Research Article

Including Nonlinguistic Processing Tasks in the Identification of Developmental Language Disorder

Kerry Danahy Ebert^a and Giang Pham^b

Purpose: Accurate identification of developmental language disorder (DLD) remains challenging, particularly for children who speak different dialects, languages, or more than 1 language. Children with DLD, on average, have shown subtle deficits on nonlinguistic cognitive processing tasks, and performance on such tasks may be minimally influenced by language experience. This study explores whether nonlinguistic cognitive processing tasks can contribute to the identification of DLD in children from diverse linguistic backgrounds.

Method: Study 1 combined data from 4 U.S.-based investigations to yield a sample of 395 children, ages 6–10 years, who spoke only English or both Spanish and English. Study 2 consisted of an international sample of 55 kindergarten children living in Vietnam. Each study included children with DLD and children with typical development. Participants completed nonlinguistic cognitive

tasks of processing speed, auditory working memory, and attentional control. Data analysis compared typically developing to DLD groups by age and language background. Then, we empirically derived cut-points to report diagnostic accuracy (sensitivity, specificity, and positive and negative likelihood ratios).

Results: For all 3 tasks, adequate sensitivity or specificity (but not both in most cases) was achieved in nearly all age groups. Likelihood ratios reached moderately to very informative levels in several instances. Diagnostic results were maintained when monolingual and bilingual samples were combined into a single group.

Conclusions: Nonlinguistic cognitive processing tasks may contribute to accurate identification of DLD in combination with other measures. Further research is needed to refine tasks, confirm cut-points established here, and extend findings to children from additional language backgrounds.

Developmental language disorder (DLD) is defined as difficulty with the acquisition of language skills in the absence of an obvious cause (Bishop, Snowling, Thompson, Greenhalgh, & CATALISE-2 Consortium, 2017). As such, the overt symptoms of the disorder vary according to the linguistic environment and the social and academic expectations for a child. An obvious example is that children learning English will show delays in the

acquisition of English grammatical structures and vocabulary words, whereas children learning Italian will exhibit problems learning aspects of Italian grammar and vocabulary (Leonard, 2014). Even within a language, however, there is variation in the overt symptoms of DLD; for example, omission of third-person present tense singular verb marking may reliably distinguish DLD among children who speak the General American English dialect (Ash & Redmond, 2014), but the same characteristic is not reliable in children who speak another American English dialect, African American English (Cleveland & Oetting, 2013). Children exposed to multiple languages add to the variability, as reduced exposure to one language may result in language skills that are indistinguishable from those of children with DLD (e.g., Kohnert, Windsor, & Ebert, 2009; Verhoeven, Steenge, van Weerdenburg, & van Balkom, 2011).

The variability in symptomatology complicates efforts to accurately identify DLD. Different approaches to this problem have been proposed, and we begin this article by reviewing advantages and disadvantages to these approaches. We then review evidence of nonlinguistic processing deficits

^aDepartment of Communication Disorders and Sciences, Rush University, Chicago, IL

^bSchool of Speech, Language, and Hearing Sciences, San Diego State University, CA

Correspondence to Kerry Danahy Ebert, who is now with the Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis: kebert@umn.edu

Editor-in-Chief: Julie Barkmeier-Kraemer

Editor: Mark Guiberson

Received September 11, 2018

Revision received February 18, 2019

Accepted April 2, 2019

https://doi.org/10.1044/2019_AJSLP-IDLL-18-0208

Publisher Note: This article is part of the Forum: Innovations in Clinical Practice for Dual Language Learners, Part 1.

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

in children with DLD in order to motivate the primary purpose of this study: to explore the value of including non-linguistic cognitive processing tasks in the identification of DLD in children from diverse linguistic backgrounds.

Approaches to Identification

A complete language assessment for DLD encompasses multiple purposes—such as identifying the presence or absence of a disorder, characterizing strengths and weaknesses, and setting goals for treatment. Here, we focus on the purpose of identification, specifically across linguistically diverse groups of children. Even within this restricted purpose, many different approaches have been explored. This section reviews strengths and weaknesses of several approaches to identification of DLD, including standardized tests, grammatical markers from language samples, and language-based processing tasks.

One of the most common approaches to the identification of DLD among speech-language pathologists is the use of omnibus norm-referenced language testing (e.g., Betz, Eickoff, & Sullivan, 2013). The pitfalls of overreliance on this approach have been well established; even when tests are administered to monolingual speakers of the standard dialect in the language for which the test was originally designed, many tests demonstrate inadequate sensitivity and specificity (Plante & Vance, 1994). In fact, in a recent review of language assessments (Spaulding, Plante, & Farinella, 2006), just five of 43 tests reported sensitivity and specificity values of at least 80% each, the minimum criterion for acceptable accuracy suggested by Plante and Vance (1994). Although four of these five tests (i.e., the Clinical Evaluation of Language Fundamentals–Fourth Edition, the Test of Early Grammatical Impairment, the Test of Narrative Language, and the Test of Language Competence–Expanded) fit the age range of interest in this study (i.e., school-age children), all four tests were designed for use with monolingual speakers of the General American English dialect. When the child to be assessed does not match the normative reference group for the test (e.g., a child who speaks a different dialect or language or more than one language), the test becomes invalid for identification (Bedore & Peña, 2008; Rodekohr & Haynes, 2001). It must be acknowledged, however, that efforts to improve norm-referenced tests and to expand the breadth of populations to which they can be applied are ongoing; for example, a norm-referenced test has recently been developed for Spanish–English bilinguals (i.e., Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2018).

Another approach to identification of DLD is to look for key linguistic characteristics of the disorder that are specific to children with a common linguistic environment. As previously mentioned, the omission of tense and agreement morphemes distinguishes monolingual speakers of General American English with DLD from their unaffected peers, within the range of about 3–9 years of age (see Ash & Redmond, 2014, for a summary). These linguistic characteristics would typically be examined in a language sample, a task that is highly useful for other purposes of clinical assessment,

such as characterizing children's language skills and goal setting (e.g., Ebert & Pham, 2017). The approach of finding key characteristics in a language sample can also be extended to different linguistic circumstances; in other words, accurate identification could be accomplished by finding surface manifestations of DLD in each dialect, language, and multilingual environment (Oetting, 2018). There are many examples of success using this approach with specific populations. In Italian, production of definite singular articles and third-person plural inflections may distinguish children with DLD (e.g., Bortolini, Caselli, Deevy, & Leonard, 2002). In Spanish, articles and direct object pronouns appear problematic for children with DLD (e.g., Bedore & Leonard, 2005; Castilla-Earls et al., 2016).

