LSHSS

Research Article

The Cultural and Diagnostic Appropriateness of Standardized Assessments for Dual Language Learners: A Focus on Jamaican Preschoolers

Rachel Wright Karem[a,b] and Karla N. Washington[a]

Purpose: The aim of this study was to investigate the appropriateness of standardized assessments of expressive grammar and vocabulary in a sample of preschool-age dual language learners (DLLs) who use Jamaican Creole (JC) and English. Adult models from the same linguistic community as these children were used to inform culturally and linguistically appropriate interpretation of children's responses to a standardized assessment.
Method: JC-English–speaking preschoolers (n = 176) and adults (n = 33) completed the Word Structure and Expressive Vocabulary subtests of the Clinical Evaluation of Language Fundamentals Preschool–Second Edition. Adults' responses were used to develop an adapted scoring procedure that considered the influence of JC linguistic features on responses. DLLs' responses scored using the standard English and adapted JC procedures were compared.
Results: JC–English DLLs and adults used similar linguistic structures in response to subtest questions. DLLs' scores differed significantly from the standardized sample on both subtests. Preschoolers received higher raw and corresponding standard scores with adapted scoring compared to standard scoring. Adapted scoring that made use of adult models yielded high classification accuracy at a rate of 93.8% for Word Structure and 92.1% for Expressive Vocabulary.
Conclusions: Adapting standardized assessment scoring procedures using adult models may offer an ecologically valid approach to working with DLL preschoolers that can support a more accurate assessment of language functioning. These findings suggest that the use of standardized assessments for bilingual JC–English speakers requires a culturally responsive approach.
Supplemental Material: https://doi.org/10.23641/asha.14403026

It is well established that young children's language functioning is inextricably linked to future academic achievement (Guiberson & Ferris, 2019; International Expert Panel on Multilingual Children's Speech [IEPMCS], 2012). The early and accurate determination of children's language abilities is an important part of their general education programming and is critical to identifying children who need special education services (Bedore & Peña, 2008; IEPMCS, 2012). For young children who speak two languages, dual language learners (DLLs), the diagnosis of developmental language disorders (DLD) can be particularly challenging (American Speech-Language-Hearing Association [ASHA], n.d.; Paradis et al., 2011). This is because a linguistic feature that marks a disorder in one language may be a natural part of another language. For DLLs, speech-language pathologists (SLPs) play a critical role in determining if a child's language profile is reflective of a *language difference* or a *language disorder*. *Language difference* occurs if there are rule-governed differences typical of dual language development. *Language disorder* occurs if there are significant discrepancies in language skills across all of a child's languages compared to what is expected at their age, experience, or developmental level (Paradis et al., 2011; Wright Karem et al., 2019).

While the range of available research-based assessment procedures to appropriately measure the complexities of dual language use has grown, there is currently a lack of comprehensive assessment tools and strategies to accurately diagnose DLD in DLLs. The Individuals with Disabilities Education

[a]Department of Communication Sciences and Disorders, University of Cincinnati, OH
[b]Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington
Correspondence to Karla N. Washington: washink2@ucmail.uc.edu

Editor-in-Chief: Holly L. Storkel
Editor: Katie Squires
Received August 14, 2020
Revision received November 30, 2020
Accepted February 5, 2021
https://doi.org/10.1044/2021_LSHSS-20-00106

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

Act (IDEA, 2006, Part B) mandates that assessments must be based on a student's language or culture and must provide the most accurate information regarding a child's developmental, academic, and functional knowledge. ASHA (n.d.) established guidelines to assist SLPs in the assessment of DLLs and to ensure that assessments not only are nondiscriminatory but also include a comprehensive assessment battery. ASHA (n.d.) states, "Given the complexities involved in bilingualism and the significant variability that exists among the linguistic skills of multilingual individuals, clinicians must be prepared to address the unique situation of each client" (p. 1). However, barriers remain that can negatively impact the accurate diagnosis of DLD in DLLs, such as the lack of appropriate assessment tools and specific knowledge regarding dual language profiles (Lewis et al., 2010; K. N. Washington et al., 2019).

