and the language sample data provide evidence upon which to make clinical decisions. The BESA Phonology scores are both low, but English is the higher score. Still, it is well below the expected range as demonstrated by her standard score of 63. LSA also provided an estimate of intelligibility based on her proportion of intelligible words and utterances. For connected speech, intelligibility was higher in Spanish than English. While her utterance level estimate of 96.5% intelligible was almost 3 *SD*s below the sample mean, it is still a high level of intelligibility and may not be clinically significant. Here, there appears to be a lack of convergence between the indicators for possible articulation or speech difficulties. One possible reason for this lack of convergence is that she used many short utterances in connected speech, and so this estimate of intelligibility may not reflect her overall intelligibility when she uses more complex language. To make a more informed conclusion about her overall speech intelligibility, it may be helpful to collect information about her responsiveness to intervention around speech production.

Overall, it can be seen that employing the full range of measures provides a rich picture of the child's performance in tasks that range in degree of structure. Parent and teacher ratings, BESA performance, and language samples pointed to relative strength in word knowledge but challenges in language form (morphosyntax and phonology). Language sample data provided converging evidence of low grammaticality but not of low intelligibility. The DA of narratives showed that she could respond to input. Her responses on the narrative task indicated that she was able to make some changes to support morphosyntactic learning (i.e., inclusion of verbs). Therefore, we concluded that she presented with a language disorder and that she was eligible for services.

Conclusion

Both ASHA guidelines for professional practice and the Individuals with Disabilities Education Act regulation support the use of converging evidence for the assessment and diagnosis of DLLs. In this tutorial, we presented information on language experience questionnaires, language sample analyses, language learning potential, and standardized testing as tools to collect valid language data for DLLs. It is important that the converging use of all measures described in this study is assessed in future research. We advocate for the use of a converging evidence approach for clinical diagnosis in DLLs. We encourage clinicians to move away from diagnostic processes and policies that over-rely on scores from standardized assessments and, instead, use a converging evidence approach in which valid assessment data are incorporated to make a diagnostic decision.

Acknowledgments

Research reported in this publication was supported by National Institute on Deafness and Other Communication Disorders Grants K23DC015835 and R15DC013670, awarded to Anny Castilla-Earls. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

American Speech-Language-Hearing Association. (2016). *Code of ethics*. http://www.asha.org/policy/

American Speech-Language-Hearing Association. (2019). *Bilingual service delivery*. http://www.asha.org/Practice-Portal/Professional-Issues/Bilingual-Service-Delivery/

Anaya, J. B., Peña, E. D., & Bedore, L. M. (2018). Conceptual scoring of expressive vocabulary: A discriminant functions analysis. *Language, Speech, and Hearing Services in Schools, 49*(1), 85–97. https://doi.org/10.1044/2017_LSHSS-16-0081

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

Barragan, B., Castilla-Earls, A., Martinez-Nieto, L., Restrepo, M. A., & Gray, S. (2018). Performance of low-income dual language learners attending English-only schools on the Clinical Evaluation of Language Fundamentals–Fourth Edition, Spanish. *Language, Speech, and Hearing Services in Schools, 49*(2), 292–305. https://doi.org/10.1044/2017_LSHSS-17-0013

Bedore, L. M., Peña, E. D., Anaya, J. B., Nieto, R., Lugo-Neris, M. J., & Baron, A. (2018). Understanding disorder within variation: Production of English grammatical forms by English language learners. *Language, Speech, and Hearing Services in Schools, 49*(2), 277–291. https://doi.org/10.1044/2017_LSHSS-17-0027

Bedore, L. M., Peña, E. D., Gillam, R., & Ho, T. (2010). Language sample measures and language ability in Spanish–English bilingual kindergarteners. *Journal of Communication Disorders, 43*(6), 498–510. https://doi.org/10.1016/j.jcomdis.2010.05.002

Bedore, L. M., Peña, E. D., Griffin, Z. M., & Hixon, J. G. (2016). Effects of Age of English Exposure, Current Input/Output, and grade on bilingual language performance. *Journal of Child Language, 43*(3), 687–706. https://doi.org/10.1017/S0305000915000811