Yet, the diversity of populations that must be studied remains a challenge. In particular, children who are exposed to more than one language present heterogeneous language profiles that may make it difficult to rely exclusively on specific grammatical characteristics for identification of DLD. For example, recent efforts to find grammatical markers in Spanish–English bilingual children have yielded some success but have also confirmed that the degree of exposure to each language affects the accuracy of the clinical markers (Bedore et al., 2018; Castilla-Earls et al., 2016).

Language-based processing tasks can be another component of an identification battery for DLD (e.g., Thordardottir & Brandeker, 2013). The premise of such tasks is that they emphasize the manipulation of linguistic material rather than knowledge of specific linguistic items, thus reducing the role of experience in task performance. The first language-based processing task to gain traction as an identifier of DLD is nonword repetition. This task has been shown to separate children with DLD from unaffected peers in speakers of General American English (Graf Estes, Evans, & Else-Quest, 2007), African American English (Rodekohr & Haynes, 2001), and a variety of languages other than English (e.g., Dispaldro, Leonard, & Deevy, 2013; Kapalková, Polišenská, & Vicenová, 2013; Pham, Ebert, Dinh, & Dam, 2018; Thordardottir & Brandeker, 2013). Sentence repetition has also received increasing attention as a language-based processing task with the ability to identify DLD. For example, a multiyear, international effort to identify tools for the diagnosis of DLD across languages (COST Action IS0804; Armon-Lotem, de Jong, & Meir, 2015) focused on the development of nonword repetition and sentence repetition tasks. Work resulting from this initiative has explored the ability of these two repetition tasks to identify DLD in Russian and Hebrew (Armon-Lotem & Meir, 2016) and in French (de Almeida et al., 2017).

However, it is clear that language-based processing tasks do not entirely circumvent the variability problem in the identification of DLD. For both nonword repetition and sentence repetition, language-specific stimuli must be developed and validated for each language and dialect. In addition, performance on these tasks is not entirely independent of language exposure. On average, sequential bilingual children score more poorly in their second language than monolingual children (de Almeida et al., 2017; Windsor,

Kohnert, Lobitz, & Pham, 2010), indicating that experience plays a role in performance. There is also evidence that dialect density (i.e., the frequency with which children use dialect-specific linguistic patterns) influences performance on nonword repetition (Moyle, Heilmann, & Finneran, 2014). As a result, several studies have shown a need to alter scoring systems or adjust cut-points on repetition tasks for children who speak nonmainstream dialects or more than one language (Armon-Lotem & Meir, 2016; de Almeida et al., 2017; Oetting, McDonald, Seidel, & Hegarty, 2016). Thus, although language-based processing tasks have shown promise for the identification of DLD, there is a substantial burden in developing, validating, and recalibrating tasks to suit the numerous linguistic backgrounds with which children present.

In this article, we consider another potential component of an identification battery for DLD that aims to be suitable across diverse linguistic circumstances. Instead of seeking language-specific characteristics that inherently vary across diverse linguistic experiences, this approach seeks to find universal features that can help identify DLD regardless of experience. To do so, we look outside the realm of language and consider the nonlinguistic deficits of children with DLD.

Nonlinguistic Processing Deficits in Children With DLD

By definition, children with DLD score broadly within the average range on standardized assessments of nonverbal intelligence. Yet, a robust literature indicates that, on average, children with DLD perform more poorly than children with typical language development on a host of tasks that have minimal linguistic content, such as mentally rotating shapes or sustaining attention to a stream of environmental noises (e.g., Ebert & Kohnert, 2011; Leonard et al., 2007; Miller, Kail, Leonard, & Tomblin, 2001). We refer here to such tasks as “nonlinguistic” because the stimuli are not language based and because the tasks themselves are designed to emphasize aspects of the cognitive processing system. However, we also acknowledge the reality that no task is completely nonlinguistic (Roebuck, Sindberg, & Ellis Weismer, 2018).

The subtle deficits found on nonlinguistic cognitive processing tasks among children with DLD have been organized and labeled in a number of different ways (e.g., Leonard, 2014; Ullman & Pierpont, 2005). Processing-based accounts of DLD (e.g., Leonard, 2014) hold that such information-processing deficits may contribute causally to the disorder; for example, subtly slower processing speed may impair the ability to process incoming linguistic information, particularly the least salient morphemes in a language. Here, we focus on two established areas of deficit—speed of processing and working memory (e.g., Ebert, Kohnert, Pham, Rentmeester Disher, & Payesteh, 2014; Leonard et al., 2007)—and one newly emerging area of potential deficit—attentional control (also referred to as *inhibition*; see Ebert, Rak, Slawny, & Fogg, 2019).

Speed of processing is an established cognitive skill that is expected to improve across childhood (Kail & Salthouse, 1994). Numerous studies have indicated that children with DLD demonstrate slower processing speed than unaffected peers, even on nonlinguistic tasks (e.g., Kohnert & Windsor, 2004; Leonard et al., 2007; Miller et al., 2001). For example, they are slower to mentally rotate shapes or to find a target shape within an array (Leonard et al., 2007; Miller et al., 2001). Even on a task as simple as detecting the appearance of a colored shape and discriminating between one of two colors, children with DLD, on average, respond more slowly (Kohnert & Windsor, 2004).

Like speed of processing, working memory improves through childhood (e.g., Zelazo, Blair, & Willoughby, 2016). Working memory is also a broad area of deficit for children with DLD, with extensive findings of working memory limitations for linguistic material (Montgomery, Magimairaj, & Finney, 2010). Meta-analysis has supported the extension of working memory deficits into the nonlinguistic domain (Vugs, Cuperus, Hendriks, & Verhoeven, 2013). Working memory assessments vary, but the most common models emphasize the simultaneous storage and processing of information (Montgomery et al., 2010). In the nonlinguistic domain, deficits associated with DLD are supported on a variety of visuospatial tasks (Vugs et al., 2013), as well as on pattern matching for auditory tone sequences (Yim, 2006).

Finally, emerging evidence suggests that children with DLD may show deficits in attentional control or the ability to tune out conflicting, irrelevant information to focus on relevant information. Tasks that assess this skill—such as flanker tasks, Simon tasks, and Stroop tasks—have also been classified as assessments of inhibition or inhibitory control (see also Ebert et al., 2019). Similar to processing speed and working memory, performance on attentional control improves with age in the childhood years (Zelazo et al., 2016). In Pauls and Archibald’s (2016) meta-analysis, they found evidence for a significant deficit in the area of inhibitory control for children with DLD in comparison to unaffected peers; both linguistic and nonlinguistic tasks were included in the effect, with no evidence that the linguistic content of the tasks explained differences among studies. In contrast, Yang and Gray (2017) found no difference between preschool children with and without DLD on a flanker task. Differences in participant ages or task parameters may have contributed to these mixed findings, and further research on the attentional control skills of children with DLD is needed.