Though research has sought to identify evidence-based strategies to appropriately assess the linguistic variability of DLLs' language profiles (Bedore & Peña, 2008; Lewis et al., 2010), limitations regarding commonly used and heavily relied upon standardized assessments remain (Finestack & Satterlund, 2018; Pearce & Williams, 2013). For example, despite the fact that the Clinical Evaluation of Language Fundamentals Preschool (CELF Preschool, Wiig et al., 1992; CELF Preschool-2, Wiig et al., 2006) has been a widely used standardized assessment to identify children with DLD for the past 30 years, few studies have examined its appropriateness with DLLs. Coupled with this, no study to date has examined the appropriateness of standardized assessments with understudied language pairings such as Jamaican Creole (JC)-English–speaking children, the third largest Caribbean-born immigrant group in the United States (U.S. Census Bureau, 2018). The growth in this population increases the possibility that SLPs will have JC–English speakers on their caseload, necessitating an understanding of how this linguistically diverse populace might perform on popularly used assessments. The aim of this study was to investigate the appropriateness of a standardized assessment of expressive grammar (i.e., morphosyntax) and vocabulary in a sample of JC-English–speaking preschoolers. These language domains are of critical importance to distinguish DLD from typical development in DLLs (Blom & Paradis, 2015; Paradis et al., 2011) and are impactful measures of oral language proficiency needed to guide educational programming (Castro et al., 2013; Greenwood et al., 2016). In this study, we informed the interpretation of standardized assessment results based on the responses of adult models from the same linguistic community to the same test items. We discuss applicable qualitative and quantitative approaches useful in characterizing adult and child language use and for developing an adapted scoring procedure that considers JC linguistic features in the English assessment context.

Distinguishing Between Language Difference and DLD

Historically, there has been a disproportionate representation of racial and ethnic groups in special education due to misdiagnosis (Dragoo, 2017). Inappropriate identification can result from both under- and overdiagnosis of language disorders (P. L. Morgan et al., 2016; Paradis et al., 2011). Misdiagnosis often occurs because *language difference* in bilinguals may overlap with indicators for *language disorder* in monolingual English speakers (Castilla-Earls et al., 2016; Paradis et al., 2011). *Language difference* occurs if there are rule-governed differences in a child's presentation of the community language (e.g., English) typical of multilingual language development. *Language disorder* occurs if there are significant discrepancies in language skills across all of a child's languages compared to defined expectations by age, experience, or developmental level that impact all languages used by the child (Paradis et al., 2011). For example, monolingual English speakers with DLD commonly omit grammatical morphemes, such as present progressive "*–ing*" or plural "*–s*" exceeding beyond typical developmental expectations. However, this same pattern is common in typically developing DLLs as they acquire English (Paradis et al., 2011; K. N. Washington et al., 2019) because their first language may not contain these same linguistic features. Therefore, typical patterns of cross-linguistic influence (i.e., interactions between each of the child's languages) and/or code mixing may be incorrectly identified as indicators of DLD rather than language difference, resulting in overdiagnosis. Conversely, overlapping patterns indicative of DLD could be incorrectly deemed as an English-learning pattern (i.e., *difference*), resulting in underdiagnosis (P. L. Morgan et al., 2016; Oetting et al., 2019). An example pattern resulting in potential misdiagnosis of JC–English DLLs might be use of the JC-influenced sentence "Im [him] a sleep" rather than the English sentence "He is sleeping." The JC structure could be described as having errors in the subject "he" and omission of "is" and "*–ing*." However, the JC-influenced sentence could be more accurately described as having cross-linguistic influence because, in JC, "Im" is the correct nominative case pronoun and "a" serves as a present tense (continuative aspect) marker (cf. K. N. Washington et al., 2019). On the other hand, a JC–English DLL's response of "Ar [her] is" rather than the English target "She is" could be incorrectly characterized as a language difference, when in fact it could be described as errored or disordered. Of importance, language disorder manifests according to the multilingual child's unique language profile; thus, patterns of language difference and language disorder can co-occur (Oetting, 2018). SLPs need specific knowledge of diverse linguistic profiles to appropriately classify language patterns to avoid misdiagnosis and guide appropriate treatment decisions (cf. De Lamo White & Jin, 2011; Paradis et al., 2011; Selin et al., 2019).