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

Bedore, L. M., Peña, E. D., Summers, C., Boerger, K., Greene, K., Resendiz, M., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–

English bilingual prekindergarten students. *Bilingualism: Language and Cognition, 15*(3), 616–629. https://doi.org/10.1017/S1366728912000090

Bohman, T. M., Bedore, L. M., Peña, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: Language performance in Spanish–English bilinguals. *International Journal of Bilingual Education and Bilingualism, 13*(3), 325–344. https://doi.org/10.1080/13670050903342019

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2013). Measuring test measurement error: A general approach. *Journal of Educational and Behavioral Statistics, 38*(6), 629–663. https://doi.org/10.3102/1076998613508584

Castilla-Earls, A., Auza, A., Pérez-Leroux, A. T., Fulcher-Rood, K., & Barr, C. (2020). Morphological errors in monolingual Spanish-speaking children with and without developmental language disorders. *Language, Speech, and Hearing Services in the Schools, 51*(2), 270–281. https://doi.org/10.1044/2019_LSHSS-19-00022

Castilla-Earls, A., Francis, D., Iglesias, A., & Davidson, K. (2019). The impact of the Spanish-to-English proficiency shift on the grammaticality of English learners. *Journal of Speech, Language, and Hearing Research, 62*(6), 1739–1754. https://doi.org/10.1044/2018_JSLHR-L-18-0324

Cohen, J. F., Korevaar, D. A., Altman, D. G., Bruns, D. E., Gatsonis, C. A., Hooft, L., Irwig, L., Levine, D., Reitsma, J. B., de Vet, H. C., & Bossuyt, P. M. (2016). STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. *BMJ Open, 6*(11), Article e012799. https://doi.org/10.1136/bmjopen-2016-012799

Dollaghan, C. A. (2004). Evidence-based practice in communication disorders: What do we know, and when do we know it? *Journal of Communication Disorders, 37*(5), 391–400. https://doi.org/10.1016/j.jcomdis.2004.04.002

Dollaghan, C. A. (2007). *The handbook of evidence-based practice in communication disorders*. Brookes.

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research, 54*(4), 1077–1088. https://doi.org/10.1044/1092-4388(2010/10-0093)

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*(1), 42–55. https://doi.org/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*(4), 337–350. https://doi.org/10.1044/2014_LSHSS-14-0034

Eligibility for Special Education Under a Specific Learning Disability Classification [PDF]. (2019). https://www.ncld.org/wp-content/uploads/2019/07/Eligibility-for-Special-Education-under-a-Specific-Learning-Disability-Classification-Final.pdf

Fabiano-Smith, L. (2019). Standardized tests and the diagnosis of speech sound disorders. *SIG 1 Perspectives of the ASHA Special Interest Groups, 4*(1), 58–66. https://doi.org/10.1044/2018_PERS-SIG1-2018-0018

Fenson, L., Marchman, V. A., Thal, D., Dale, P. S., Reznick, J. S., & Bates, E. (2006). *MacArthur Bates Communicative Development Inventories*. Brookes. https://doi.org/10.1037/t11538-000

Fuchs, D., & Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice, 22*(2), 129–136. https://doi.org/10.1111/j.1540-5826.2007.00237.x

Fulcher-Rood, K., Castilla-Earls, A. P., & Higginbotham, J. (2018). School-based speech-language pathologists' perspectives on diagnostic decision making. *American Journal of Speech-Language Pathology, 27*(2), 796–812. https://doi.org/10.1044/2018_ajslp-16-0121

Fulcher-Rood, K., Castilla-Earls, A. P., & Higginbotham, J. (2019). Diagnostic decisions in child language assessment: Findings from a case review assessment task. *Language, Speech, and Hearing Services in Schools, 50*(3), 385–398. https://doi.org/10.1044/2019_LSHSS-18-0044

Genesee, F., Paradis, J., & Crago, M. (2004). *Dual language development and disorders*. Brookes.