In summary, there are numerous tasks with minimal linguistic content on which children with DLD show deficits at the group level. Because linguistic experience presumably has little effect on performance on these tasks, it is possible that including one or more nonlinguistic cognitive processing tasks could improve an identification battery for DLD. However, it is necessary to investigate the diagnostic accuracy of these tasks for identifying children with DLD. In other words, the existence of group deficits on nonlinguistic tasks does not necessarily mean they are suitable for identifying

individual children; instead, diagnostic accuracy analyses (such as sensitivity, specificity, and likelihood ratios [LRs]) are needed to explore their value in the identification of DLD (Dollaghan, 2007; Šimundić, 2008).

The Present Investigation

Despite substantial attention in the literature, accurate identification of DLD across linguistically diverse backgrounds continues to be challenging. Nonlinguistic tasks could potentially supplement existing approaches to identification but have not yet been fully explored in terms of diagnostic accuracy. The purpose of this study is to explore the ability of nonlinguistic cognitive processing tasks to identify school-age children with DLD across a range of linguistic backgrounds. We conduct rigorous diagnostic accuracy analyses (including sensitivity, specificity, and LR) in two separate studies that include diverse groups of children. Study 1 considers the ability of nonlinguistic cognitive processing tasks to discriminate between children with and without DLD in two samples based in the United States: English-only monolinguals ($n = 194$) and Spanish–English bilinguals ($n = 201$). We consider each group separately and then in combination to test whether diagnostic accuracy will hold for a mixed sample. The purpose is to determine whether these nonlinguistic cognitive processing tasks can maintain their accuracy across monolinguals and bilinguals, a diagnostic goal that is very challenging for language-based measures. In Study 2, we extend our analyses outside the United States by examining the ability of the same nonlinguistic tasks to identify DLD in a sample of monolingual Vietnamese-speaking children living in Vietnam ($n = 55$). Together, these studies provide a rigorous test of our hypothesis across diverse contexts.

Study 1

Study 1 examined whether nonlinguistic cognitive processing tasks could discriminate between children with and without DLD in monolingual children, bilingual children, and a combined sample of monolinguals and bilinguals.

Method

Participants

The data set for Study 1 was collected over the course of four prior investigations of school-age children with and without DLD (Ebert et al., 2014, 2019; Kohnert & Windsor, 2004; Windsor et al., 2010). These studies recruited both English-only monolinguals and Spanish–English bilinguals from urban areas in the north central United States. The data set consisted of children from these investigations who were aged 6;0–10;11 (years;months), spoke only English or both Spanish and English, and met criteria for DLD or for typical development. Table 1 shows the age range and language groups included in each of the original four investigations. Across all studies, disabilities and developmental

concerns outside DLD were ruled out. Children were required to pass a hearing screening and to score within the average range on a test of nonverbal intelligence. Parent report was used to exclude children with other disabilities, such as autism spectrum disorder, seizure disorder, or brain injuries.

There were a total of 194 English-only monolinguals. Seventy-three of these children were diagnosed with DLD based on a combination of language assessment measures and parent or teacher report. The remaining 121 children demonstrated language skills that were typically developing (TD; see original studies for specific eligibility criteria). Children with significant exposure to a language other than English at home or at school were excluded from the sample. Measures of socioeconomic status varied across the source studies. In two studies (Ebert et al., 2019; Kohnert & Windsor, 2004), maternal education was utilized, and this measure averaged “some college” for monolingual participants in both source studies. Participants in the remaining two studies (Ebert et al., 2014; Windsor et al., 2010) were recruited from a large urban school district in which the average percentage of students eligible for free or reduced lunch was 58.6% across elementary schools.

There were 201 Spanish–English bilinguals. In all studies, Spanish was the language spoken in the home, and English was learned subsequently via community and school exposure. Ninety-two of the Spanish–English bilingual children met criteria for DLD. In all studies, these children completed language assessments in both Spanish and English. Children included in the Spanish–English DLD group scored below the average range in both languages and also had either parent- or school-reported concerns regarding language development. The 109 TD Spanish–English bilingual children all scored within the average range in at least one of their two languages; scores within the average range in at least one language are considered sufficient to rule out DLD (e.g., Thordardottir, 2015). Maternal education for bilingual participants averaged “completed high school” in both source studies that collected this information (Ebert et al., 2019; Kohnert & Windsor, 2004). Bilingual participants from Windsor et al. (2010) and Ebert et al. (2014) were recruited from the same urban school district as the monolingual participants.

Tasks

All participants completed at least one of the three nonlinguistic cognitive tasks described below. All tasks were administered via computer using E-Prime software (Schneider, Eschman, & Zuccolotto, 2007). Responses were collected via button press on an external response box.

Visual Detection

All participants completed the same visual detection task (Kohnert & Windsor, 2004) as a measure of processing speed. Children were seated in front of a computer and were instructed to respond as quickly as possible to the appearance of a red or blue circle on the screen by pressing the button corresponding to the circle’s color. Before beginning the task,

Table 1. Summary of four source investigations for Study 1.

Source investigation	Age range (years;months)	Visual detection	Auditory pattern matching	Flanker	English only		Spanish–English	
					TD	DLD	TD	DLD
Kohnert & Windsor (2004)	7;10–10;11	X	X		29	16	18	—
Windsor et al. (2010)	6;0–10;11	X	X		67	35	66	17
Ebert et al. (2014)	6;0–10;7	X	X		—	—	—	46
Ebert et al. (2019)	6;0–8;11	X		X	25	22	25	29
Totals					121	73	109	92

Note. Em dashes indicate that a participant group was not included in the source study. TD = typically developing; DLD = developmental language disorder.

children completed a training phase of at least eight trials in which they were given feedback on the accuracy of their responses. The task consisted of 25 trials without feedback.

The dependent variable from the visual detection task was reaction time (RT). RT was calculated for correct trials only. In all studies, RT outliers were eliminated from the data set before calculating a mean RT for each participant. This procedure involved removing trials that were less than 50 ms, greater than 2,000 ms, or outside the individual participant's typical response range (defined as 2 *SDs* around the individual mean).

Auditory Pattern Matching

A subset of participants in both monolingual and bilingual groups (specifically those recruited from Ebert et al., 2014; Kohnert & Windsor, 2004; Windsor et al., 2010) completed an auditory pattern-matching task (originally described in Yim, 2006) that was designed as an assessment of working memory. On each trial, children listened to a pair of tone sequences and were asked to determine whether the two sequences matched. Trials began with two tones per sequence and progressed to include up to five tones per sequence. The original versions of the task presented 15 trials at each level of difficulty (i.e., two, three, four, and five tones per sequence; 25 trials per level were presented in Kohnert & Windsor, 2004), for a total of 60 trials (100 trials in Kohnert & Windsor, 2004). The dependent variable was the percentage of correct responses across all trials.