Standardized Assessments

Norm-referenced, standardized assessments have been considered a cornerstone of diagnosis in the field of speech-language pathology (De Lamo White & Jin, 2011). Reliance on such assessments could be attributed to multiple factors, including (a) the ability to assess multiple

language domains, (b) requirement of a standardized score for qualification for special education services in school-based settings, (c) ease and efficiency of administration and scoring, and (d) comparison to a norm-referenced sample (Barragan et al., 2018; De Lamo White & Jin, 2011). However, the potential for misdiagnosis of DLLs through use of standardized assessments has been clearly demonstrated. Pearce and Williams (2013) found that when assessed using the CELF-4 Australian Standardized Edition (Semel et al., 2006a), children who used Aboriginal English were negatively impacted, with assessment scores not aligning with educator ratings of language skills. The reduced accuracy of the measure was due to scoring not capturing cross-linguistic language features, potentially resulting in misdiagnosis. G. P. Morgan et al. (2013) found that the CELF-4 Spanish (Semel et al., 2006b) underidentified monolingual Spanish speakers with language impairment, while Barragan et al. (2018) found that the CELF-4 Spanish overidentified low-income Spanish–English DLLs. Thus, research continues to demonstrate a lack of validity in current standardized assessments for DLLs, with notable challenges in their ability to capture contextualized factors (e.g., socioeconomic status [SES], bilingual typology) and cross-linguistic variation that shape dual language development.

The Adaptation of Standardized Assessments

One approach to improve the utility and ecological validity of standardized assessments for DLLs is to adapt them to account for cross-linguistic differences (Paradis et al., 2011), a strategy that is commonly used in research studies (Wright Karem et al., 2019). Gross et al. (2014) used conceptual scoring of children's responses on the Peabody Picture Vocabulary Test–Third Edition (Dunn & Dunn, 1997) and the Spanish version (Test de Vocabulario en Imágenes Peabody; Dunn et al., 1986) where the child was given another opportunity to respond in the nontarget language for incorrect responses. If a correct response was given in the nontarget language, the item was rescored as correct, obtaining the conceptual score. Children's scores improved when conceptual scoring, rather than standard scoring, was used. Gorman et al. (2016) developed a Spanish adaptation of the English Narrative Assessment Protocol (Justice et al., 2010) to include language-specific features of Spanish. By comparing Spanish–English DLLs' performance between the original English protocol and the Spanish-adapted protocol, they found that inclusion of language-specific features benefited the validity of the tool to analyze children's language abilities. Hendricks and Adlof (2017) investigated how modified scoring procedures that considered known features of African American English (AAE) affected the diagnostic accuracy of the CELF-4 (Semel et al., 2003). The researchers found that modified scoring yielded higher scores for AAE speakers; however, false negatives (i.e., resulting in underidentification) using this approach did also occur.

More recently, Oetting et al. (2019) examined three different scoring approaches to detail the accuracy and utility of adapted scoring measures for dialectal speakers (i.e., AAE and Southern White English) informed by knowledge of dialectal language patterns (e.g., zero marking of regular past tense). In this study, three different scoring procedures were examined: (a) unmodified, that is, counting only mainstream overt forms; (b) modified, that is, counting mainstream and nonmainstream overt and zero forms; and (c) strategically modified, that is, counting only mainstream and nonmainstream overt forms and excluding responses that did not obligate the targeted tense and agreement structure. In line with the structure of traditional standardized assessments, this study focused on dialect-informed language probes rather than language sampling to control for structures elicited. The researchers found that all scoring approaches revealed differences between typically developing dialectal speakers and dialectal speakers with DLD; however, strategically modified scoring yielded the highest levels of sensitivity and specificity. Importantly, unmodified scoring led to overidentification, and modified scoring led to underidentification. Study findings highlighted the importance of using strategic scoring approaches unique to linguistic properties of the child's languages and the need for additional strategies to inform scoring procedures for linguistically diverse children.