Gillam, R. B., Peña, E. D., Bedore, L. M., Bohman, T. M., & Pérez, A. M. (2013). Identification of specific language impairment in bilingual children, Part 1: Assessment in English. *Journal of Speech, Language, and Hearing Research, 56*(6), 1813–1823. https://doi.org/10.1044/1092-4388(2013/12-0056)

Glaspey, A. (2018). *The Glaspey dynamic assessment of phonology (GDAP)*. Academic Therapy Publications.

Glaspey, A., & Stoel-Gammon, C. (2007). A dynamic approach to phonological assessment. *Advances in Speech Language Pathology, 9*(4), 286–296. https://doi.org/10.1080/14417040701435418

Goldstein, B. A. (2006). Clinical implications of research on language development and disorders in bilingual children. *Topics in Language Disorders, 26*(4), 305–321. https://doi.org/10.1097/00011363-200610000-00004

Goldstein, B. A. (2012). *Bilingual language development and disorders in Spanish–English speakers* (2nd ed.). Brookes.

Guiberson, M., & Atkins, J. (2012). Speech-language pathologists' preparation, practices, and perspectives on serving culturally and linguistically diverse children. *Communication Disorders Quarterly, 33*(3), 169–180. https://doi.org/10.1177/1525740110384132

Gusewski, S., & Rojas, R. (2017). Tense marking in the English narrative retells of dual language preschoolers. *Language, Speech, and Hearing Services in Schools, 48*(3), 183–196. https://doi.org/10.1044/2017_LSHSS-16-0093

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics, 24*(2), 267–288. https://doi.org/10.1017/S0142716403000158

Gutiérrez-Clellen, V. F., Restrepo, M. A., Bedore, L., Peña, E., & Anderson, E. R. (2001). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools, 31*(1), 88–98. https://doi.org/10.1044/0161-1461.3101.88

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2007). The discriminant accuracy of a grammatical measure with Latino English-speaking children. *Journal of Speech, Language, and Hearing Research, 50*(4), 968–981. https://doi.org/10.1044/1092-4388(2007/068)

Hammer, C. S., Detwiler, J. S., Detwiler, J., Blood, G. W., & Qualls, C. D. (2004). Speech-language pathologists' training and confidence in serving Spanish–English bilingual children. *Journal of Communication Disorders, 37*(2), 91–108. https://doi.org/10.1016/j.jcomdis.2003.07.002

Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice: Clinical and educational applications*. Cambridge University Press.

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*(2), 178–188. https://doi.org/10.1097/01.TLD.0000318937.39301.76

Heilmann, J. J., Miller, J. F., & Nockerts, A. (2010). Using language sample databases. *Language, Speech, and Hearing Services in Schools, 41*(1), 84–95. https://doi.org/10.1044/0161-1461(2009/08-0075)

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*(3), 339–345. https://doi.org/10.1111/1460-6984.12201

Henderson, D., Restrepo, M. A., & Aiken, L. (2018). Dynamic assessment of narratives among Navajo preschoolers. *Journal of Speech, Language, and Hearing Research, 61*(10), 2547–2560. https://doi.org/10.1044/2018_JSLHR-L-17-0313

Huang, R., Hopkins, J., & Nippold, M. A. (1997). Satisfaction with standardized language testing. *Language, Speech, and Hearing Services in Schools, 28*(1), 12–29. https://doi.org/10.1044/0161-1461.2801.12

Kapantzoglou, M., Restrepo, M. A., & Thompson, M. S. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in Schools, 43*(1), 81–96. https://doi.org/10.1044/0161-1461(2011/10-0095)

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., Zhang, J., Rathbun, A., Wilkinson-Flicker, S., Diliberti, M., Barmer, A., Bullock Mann, F., & Dunlop Velez, E. (2016). *The condition of education 2016 (NCES 2016-144)*. U.S. Department of Education, National Center for Education Statistics. http://nces.ed.gov/pubsearch

Kramer, K., Mallett, P., Schneider, P., & Hayward, D. (2009). Dynamic assessment of narratives with Grade 3 children in a First Nations community. *Canadian Journal of Speech-Language Pathology and Audiology, 33*(3), 119–128.