Flanker. A subset of children in both monolingual and bilingual groups (i.e., those recruited from Ebert et al., 2019) completed a flanker task based on the children's attentional network task (Rueda et al., 2004) as a measure of nonlinguistic attentional control. Each trial presented a linear array of five fishes on the screen, and participants were instructed to report the orientation of the center fish. On congruent trials, the surrounding four fishes faced the same direction as the center fish, and on incongruent trials, the surrounding four fishes faced the opposite direction. Each trial was preceded by a single fixation cross; unlike in the study of Rueda et al. (2004), there were no additional cues presented. Before completing the task, children completed a training phase with a minimum of four trials with only the center fish (to learn to report the direction of the center

fish) followed by a minimum of eight trials with a row of five fishes. Feedback was given during the training phase. Following the training phase, children completed the main flanker task, consisting of 64 trials without feedback.

Of key interest was the difference between incongruent and congruent trials because the performance decrement associated with incongruent trials is considered a measure of attentional control (Rueda et al., 2004). Scores for congruent and incongruent trials combined both accuracy and RT using procedures outlined in Zelazo et al. (2013; see also Ebert et al., 2019). The purpose is to allow scores from children with relatively poor accuracy who are not yet demonstrating a speed–accuracy trade-off to be included along with scores from children who do exhibit evidence of a speed–accuracy trade-off (Zelazo et al., 2013). Individual accuracy and RT scores were transformed to a 0- to 5-point scale, separately for congruent and incongruent trials. For children who exhibited task accuracy below 80%, the RT score was not included in the overall score. For example, a child who completed 60% of incongruent trials correctly would receive a score of 3 (i.e., 60% of the 5 possible accuracy points and 0 points for RT) for those trials. For children who exhibited task accuracy of 80% or higher, a scaled RT score is calculated and summed with the accuracy score. The range of possible RTs is determined (in this case, 500–2,000 ms), and the scaled score is based on this range. For example, a child who completed 80% of incongruent trials correctly and had a mean RT of 1100 ms would get 4 points for accuracy and 3 points for RT (because 1,100 ms is 3/5 of the way through the 500- to 2,000-ms range), resulting in a total score of 7 for incongruent trials (see Ebert et al., 2019, for further illustration).

Data Analyses

Age was expected to influence performance on all three tasks (Kail & Salthouse, 1994; Zelazo et al., 2016), and the data set included children across a wide age range. Therefore, children were first divided into 1-year age bands (i.e., 6-, 7-, 8-, 9-, and 10-year-olds) so that children with DLD could be compared to same-age peers.

We first conducted analyses separately for each sample in order to compare children with DLD to peers of the same

age and same linguistic background. In other words, English-only monolinguals were compared to other English-only monolinguals and Spanish–English bilinguals. As a preliminary step, within each sample, we first examined group-level differences between children with and without DLD on each task. Within each age band, *t*-test comparisons were conducted with DLD status as the independent variable. Levene's test of equality of variances was used to check the assumption of homogeneity of variances, and adjusted results were used for comparisons in which the assumption was violated. Hedges' *g* (Hedges & Olkin, 1985) was calculated as the effect size for these comparisons because sample sizes differed between DLD and TD groups within each age band. These effect sizes were interpreted using Cohen's (1988) suggestions: *g* = 0.2 represents a small effect, *g* = 0.5 represents a medium effect, and *g* = 0.8 represents a large effect.

Diagnostic accuracy analyses were then conducted using the OptimalCutpoints package in *R* (López-Ratón, Rodríguez-Álvarez, Cadarso-Suárez, & Gude-Sampedro, 2014). First, optimal cut-points were derived for each age group and task using the Youden method. The Youden method is a mathematical approach to simultaneously maximize sensitivity and specificity (López-Ratón et al., 2014). The cut-points were then used to calculate group sensitivity and specificity as well as positive and negative LR for each task. The 95% confidence intervals for the LR were also generated in these analyses.

We interpret sensitivity and specificity values using the guidelines for acceptable diagnostic accuracy proposed by Plante and Vance (1994): Sensitivity and specificity values below .80 are not useful diagnostically, values of .80–.89 are considered fair, and values of .90 and higher are considered good. We interpret positive and negative LR using the guidelines discussed by Dollaghan (2007): For positive LR, 1 is a neutral or uninformative test, 3 is a moderately positive test that may be suggestive but insufficient to confirm a diagnosis, and 10 or greater is a very positive test; for negative LR, 1 is again a neutral or uninformative test, 0.30 is a moderately negative test that may be suggestive but insufficient to rule out a diagnosis, and 0.10 or smaller is a very negative test. The 95% confidence intervals for the LR provide a very stringent test of whether the LR fall within these informative ranges.

Diagnostic accuracy analyses were first performed separately for each group in order to determine whether non-linguistic cognitive processing tasks can identify children with DLD from a group of peers from the same linguistic background. Analyses were then repeated with the combined group of both monolingual and bilingual children in order to determine whether diagnostic accuracy can be maintained in a diverse sample.

Results

Group Comparisons

Results of the *t*-test comparisons for each of the tasks, by age group, appear in Figure 1. On the visual detection task,

group differences favored the TD group, meaning that the DLD group was slower in every age band. Group differences were significant for three comparisons: 8-year-old English-only monolinguals, $t(16.74) = 3.19, p = .005, g = 1.27$; 9-year-old English-only monolinguals, $t(40) = 4.52, p < .001, g = 1.44$; and 8-year-old Spanish–English bilinguals, $t(45) = 2.74, p = .009, g = 0.83$. Effect sizes were large for these comparisons.

On the auditory pattern-matching test, group differences favored the TD group in all age bands for the English-only monolinguals. The comparison for 6-year-olds reached statistical significance, $t(17) = -2.42, p = .027, g = 1.52$. For the Spanish–English bilinguals, the group mean score favored the TD group for the 6- and 7-year-olds and the DLD group within the oldest three age bands. However, none of these comparisons reached statistical significance.

On the flanker task, group differences favored the TD group across all age bands in all samples, meaning that the TD group had smaller flanker effects (i.e., this group was less affected by incongruent trials on the task). Group differences did not reach statistical significance.

Diagnostic Accuracy

The task cut-points and resulting sensitivity and specificity appear in Table 2 for the English-only, Spanish–English, and combined samples. The table shows that the visual detection task had good specificity in the middle-age bands (7-, 8-, and 9-year-olds) of the English-only monolingual sample. The specificity values were generally maintained in the bilingual and combined samples, suggesting the task was relatively good at confirming typical development status for TD children, monolingual or bilingual, within these age bands.