Another method that has received relatively little attention in the literature and could address concerns of underidentification when using adapted scoring procedures is the use of adult models to inform scoring of DLLs on standardized assessments (Canagarajah, 2006; Paradis, 2016). Terrell et al. (1992) used the Parent–Child Comparative Analysis to compare a child's response on English standardized assessments to the parent's response in order to consider home language influences. Child responses that deviated from the English target were compared with parent responses to determine if the production was cross-linguistically appropriate or an error. Using the parent model to interpret assessment results in the English context provided pivotal information on the child's true language abilities, thus avoiding misdiagnosis. Other research has suggested that children's productions may differ from that of their parents and that a wider body of speakers should be considered, such as community models (Canagarajah, 2006; J. A. Washington, 1996). Masso et al. (2019) found that English-speaking SLPs' transcription accuracy of child speech samples in Vietnamese improved when adult models were used as a point of comparison. Lockart and McLeod (2013) investigated non–Cantonese-speaking clinicians' transcription of typical and atypical Cantonese speech samples, finding improved accuracy following training with adult models. These studies document the potential utility of adult models to inform analysis for bilingual populations.

In a review, Canagarajah (2006) documents the immense number of varieties of English both within communities and across the globe, suggesting a multilayering of local norms to inform assessment is needed for interpretation in English contexts. Norms may vary according to the communicative context; thus, local models obtained in specific contexts could support identifying varying linguistic patterns across speakers. To this end, in a recent tutorial,

McLeod et al. (2017) suggested that adults from the same linguistic community may provide appropriate guidance in determining if responses are consistent with child models. The use of adult models to interpret performance is an underutilized resource, yet it has promising implications for informing future research and practice to document the unique variability in dual language use.

Researchers are recognizing promise in adapting standardized assessments and the need to identify DLLs' specific linguistic patterns in English assessment contexts, where misdiagnosis is most often likely to occur (McLeod et al., 2017). Though emerging evidence has suggested the utility of adapted standardized assessment, concerns persist regarding how to best adapt procedures to capture the complexity and variability of multilingual speakers, while also avoiding underidentification (Gross et al., 2014; Hendricks & Adlof, 2017; Oetting et al., 2019). For understudied populations, such as JC-English–speaking children, this concern is heightened as little is known about their linguistic profiles (K. N. Washington, 2012), making the application of adult models an ecologically valid approach that supports accurate diagnoses. Accordingly, we focused on JC-English–speaking bilinguals' performance in the English assessment context. Identifying patterns of language difference could guide the adaptation of standardized assessments to reduce cultural and linguistic bias.

This Study

Jamaican children are particularly at risk for misdiagnosis of language disorders in English-dominant communities (K. N. Washington et al., 2019). While recent evidence has described the speech and language profiles of JC-English–speaking children (Abu El Adas et al., 2020; K. N. Washington et al., 2019, 2017), the appropriateness of commonly used assessment tools has not been described for this population. To better understand the appropriateness of commonly used assessment tools for understudied DLL populations, we seek to investigate the Word Structure and Expressive Vocabulary subtests of the CELF Preschool-2 for JC-English–speaking preschoolers. Standards of practice indicate that measuring these language domains using standardized assessment may serve as pivotal components to a comprehensive assessment battery (i.e., also including parent interview, language sampling, dynamic assessment; ASHA, n.d.; Lewis et al., 2010). Scoring of these subtests was informed by the responses to test items of adult models from the same linguistic community. As such, we took the approach of conducting two separate studies examining JC–English bilinguals' performance in the English context, each with its own set of research questions (RQs), to meet our research purpose. Study I used a qualitative approach to formulate the foundational information needed to characterize adult linguistic patterns and compare them to child responses. To complement the more prompted level of data obtained using the CELF Preschool-2, we also examined children's spontaneous samples for morphosyntactic and lexical variations. Including these data provided further confirmatory evidence of the linguistic patterns identified in adult models. Qualitative analysis of linguistic patterns in Study I provided the necessary foundation for Study II to develop an adapted scoring procedure that considered JC linguistic features. Study II then used a quantitative approach to compare differences in preschoolers' subtest scaled scores between the standard English scoring protocol and the adapted JC scoring procedure. The application of a mixed-methods analysis provided the advantage of having a deeper understanding of the linguistic profile of the JC–English DLL child and to also inform adapted scoring procedures. The following RQs were addressed:

Study I

1. What are the patterns of response of JC-English–speaking adults for morphosyntax (i.e., Word Structure) and lexical items (i.e., Expressive Vocabulary)?
2. How do the patterns of responses of JC-English–speaking preschoolers for Word Structure and Expressive Vocabulary subtests compare to those of JC-English–speaking adults?