Kohnert, K., & Medina, A. (2009). Bilingual children and communication disorders: A 30-year research retrospective. *Seminars in Speech and Language, 30*(4), 219–233. https://doi.org/10.1055/s-0029-1241721

Lazewnik, R., Creaghead, N., Smith, A., Predenville, L., Raisor-Becker, L., & Silbert, N. (2019). Identifiers of language impairments for Spanish–English dual language learners. *Language, Speech, and Hearing Services in Schools, 50*(1), 126–137. https://doi.org/10.1044/2018_LSHSS-17-0046

Lidz, C. S. (1987). *Dynamic assessment: An interactional approach to evaluating learning potential*. Guilford Press.

Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment*. Guilford Press.

Lidz, C. S., & Peña, E. D. (2009). Response to intervention and dynamic assessment: Do we just appear to be speaking the same language? *Seminars in Speech and Language, 30*(2), 121–134. https://doi.org/10.1055/s-0029-1215719

Mann, W., Peña, E. D., & Morgan, G. (2015). Child modifiability as a predictor of language abilities in deaf children who use American Sign Language. *American Journal of Speech-Language Pathology, 24*(3), 374–385. https://doi.org/10.1044/2015_AJSLP-14-0072

Martin, N., & Brownell, R. (2013). *Expressive One-Word Picture Vocabulary Test–4: Spanish-Bilingual Edition (EOWPVT-4 SBE)*. Academic Therapy Publications.

Mayer, M. (1969). *Frog, where are you?* The Dial Press.

Mayer, M. (1973). *Frog on his own*. The Dial Press.

Mayer, M. (1974). *Frog goes to dinner*. The Dial Press.

McCauley, R., & Swisher, L. (1994). Psychometric review of language and articulations tests for preschool children. *Journal of Speech and Hearing Disorders, 49*(1), 34–42. https://doi.org/10.1044/jshd.4901.34

Miller, J. F., Andriacchi, K., & Nockerts, A. (2019). *Assessing language production using SALT software: A clinician's guide to language sample analysis* (3rd ed.). SALT Software.

Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research & Practice, 21*(1), 30–43. https://doi.org/10.1111/j.1540-5826.2006.00205.x

Miller, J. F., & Iglesias, A. (2019). *Systematic Analysis of Language Transcripts (SALT): Version 20* [Computer software]. SALT Software.

Muñoz, M. L., Gillam, R. B., Peña, E. B., & Gulley-Faehnle, A. (2003). Measures of language development in fictional narratives of Latino children. *Language, Speech, and Hearing Services in Schools, 34*(4), 332–342. https://doi.org/10.1044/0161-1461(2003/027)

Paradis, J., & Crago, M. (2000). Tense and temporality: A comparison between children learning a second language and children with SLI. *Journal of Speech, Language and Hearing Research, 43*(4), 834–848. https://doi.org/10.1044/jslhr.4304.834

Paradis, J., Crago, M., Genesee, F., & Rice, M. (2003). French–English bilingual children with specific language impairment: How do they compare with their monolingual peers? *Journal of Speech, Language, and Hearing Research, 46*(1), 113–127. https://doi.org/10.1044/1092-4388(2003/009)

Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders, 43*(6), 474–497. https://doi.org/10.1016/j.jcomdis.2010.01.002

Paradis, J., Rice, M., Crago, M., & Marquis, J. (2008). The acquisition of tense in English: Distinguishing child second language from first language and specific language impairment. *Applied Psycholinguistics, 29*(4), 689–722. https://doi.org/10.1017/S0142716408080296

Potapova, I., Kelly, S., Combiths, P., & Pruitt-Lord, S. (2018). Evaluating English morpheme accuracy, diversity, and productivity measures in language samples of developing bilinguals. *Language, Speech, and Hearing Services in Schools, 49*(2), 260–276. https://doi.org/10.1044/2017_LSHSS-17-0026

Paul, R., & Norbury, C. F. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (4th ed.). Elsevier.