In contrast, the auditory pattern-matching task generally had higher sensitivity than specificity, indicating that the task was relatively good at identifying DLD status for children with DLD. Adequate to perfect sensitivity was maintained in three age bands in the combined sample (6-, 9-, and 10-year-olds). This task also had adequate specificity in the 7-year-old age band for the Spanish–English bilingual and combined samples. Finally, the flanker task showed a less consistent pattern but did show adequate to perfect specificity for the oldest age band (8-year-olds) across the monolingual, bilingual, and combined samples.

Positive and negative LR for the English-only monolinguals appear in Figure 2. A total of six out of 13 positive LR were considered moderately positive (i.e., ≥ 3) to very positive (i.e., ≥ 10). Two LR were very positive (not depicted in the figure): the visual detection task for 9-year-olds ($LR+ = 16.25$) and the flanker task for 8-year-olds ($LR+ = \infty$). Four values were moderately positive: the visual detection task for 7- and 8-year-olds ($LR+ > 8$ for each age band), the auditory pattern-matching task for 6-year-olds ($LR+ = 4$), and the flanker task for 7-year-olds ($LR+ = 3$).

For negative LR, four of 13 values were moderately to very negative. Three negative LR had a value of 0, namely, the auditory pattern-matching task for 6-, 7-, and 8-year-olds. One negative LR was moderately negative, the

Figure 1. Group comparisons for each nonlinguistic cognitive processing task in Study 1. Means and standard error bars are shown for each task by age (in years), language group (English-only monolinguals or Spanish–English bilinguals), and language status of typically developing (TD) or developmental language disorder (DLD). * $p < .05$, ** $p < .01$, *** $p < .001$.

visual detection task for 10-year-olds ($LR- = 0.24$). With the exception of the four perfect LRs (i.e., negative LR of 0 or positive LR of infinity), the 95% confidence intervals extended outside the moderately informative range for both the positive and negative LRs.

Positive and negative LRs for the Spanish–English bilinguals appear in Figure 3. Three positive LRs were at or above the threshold for being moderately positive (i.e., ≥ 3): the visual detection task for 7- and 8-year-olds and the flanker task for 8-year-olds. Three negative LRs fell in the moderately to very negative range: The auditory pattern-matching task for 6- and 10-year-olds had perfect values of 0, and the flanker task for 7-year-olds had a moderately negative value of 0.23.

Positive and negative LRs for the combined sample appear in Figure 4. Three positive LR values were within the informative range: the visual detection task for 8-year-olds

($LR+ = 6.2$) and 9-year-olds ($LR+ = 4.4$) and the flanker task for 8-year-olds ($LR+ = 5.4$). For negative LRs, two values were perfect ($LR- = 0$): the auditory pattern-matching task for 6-year-olds and 10-year-olds.

Discussion

Study 1 examines the ability of three nonlinguistic cognitive processing tasks to identify children with DLD in monolingual, bilingual, and combined samples. There were many instances in which either sensitivity or specificity values were fair to excellent. Notably, these values were generally maintained in the combined sample that included monolinguals and bilinguals (see Table 2).

The LR analyses then incorporate sensitivity and specificity values into indices that show how much a test result can increase confidence that an individual does or

Table 2. Sensitivity and specificity for Study 1.

Samples		English only					Spanish–English					Combined				
		TD, n	DLD, n	Cut	Sens	Spec	TD, n	DLD, n	Cut	Sens	Spec	TD, n	DLD, n	Cut	Sens	Spec
Visual detection	6	24	15	970	.47	.71	26	20	782	.90	.27	50	35	960	.43	.74
	7	27	12	1,069	.33	.96	24	31	906	.42	.92	51	43	941	.40	.82
	8	23	14	863	.71	.91	30	17	878	.41	.90	53	31	863	.58	.91
	9	26	16	765	.63	.96	16	18	721	.44	.75	42	34	763	.53	.88
	10	20	13	617	.85	.65	11	7	697	.43	.73	31	20	661	.60	.74
Auditory pattern matching	6	16	3	0.70	1.00	.75	11	5	0.76	1.00	.55	27	8	0.76	1.00	.56
	7	13	2	0.76	1.00	.62	16	10	0.67	.40	.81	29	12	0.71	.33	.83
	8	20	10	0.76	1.00	.40	22	8	0.86	.88	.23	42	18	0.77	.78	.43
	9	26	14	0.76	.64	.58	16	16	0.82	.81	.31	42	30	0.82	.80	.36
	10	21	12	0.79	.67	.57	11	5	0.85	1.00	.09	32	17	0.88	1.00	.09
Flanker	6	8	10	0.82	.60	.63	9	8	0.64	.38	.78	17	18	0.79	.44	.71
	7	14	7	0.65	.43	.86	7	15	0.12	.87	.57	21	22	0.59	.59	.76
	8	3	5	0.74	.40	1.00	7	8	0.61	.50	.86	10	13	0.61	.54	.90

Note. Bold-only text is used to indicate sensitivity (Sens) and specificity (Spec) values that reached the adequate criterion of .8 or higher (Plante & Vance, 1994). Bold and underlined text is used for sensitivity and specificity values that reached the good criterion of .9 or higher. Cut-points (Cut) for visual detection are reported in milliseconds. Cut-points for auditory pattern matching are reported in percent correct. Cut-points for flanker are the difference between incongruent and congruent trials in scaled score points. TD = typically developing; DLD = developmental language disorder.

Figure 2. Likelihood ratios (LRs) for English-only monolinguals in Study 1. Markers indicate the positive and negative LR values, and horizontal bars indicate the 95% confidence interval for each value. Values farther from the center in both plots indicate more informative LRs. Boxed and shaded areas indicate LRs that are considered moderately to very informative (i.e., 0.3 or less for negative LRs, 3 or more for positive LRs). The maximum value on the horizontal axis of the positive LR plot was set at 10 in order to accurately depict the majority of the positive LRs. However, some values extended beyond this maximum and are truncated here: The positive LR for visual detection for 9-year-olds is 16; the upper bounds of the confidence intervals for the visual detection task are 72 for 7-year-olds, 32 for 8-year-olds, and 115 for 9-year-olds. For the flanker task, the upper bound of the 7-year-old interval is 14. The positive LR for 8-year-olds on the flanker task is infinite and cannot be depicted. Aud. = auditory.

does not have a disorder. The nonlinguistic tasks examined here again achieve some notable successes. For example, there were several informative positive LRs on the visual detection task, meaning that performance slower than the cut-point on this task could substantially increase the confidence that a child has DLD. On the auditory pattern-matching task, the negative LRs are generally more informative than the positive LRs, suggesting that performance better than the cut-point could be helpful in confirming TD status. The flanker task could be most successful in confirming suspicions of DLD in the oldest age group that completed it (i.e., 8-year-olds who had informative positive LRs across monolingual, bilingual, and combined samples). Given the pattern here, exploring the flanker task in 9- and 10-year-old children may be worthwhile. We return to the strengths and limitations of our results in the general discussion below.