Study II

3. How does the performance of JC-English–speaking preschoolers' expressive language in English on a standardized measure of morphosyntax and vocabulary compare with the standardization sample?
4. Does adapted scoring that considers known linguistic features of JC change the standardized test scores for JC-English–speaking preschoolers?

Method—Study I

Ethical Approval

Study approval was obtained from the institutional review board of the University of Cincinnati. The Early Child Commission, Government of Jamaica, and each participating early learning center provided permission. Speech therapy licensure in Jamaica was obtained from the Council of Professions Allied to Medicine.

Participants

Adult and child participants in this study were drawn from the Jamaican Creole Language Project (cf. K. N. Washington et al., 2019, 2017) cohort assessed during 2013–2018. Participants in the Jamaican Creole Language Project were recruited from Kingston, Jamaica. Child participants were recruited from three public schools recommended by the Early Childhood Commission as containing speakers of JC and English. Characteristics of adult and child participants are further described in the following sections.

Adult Participants

Forty adults were invited to participate by way of flyers and through a community meeting. Informed consent was obtained, and data were collected from 33 JC-English–speaking adults aged 24–51 years who were from the same linguistic community as the children and completed the CELF Preschool-2. Adults were not parents of children participating in this study, reported no history of speech-language difficulties, and were either self-identified or identified by their employers as proficient speakers of JC and English. Adult participants were from varied socioeconomic backgrounds, as evidenced through education levels and current employment (see Table 1), with the majority (n = 30, 90.9%) being employed at the time of the study. Table 1 provides a summary of adult characteristics.

Child Participants

The parents of 214 children were invited to have their children participate in the Jamaican Creole Language Project using flyers and through parent–teacher meetings. Informed consent was obtained, and data were collected for 211 children. Data describing a subset of 176 preschoolers were extracted from the Jamaican Creole Language Project database based on the following inclusion criteria: (a) simultaneous DLL of JC–English, based on parent and teacher report obtained via questionnaires; (b) aged 4;0–5;11 (years; months); (c) passed a binaural hearing screening at 25 dB for 1, 2, and 4 kHz (using a MAICO MA 1 audiometer and headphones fitted with Peltor cups); (d) able to use spoken language; and (e) complete subtest data for the CELF Preschool-2. We applied a standard approach to our inclusion criteria that is used when developing standardized assessments. As such, our approach modeled standardized assessments of children's language (e.g., CELF Preschool-2) to ensure we had a representative sample of JC-English–speaking children. Thus, our sample included children with a range of speech and language capabilities. For the purposes of our study, we focused on preschoolers aged 4;0–5;11, as this age holds special sensitivity for informing the early and accurate identification of DLD in academic and clinical settings (Bishop & Edmundson, 1987; Paradis et al., 2011). Participants included in this study were aged 4;2–5;11 (M = 4;11) and included 82 boys and 94 girls from varied socioeconomic backgrounds, as indexed by maternal education levels and household income (see Table 1). Children represented the following age groups: 4;0–4;11 (n = 105; 50 boys, 55 girls) and 5;0–5;11 (n = 71; 31 boys, 40 girls).

Information about children's development and language use was obtained via questionnaires. Most preschoolers' parents (n = 141, 80.1%) reported little to no concerns about their child's talking, while 31 (17.6%) parents reported concerns and four (2.3%) parents did not respond. Parents reported that preschoolers used JC "very well" (n = 65, 36.9%), "somewhat well" (n = 65, 36.9%), and "not very well" (n = 21, 11.9%), and 25 (14.2%) did not respond. Parents reported that preschoolers used English "very well" (n = 46, 26.1%), "somewhat well" (n = 98, 55.7%), and "not very well" (n = 21, 11.9%), and 11 (6.3%) did not respond. Parents anecdotally reported that their perception of "very well" included "near perfect production" of the language, without code mixing. Preschoolers' classroom teachers also reported little to no concerns regarding most children's communication (n = 160,

Table 1. Child (n = 176) and adult (n = 33) participants' characteristics.