Pavelko, S. L., Owens, R. E., Ireland, M., & Hahs-Vaughn, D. L. (2016). Use of language sample analysis by school-based SLPs: Results of a nationwide survey. *Language, Speech, and Hearing Services in Schools, 47*(3), 246–258. https://doi.org/10.1044/2016_LSHSS-15-0044

Peña, E. D., Bedore, L. M., & Krester, E. S. (2015). Discriminant accuracy of a semantics measure with Latino English-speaking, Spanish-speaking, and English–Spanish bilingual children. *Journal of Communication Disorders, 53,* 30–41. https://doi.org/10.1016/j.jcomdis.2014.11.001

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*(6), 2208–2220. https://doi.org/10.1044/2014_JSLHR-L-13-0151

Peña, E. D., Gillam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research, 49*(5), 1037–1057. https://doi.org/10.1044/1092-4388(2006/074)

Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2018). *Bilingual English–Spanish Assessment (BESA)*. Brookes.

Peña, E. D., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology, 10*(2), 138–154. https://doi.org/10.1044/1058-0360(2001/014)

Peña, E. D., & Méndez-Pérez, A. (2006). Individualistic and collectivistic approaches to language learning. *Zero to Three (J), 27*(1), 34–45.

Peña, E. D., Reséndiz, M., & Gillam, R. B. (2007). The role of clinical judgements of modifiability in the diagnosis of language impairment. *International Journal of Speech-Language Pathology, 9*(4), 332–345. https://doi.org/10.1080/14417040701413738

Petersen, D. B., Chanthongthip, H., Ukrainetz, T. A., Spencer, T. D., & Steeve, R. W. (2017). Dynamic assessment of narratives: Efficient, accurate identification of language impairment in bilingual students. *Journal of Speech, Language, and Hearing Research, 60*(4), 983–998. https://doi.org/10.1044/2016_JSLHR-L-15-0426

Petersen, D. B., & Spencer, T. D. (2014). *Predictive early assessment of reading and learning (PEARL)*. Language Dynamics Group.

Pew Research Center. (n.d.). *Key facts about refugees to the U.S.* Retrieved November 5, 2019, from https://www.pewresearch.org/fact-tank/2019/10/07/key-facts-about-refugees-to-the-u-s/

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development, 82,* 1834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x

Plante, E., & Vance, R. (1994). Selection of preschool language tests. *Language, Speech, and Hearing Services in Schools, 25*(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15

Pratt, A., Adams, A., Peña, E. D., & Bedore, L. M. (in press). Exploring the use of parent and teacher questionnaires to screen for language disorders in bilingual children, topics in early childhood special education.

Potopova, I., Kelly, S., Combiths, P., & Pruitt-Lord, S. (2018). Evaluating English morpheme accuracy, diversity, and productivity measures in language samples of developing bilinguals. *Language, Speech, and Hearing Services in Schools, 49,* 260–276. https://doi.org/10.1044/2017_LSHSS-17-0026

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

Restrepo, M. A., Castilla, A. P., Schwanenflugel, P., Neuhart Pritchett, S., Hamilton, C., & Arboleda, A. (2010). Sentence length, complexity and grammaticality growth in Spanish-speaking children attending English-only and bilingual preschool programs. *Language, Speech, and Hearing Services in Schools, 41,* 3–13. https://doi.org/10.1044/0161-1461(2009/06-0017)

Restrepo, M. A., Morgan, G., & Thompson, M. (2013). Vocabulary intervention in bilingual preschoolers with language impairment. *Journal of Speech, Language, and Hearing Research, 56,* 748–765. https://doi.org/10.1044/1092-4388(2012/11-0173)

Restrepo, M. A., & Silverman, S. (2001). The validity of the Spanish Preschool Language Scale for use with bilingual children. *American Journal of Speech-Language Pathology, 10,* 382–393. https://doi.org/10.1044/1058-0360(2001/032)

Restrepo, M. A., Spencer, T., Thompson, M., & Petersen, D. (2018, June). *Does CELF-P English and Spanish predict performance on response to intervention?* [Conference session]. BiSLI Conference, Reading, England.

Rhoades, K., & Madaus, G. (2003). *Errors in standardized tests: A systemic problem*. Boston College.

Rice, M. L., & Wexler, K. (2001). *Test of Early Grammatical Impairment*. The Psychological Corporation.