Study 2

The greatest attraction in using nonlinguistic cognitive processing tasks to assist in the identification of DLD is the promise of using the same approach universally. Therefore, in Study 2, we repeat the consideration of diagnostic accuracy within a different, international context. Study 2 includes monolingual speakers of Vietnamese who were living in Vietnam.

Method

Participants

Fifty-five Vietnamese-only monolingual children were recruited in a single study conducted in Vietnam (Pham et al., 2019). Children were recruited from kindergarten programs at four schools in Hanoi. The resulting group ranged in age from 5;2 to 6;2. Participating families reported a higher socioeconomic status than the average for Vietnam, and 85% of parents had a college degree (see Pham et al., 2019). As in Study 1, children were required to score within the average range on a test of nonverbal intelligence, and parent report was used to exclude other disabilities. Parent and teacher reports were also used to confirm normal hearing status.

Ten children met criteria for DLD. In the absence of standardized assessments in Vietnamese, children in this study were diagnosed with DLD through a convergence of parent report, teacher report, and direct assessment of Vietnamese language skills such as vocabulary and grammar (see Pham et al., 2019). Children were classified as having DLD if they met the following three criteria: (a) a score below 1 *SD* from the mean on a report measure (i.e., parent or teacher concern), (b) scores below 1 *SD* from the mean on at least three direct language measures, and (c) low language performance that spanned two of three domains

Figure 3. Likelihood ratios (LRs) for Spanish–English bilinguals in Study 1. Markers indicate the positive and negative LR values, and horizontal bars indicate the 95% confidence interval for each value. Values farther from the center in both plots indicate more informative LRs. Boxed and shaded areas indicate LRs that are considered moderately to very informative (i.e., 0.3 or less for negative LRs, 3 or more for positive LRs). The maximum value on the horizontal axis of the positive LR plot was set at 10 in order to accurately depict the majority of the positive LRs; however, three confidence intervals extend beyond this maximum value and are truncated here. The upper bounds of the confidence intervals for the visual detection task are 20 for 7-year-olds and 14 for 8-year-olds. For the flanker task, the upper bound of the 8-year-old interval is 14. The maximum value on the horizontal axis of the negative LR plot was set at 2 in order to accurately depict the majority of the negative LRs; however, two confidence intervals extend beyond this maximum value and are truncated here. The upper bounds of the confidence intervals for the auditory (Aud.) pattern-matching task are 4.7 for 8-year-olds and 2.1 for 9-year-olds.

of vocabulary, grammar, and/or narratives. Children who scored within the average range on all indicators (i.e., parent report, teacher report, and direct language measures) qualified for the TD group ($n = 45$).

Tasks and Analysis

Participants in Study 2 also completed nonlinguistic cognitive processing tasks implemented within E-Prime (Schneider et al., 2007). Responses were collected via an external keypad. The visual detection and flanker tasks were identical to those in Study 1. The auditory pattern-matching task was modified slightly to improve suitability for young children. Instead of automatically advancing the task difficulty (i.e., the length of the tone sequences) across trials, the modified version only advanced task difficulty when children demonstrated sufficient accuracy at the current tone sequence length (defined as four consecutively correct trials; see also Ebert, 2014, for complete details regarding this version of the task). In the modified version, the dependent variable remained an accuracy score; however, to account for difficulty level, the accuracy score was calculated as the number of correct trials multiplied by the length of the tone sequence for each level, summed across all levels (see Ebert, 2014). For example, if a child correctly responded to four items at the two-tone difficulty level, seven items at the three-tone difficulty level, and

one item at the four-tone difficulty level, their total score would be $(4 \times 2) + (7 \times 3) + (1 \times 4) = 33$. Analyses for Study 2 mirrored those of Study 1. Children in Study 2 were all recruited from the same grade and were not further subdivided by age.

Results

Group comparisons were first conducted using t tests. In Study 2, Levene's test was significant for all comparisons, and adjusted results were used. On the visual detection task, the group difference favored the TD group and approached significance, $t(9.33) = 2.26$, $p = .05$, $g = 1.48$. On the modified auditory pattern-matching task, the TD group was more accurate, $t(37.65) = -8.84$, $p < .001$, $g = 1.85$. On the flanker task, the TD group had a smaller mean flanker effect than the DLD group, but the difference was not significant, $t(9.53) = 1.28$, $p = .23$, $g = 0.78$. Effect sizes were large for all comparisons.

Next, diagnostic accuracy analyses were conducted. For the visual detection task, the optimal cut-point fell at 1,203 ms, resulting in a sensitivity of .70 and a specificity of .71. For the modified auditory pattern-matching task, the optimal cut-point was 26, resulting in a sensitivity of .90 and a specificity of .80. For the flanker task, the optimal cut-point was 0.98, resulting in a sensitivity of .60 and a specificity of .82.

Figure 4. Likelihood ratios (LRs) for the combined sample of English-only monolinguals and Spanish–English bilinguals in Study 1. Markers indicate the positive and negative LR values, and horizontal bars indicate the 95% confidence interval for each value. Values farther from the center in both plots indicate more informative LRs. Boxed and shaded areas indicate LRs that are considered moderately to very informative (i.e., 0.3 or less for negative LRs, 3 or more for positive LRs). The maximum value on the horizontal axis of the positive LR plot was set at 10 in order to accurately depict the majority of the positive LRs; however, three confidence intervals extend beyond this maximum value and are truncated here. The upper bounds of the confidence intervals for the visual detection task are 15 for 8-year-olds and 10.7 for 9-year-olds. For the flanker task, the upper bound of the 8-year-old interval is 37. Aud. = auditory.

Positive and negative LRs for Study 2 appear in Figure 5. In this sample, both the positive and negative LRs for the modified auditory pattern-matching task fall within the moderately informative range. The positive LR for the flanker task was also moderately informative.