Variable	Child characteristics			Adult characteristics
	Total	TD	DLD	
Age				
M (*SD*)	4;11 (0;6)	4;10 (0;6)	4;11 (0;6)	38;10 (11;0)
Range	4;2–5;11	4;2–5;11	4;2–5;10	24;4–51;8
Sex				
Female	94	82	12	24
Male	82	74	8	9
Highest education level, *n* (%)		Maternal education		Own education
University	53 (30.1)	47 (30.1)	6 (30)	12 (36.4)
College	38 (21.6)	35 (22.4)	3 (15)	6 (18.1)
Trade school	19 (10.7)	17 (10.9)	2 (10)	2 (6.1)
High school	62 (35.2)	53 (34)	9 (45)	12 (36.4)
Primary school	2 (1.2)	2 (1.3)	0 (0)	1 (3)
No response	2 (1.2)	2 (1.3)	0 (0)	—
Household income, *n* (%)				
Employed	168 (95.5)	149 (95.5)	19 (95)	30 (90.9)
Dual income	122 (69.3)	107 (68.5)	15 (75)	—
Single income	46 (26.1)	42 (26.9)	4 (20)	—
Unemployed	4 (2.3)	4 (2.6)	0 (0)	3 (9.1)
No response	4 (2.3)	3 (1.9)	1 (5)	—

Note. Ages are displayed as years;months. TD = typically developing; DLD = developmental language disorder.

90.9%), while concerns were reported for 14 (8%) preschoolers, with no teacher response for two (1.1%) preschoolers. Consistent with previous works (cf. León et al., 2021), children were classified as typically developing or language disordered based on a consensus of concerns of (a) parents and (b) teachers or SLPs (Restrepo, 1998). As such, the informant answered the question, "Are you concerned about this child's talking?"—responding "yes/no/a little" (cf. Glascoe, 2000). Children who had a "yes" response from two sources for both languages were classified as language disordered to inform analysis procedures. This "two-source approach" classified 20 JC–English bilinguals as language disordered (i.e., 11.4%), offering an estimate that approximates the expected prevalence rate of DLD in bilingual children (i.e., 7%–10%; Kohnert, 2010; Nayeb et al., 2021).

Data were also extracted from the Jamaican Creole Language Project database (cf. K. N. Washington et al., 2019, 2017) concerning children's performance on measures of nonverbal intelligence (Primary Test of Nonverbal Intelligence [PTONI]; Ehrler & McGhee, 2008), oral motor skills (Diagnostic Evaluation of Articulation and Phonology [DEAP] Oral Motor Screening; Dodd et al., 2009), and intelligibility in English (Intelligibility in Context Scale [ICS]; McLeod et al., 2012a) and JC (ICS-JC; McLeod et al., 2012b) to further describe this study's sample. These participants had a mean PTONI score of 104.2 (SD = 16.36, range: 65–147). The majority of children (n = 173, 98.3%) met the age-based criterion according to the DEAP Oral Motor Screening. The majority of children showed age-appropriate speech intelligibility (cf. León et al.,2021)with a mean total score of 4.7 on the ICS for English (n = 132, 75%) and a mean total score of 4.6 on the ICS-JC (McLeod et al., 2012b) for JC (n = 110 out of 139 available scores due to ICS-JC availability, 79.1%). Table 1 also provides a summary of child participant characteristics.

Materials

CELF Preschool-2

The CELF Preschool-2 (Wiig et al., 2006) is a valid and reliable, norm-referenced, standardized language assessment of receptive and expressive language designed specifically for monolingual English-speaking preschoolers aged 3;0–6;11. This assessment is routinely used by SLPs for the identification, diagnosis, and follow-up evaluation of language deficits in preschool-age children (Finestack & Satterlund, 2018). CELF Preschool-2 subtests can be used individually to assess a specific language domain or collectively to provide information about a child's language abilities using receptive, expressive, or core language scores (Wiig et al., 2006). In this study, two expressive language subtests were administered to adults and children to identify linguistic patterns of responses for morphosyntactic (i.e., Word Structure) and lexical items (i.e., Expressive Vocabulary). These domains offer critical information to distinguish between overlapping linguistic features of language difference and DLD (Blom & Paradis, 2015) and to establish DLLs' oral language proficiency to guide academic placement (Castro et al., 2013; Greenwood et al., 2016). Children's morphosyntactic and vocabulary skills can also predict future literacy achievement and academic success (Greenwood et al., 2016). As such, these subtests were deemed as critical components within this standardized measure to characterize adults' and children's morphosyntactic and lexical linguistic patterns.