Rojas, R., Hiebert, L., Gusewski, S., & Francis, D. J. (2019). Moving forward by looking back: Understanding why some Spanish-speaking English learners fall behind. *New Directions for Child and Adolescent Development, 166,* 43–77. https://doi.org/10.1002/cad.20305

Rojas, R., & Iglesias, A. (2009). Making a case for language sampling. *The ASHA Leader, 14*(3), 10–13. https://doi.org/10.1044/leader.FTR1.14032009.10

Rojas, R., & Iglesias, A. (2013). The language growth of Spanish-speaking English language learners. *Child Development, 84*(2), 630–646. https://doi.org/10.1111/j.1467-8624.2012.01871.x

Rojas, R., & Iglesias, A. (2019). Assessing the bilingual (Spanish/English) population. In J. F. Miller, K. Andriacchi, & A. Nockerts (Eds.), *Assessing language production using SALT software: A clinician's guide to language sample analysis* (3rd ed., pp. 115–127). SALT Software.

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. https://doi.org/10.3109/17549507.2015.1081290

Semel, E., Wiig, E. H., & Secord, W. A. (2004). *Clinical Evaluation of Language Fundamentals-Preschool-2 (CELF-Preschool-2)*. Pearson.

Semel, E., Wiig, E. H., & Secord, W. A. (2009). *Clinical Evaluation of Language Fundamentals-Preschool-2 Spanish (CELF-Preschool-2 Spanish)*. Pearson.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders, 47*(3), 256–270. https://doi.org/10.1044/jshd.4703.256

Spaulding, T., Plante, E., & Farinella, K. (2006). Eligibility criteria for language impairment. *Language, Speech, and Hearing Services in Schools, 37*(1), 61–72. https://doi.org/10.1044/0161-1461(2006/007)

Spencer, T. D., & Petersen, D. B. (2012). *Story Champs*. Language Dynamics Group.

Spencer, T. D., Petersen, D. B., & Adams, J. L. (2015). Tier 2 language intervention for diverse preschoolers: An early-stage randomized control group study following an analysis of response to intervention. *American Journal of Speech-Language Pathology, 24*(4), 619–636. https://doi.org/10.1044/2015_AJSLP-14-0101

Stertzbach, J., & Gildersleeve-Neumann, C. (2005, November). *Parent report as a screening tool of speech disorders in Spanish-speaking preschool children* (Unpublished manuscript). Technical presentation presented at the annual convention of the American Speech-Language-Hearing Association, San Diego, CA.

Thompson, C. A., Craig, H. K., & Washington, J. A. (2004). Variable production of African American English across oracy and literacy contexts. *Language, Speech, and Hearing Services in Schools, 35,* 269–282. https://doi.org/10.1044/0161-1461(2004/025)

Troia, G. A. (2005). Writing instruction for students with learning disabilities. In A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 129–156). Guilford.

Ukrainetz, T. A., Harpell, S., Walsh, C., & Coyle, C. (2000). A preliminary investigation of dynamic assessment with Native American kindergartners. *Language, Speech, and Hearing Services in Schools, 31*(2), 142–154. https://doi.org/10.1044/0161-1461.3102.142

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics, 35*(04), 765–805. https://doi.org/10.1017/S0142716412000574

Whittaker, M., & Batsche, G. (n.d.). *Data-based problem solving: Effective implementation of MTSS, RTI, and PBIS*. National Center for Learning Disabilities. https://www.ncld.org/wp-content/uploads/2019/09/Essential-Components-of-Data-Based-Problem-Solving-Approaches.pdf

Whittaker, M., & Burns, M. (n.d.). *Evaluation for specific learning disabilities: Allowable methods of identification & their implications* [PDF file]. https://www.ncld.org/wp-content/uploads/2019/09/Evaluation-for-Specific-Learning-Disabilities-Allowable-Methods-of-Identification-Their-Implications.pdf

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2012). *Preschool Language Scale–Fourth Edition (PLS-5) Spanish Edition*. Harcourt. https://doi.org/10.1037/t15141-000