Discussion

In Study 2, we utilized the same nonlinguistic cognitive processing tasks with a group of kindergarten monolingual speakers of Vietnamese. These initial results show promise. The visual detection task appears less diagnostically useful than in Study 1, perhaps because the children in Study 2 are younger. The auditory pattern-matching task, however, simultaneously achieves moderate to good sensitivity and specificity, as well as informative LRs for both the positive and negative values. These results indicate that the auditory pattern-matching task could have value both for confirming the presence of a disorder for children with DLD and for confirming typical language status for TD children. In Study 1, auditory pattern matching was a sensitive task for the youngest age groups but was not specific, meaning that some TD children did poorly on it. In Study 2, participants completed a modified version of this task, in which task difficulty was adjusted depending on the child's performance. This modification appears to have improved the task's

specificity while maintaining its sensitivity, and it serves as an example of how the existing tasks could be improved.

In Study 2, the participating age group was more restricted than in Study 1, and the overall socioeconomic status of participants was somewhat higher than in Study 1. In addition, the qualification process for DLD was somewhat different from the U.S.-based samples. However, the qualification process was altered due to the lack of standardized assessments for Vietnamese, and this scenario—needing to identify DLD in a different country that speaks an underresearched language—is one in which universally validated tasks would be most useful.

General Discussion

This investigation is the first step in exploring the diagnostic accuracy of nonlinguistic cognitive processing tasks in DLD identification. We culled information from five source investigations (four investigations in Study 1 and one in Study 2) to evaluate the performance of children from a wide age range (5- to 10-year-olds) who spoke one or two out of three distinct languages (English, Spanish, and Vietnamese) and who lived in distinct countries (United States and Vietnam). Results suggest both promise and room for improvement. One highlight is the maintenance of diagnostic accuracy results when monolingual and bilingual children

Figure 5. Likelihood ratios (LRs) for Vietnamese-only monolinguals in Study 2. Markers indicate the positive and negative LR values, and horizontal bars indicate the 95% confidence interval for each value. Values farther from the center in both plots indicate more informative LRs. Boxed and shaded areas indicate LRs that are considered moderately to very informative (i.e., 0.3 or less for negative LRs, 3 or more for positive LRs).

were combined in Study 1. These results demonstrated the potential for a single cut-point from a nonlanguage task to show fair to good specificity or sensitivity within a sample of children with diverse backgrounds and experiences. Indeed, even existing language-based tasks have needed different cut-points for children with different linguistic experiences (Armon-Lotem & Meir, 2016; de Almeida et al., 2017; Oetting et al., 2016).

However, there is also room for improvement. Both sensitivity and specificity that reach fair to good levels are needed for diagnostic accuracy, and this standard is not yet met for most of the nonlinguistic tasks analyzed here. Similarly, in most instances, either a positive or negative LR was informative, but not both. A few notable exceptions involved the auditory pattern-matching task. In Study 1, this task yielded informative positive and negative LRs for English-only 6-year-olds living in the United States (see Figure 2). Similarly, performance on a modified version of this task yielded informative positive and negative LRs for Vietnamese-only monolinguals living in Vietnam (see Figure 5). High diagnostic accuracy for younger children (5- to 6-year-olds) across two highly distinct languages and countries is an illustration of how useful nonlinguistic cognitive processing tasks could be for clinical identification. Moreover, the modified version of auditory pattern-matching task used in Study 2 was tailored to the difficulty level achieved by individual children. Adaptive task designs such as this one could reduce the amount of time required to complete such tasks, which in turn increases the feasibility of incorporating them into an identification protocol.

For tasks that had either positive or negative LRs within the informative range but not both, results indicate the task could help to either confirm or rule out DLD. In these cases, the nonlinguistic processing task would need to be used as a component of an identification battery rather than a stand-alone test. For example, performance slower than the cut-point on the visual detection task by an 8-year-old child could be used, in combination with linguistic errors in a language sample or poor performance on a language test, to confirm a diagnosis of DLD. This would be particularly useful when it is not clear whether poor performance on language measures can be attributed to differences in language exposure or to DLD. In this case, poor performance on a nonlinguistic cognitive processing task could confirm the presence of a disorder. In contrast, a performance faster than the cut-point within this age group could be consistent with either TD or DLD. The nonlinguistic cognitive assessment would be minimally informative in this case, and further assessments would be necessary to determine diagnostic status. Of course, further evaluation would always be needed for assessment purposes beyond identification such as characterizing strengths and weaknesses or setting goals for intervention.

Across all analyses, one limitation in the LRs is the size of the 95% confidence intervals. Except in instances in which the LR was perfect (i.e., 0 for a negative ratio or infinity for a positive ratio), the confidence intervals extended outside the range of informative values. Ideally, confidence intervals would fall entirely within the diagnostically informative range (Dollaghan, 2007). However,

to our knowledge, this highly rigorous standard of diagnostic accuracy has not yet been met in investigations of DLD, even when English language measures are examined within relatively homogenous English-only samples.

In the search for tools to identify DLD across a wide variety of languages and cultures, this investigation is the first step of many toward incorporating nonlinguistic cognitive processing tasks in assessment. Further work is needed to refine the most appropriate tasks for each age band and to verify the cut-points established in this investigation. Moreover, continued research on the connections between DLD and nonlinguistic cognitive processing skills could lead to new tasks that are more accurate than the tasks explored here. To conclude, the potential contribution of nonlinguistic cognitive processing tasks in improving diagnostic accuracy, particularly for linguistically diverse populations, is intriguing. The next steps will dictate whether a set of nonlinguistic cognitive processing tasks can be refined and validated for clinical practice.

Acknowledgments

Funding sources for the data analyzed in this study include National Institutes of Health Awards R01DC004437 (J. Windsor), R03DC004442 (K. Kohnert), R21HD053222 (J. Windsor), R21DC010868 (K. Kohnert), R03DC013760 (K. Ebert), and K23DC014750 (G. Pham). Portions of this database were collected with support from the University of Minnesota. We thank Kathryn Kohnert and Jennifer Windsor for access to data, and we are grateful to the many research staff and participants in the studies considered here.

References

- Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol, United Kingdom: Multilingual Matters.
- Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders*, 51(6), 715–731.
- Ash, A. C., & Redmond, S. M. (2014). Using finiteness as a clinical marker to identify language impairment. *SIG 1 Perspectives on Language Learning and Education*, 21(4), 148–158.
- Bedore, L. M., & Leonard, L. B. (2005). Verb inflections and noun phrase morphology in the spontaneous speech of Spanish-speaking children with specific language impairment. *Applied Psycholinguistics*, 26(2), 195–225.
- Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, 11(1), 1–29.
- 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.
- Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language, Speech, and Hearing Services in Schools*, 44, 133–146.
- Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *The Journal of Child Psychology and Psychiatry*, 58(10), 1068–1080.
- Bortolini, U., Caselli, M. C., Deevey, P., & Leonard, L. B. (2002). Specific language impairment in Italian: The first steps in the search for a clinical marker. *International Journal of Language & Communication Disorders*, 37(2), 77–93.
- Castilla-Earls, A. P., Restrepo, M. A., Perez-Leroux, A. T., Gray, S., Holmes, P., Gail, D., & Chen, Z. (2016). Interactions between bilingual effects and language impairment: Exploring grammatical markers in Spanish-speaking bilingual children. *Applied Psycholinguistics*, 37(5), 1147–1173.
- Cleveland, L. H., & Oetting, J. B. (2013). Children's marking of verbal –s by nonmainstream English dialect and clinical status. *American Journal of Speech-Language Pathology*, 22(4), 604–614.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- de Almeida, L., Ferré, S., Morin, E., Prévost, P., dos Santos, C., Tuller, L., ... Barthez, M.-A. (2017). Identification of bilingual children with specific language impairment in France. *Linguistic Approaches to Bilingualism*, 7(3), 331–358.
- Dispaldro, M., Leonard, L. B., & Deevey, P. (2013). Real-word and nonword repetition in Italian-speaking children with specific language impairment: A study of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, 56(1), 323–336.
- Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore, MD: Brookes.
- Ebert, K. D. (2014). Nonlinguistic cognitive effects of language treatment for children with primary language impairment. *Communication Disorders Quarterly*, 35(4), 216–225.
- Ebert, K. D., & Kohnert, K. (2011). Sustained attention in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 54, 1372–1384.
- Ebert, K. D., Kohnert, K., Pham, G., Rentmeester Disher, J., & Payesteh, B. (2014). Three treatments for bilingual children with primary language impairment: Examining cross-linguistic and cross-domain effects. *Journal of Speech, Language, and Hearing Research*, 57, 172–186.
- Ebert, K. D., & Pham, G. (2017). Synthesizing information from language samples and standardized tests in school-age bilingual assessment. *Language, Speech, and Hearing Services in Schools*, 48, 42–55.
- Ebert, K. D., Rak, D., Slawny, C. M., & Fogg, L. (2019). Attention in bilingual children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 62, 979–992.
- Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50(1), 177–195.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, 86, 199–225.
- Kapalková, S., Polišínská, K., & Vicienová, Z. (2013). Non-word repetition performance in Slovak-speaking children with and without SLI: Novel scoring methods. *International Journal of Language & Communication Disorders*, 48(1), 78–89.
- Kohnert, K., & Windsor, J. (2004). The search for common ground: Part II. Nonlinguistic performance by linguistically

- diverse learners. *Journal of Speech, Language, and Hearing Research*, 47(4), 891–903.
- Kohnert, K., Windsor, J., & Ebert, K. D. (2009). Primary or “specific” language impairment and children learning a second language. *Brain and Language*, 109, 101–111.
- Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.
- Leonard, L. B., Ellis Weismer, S., Miller, C. A., Francis, D. J., Tomblin, J. B., & Kail, R. V. (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research*, 50(2), 408–428.
- López-Ratón, M., Rodríguez-Álvarez, M. X., Cadarso-Suárez, C., & Gude-Sampedro, F. (2014). OptimalCutpoints: An R package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software*, 61(8), 1–36.
- Miller, C. A., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 44(2), 416–433.
- Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working memory and specific language impairment: An update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology*, 19(1), 78–94.
- Moyle, M. J., Heilmann, J. J., & Finneran, D. A. (2014). The role of dialect density in nonword repetition performance: An examination with at-risk African American preschool children. *Clinical Linguistics & Phonetics*, 28(9), 682–696.
- Oetting, J. B. (2018). Prologue: Toward accurate identification of developmental language disorder within linguistically diverse schools. *Language, Speech, and Hearing Services in Schools*, 49(2), 213–217.
- Oetting, J. B., McDonald, J. L., Seidel, C. M., & Hegarty, M. (2016). Sentence recall by children with SLI across two non-mainstream dialects of English. *Journal of Speech, Language, and Hearing Research*, 59(1), 183–194.
- Pauls, L. J., & Archibald, L. M. (2016). Executive functions in children with specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 59(5), 1074–1086.
- Peña, E., Gutiérrez-Clellen, V., Iglesias, A., Goldstein, B., & Bedore, L. (2018). *Bilingual English–Spanish Assessment*. Baltimore, MD: Brookes.
- Pham, G., Ebert, K. D., Dinh, K. T., & Dam, Q. (2018). Nonword repetition stimuli for Vietnamese-speaking children. *Behavior Research Methods*, 50, 1–16.
- Pham, G., Pruitt-Lord, S., Snow, C. E., Nguyen, H. T. Y., Pham, B., Dao, T. B. T., ... Dam, Q. D. (2019). Identifying developmental language disorder in Vietnamese children. *Journal of Speech, Language, and Hearing Research*, 62(5), 1452–1467.
- Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, 25(1), 15–24.
- Rodekohr, R. K., & Haynes, W. O. (2001). Differentiating dialect from disorder: A comparison of two processing tasks and a standardized language test. *Journal of Communication Disorders*, 34(3), 255–272.
- Roebuck, H., Sindberg, H., & Weismer, S. E. (2018). The role of language in nonlinguistic stimuli: Comparing inhibition in children with language impairment. *Journal of Speech, Language, and Hearing Research*, 61(5), 1216–1225.
- Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, 42(8), 1029–1040.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2007). E-Prime 2.0 Professional [Computer software]. Pittsburgh, PA: Psychology Software Tools.
- Šimundić, A. M. (2008). Measures of diagnostic accuracy: Basic definitions. *Medical and Biological Sciences*, 22(4), 61–65.
- Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools*, 37(1), 61–72.
- Thordardottir, E. (2015). Proposed diagnostic procedures and criteria for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol, United Kingdom: Multilingual Matters.
- Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, 46(1), 1–16.
- Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, 41(3), 399–433.
- Verhoeven, L., Steenge, J., van Weerdenburg, M., & van Balkom, H. (2011). Assessment of second language proficiency in bilingual children with specific language impairment: A clinical perspective. *Research in Developmental Disabilities*, 32(5), 1798–1807.
- Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: A meta-analysis. *Research in Developmental Disabilities*, 34(9), 2586–2597.
- Windsor, J., Kohnert, K., Lobitz, K. F., & Pham, G. T. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*, 19(4), 298–310.
- Yang, H., & Gray, S. (2017). Executive function in preschoolers with primary language impairment. *Journal of Speech, Language, and Hearing Research*, 60(2), 379–392.
- Yim, D. (2006). *Auditory and visual pattern learning and language skills in children and adults* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 3240493).
- Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH toolbox cognition battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, 78(4), 16–33.
- Zelazo, P. D., Blair, C. B., & Willoughby, M. T. (2016). *Executive function: Implications for education. NCER 2017-2000*. Washington, DC: National Center for Education Research.