Language Sample Stimuli

As part of the Jamaican Creole Language Project (K. N. Washington et al., 2019, 2017), play-based language samples were collected for all children in English. For this study, a subset (n = 35, 20%) of children's language samples were analyzed to provide confirmatory evidence of the linguistic patterns identified in adult responses. Language sample stimuli in the English context included a castle, Mike the Knight set, Disney princesses, and toy food objects.

Procedure

As previously described, the Jamaican Creole Language Project (K. N. Washington et al., 2019, 2017) involves adult and preschooler speakers of JC and English who are from Kingston and surrounding areas in Jamaica. Adult participants provided written consent and completed a background questionnaire. For child participants, parental consent and child assent were obtained, and parents completed a questionnaire describing children's communication and development. Adults and children participated in a number of assessments in English and/or JC, depending on the task requirements of the Jamaican Creole Language Project (K. N. Washington et al., 2019, 2017). Performance on measures of children's nonverbal intelligence (i.e., PTONI), oral motor skills (i.e., DEAP Oral Motor Screening), and intelligibility in English (i.e., ICS) and JC (i.e., ICS-JC) is reported to describe the study sample and provide applicable contextual details. However, only assessments and procedures relevant to this study are described in detail in this article. Administration in English only is described, since this assessment context was needed to answer our RQs and to document potential misdiagnosis in the English context. All assessments were conducted in an authentic school environment for adults and children.

CELF Preschool-2 Administration

The Word Structure and Expressive Vocabulary subtests were administered to adults and children following instructions from the CELF Preschool-2 manual. The Word Structure subtest requires the participant to provide a one- or two-word response to complete a sentence using a particular syntactic frame as part of a cloze procedure (i.e., prompt: "This boy is standing" [photograph of boy standing], "This boy is _____" [photograph of boy sitting]).

The 24 items of this subtest evaluate morphological knowledge by requiring the participant to use inflections, derivations, comparisons, possessive forms, and pronouns to refer to people/objects. The Expressive Vocabulary subtest requires the participant to name pictured stimuli in response to a verbal prompt. The 20 items of this subtest evaluate expressive vocabulary via referential naming and labeling of people, objects, and actions (Wiig et al., 2006). For the purposes of this study, in administration, basal rules were applied (i.e., subtest basal rules indicated to start at the first item), while ceiling rules were not so that all subtest items would be administered and responses to all items would be collected. This approach was applied to ensure all necessary data needed to address original and adapted scoring methodological approaches (see Study II) were available.

Consistent with the Jamaican Creole Language Project (K. N. Washington et al., 2019, 2017), the English CELF Preschool-2 administration was provided by an English-speaking SLP unknown to the adult and child participants. Elicitors used only the target language (i.e., English) as a strategy to encourage use of the target language during assessment (Paradis et al., 2011). For example, at the beginning of assessment, examiners said, "Remember I want you to talk to me in English" (cf. Morren & Morren, 2007; K. N. Washington et al., 2019). This strategy was used to encourage productions that would most closely align with the English targets of the assessment. For adults, each subtest took approximately 5–7 min to complete. For children, each subtest took approximately 7–10 min to complete. All adults completed the subtests on the same day, while most children completed both subtests on the same day. All participants' responses to the Word Structure and Expressive Vocabulary subtests were video- and audio-recorded.

Child Language Samples

Consistent with the Jamaican Creole Language Project (K. N. Washington et al., 2019, 2017), all children completed a 15-min, play-based language sample in English that was video- and audio-recorded. A randomly selected subset of samples (i.e., $n = 35$, 20%) was analyzed for the purposes of this article to offer confirmatory evidence about children's morphosyntactic and lexical language. By also examining spontaneous productions, described as a gold standard in language assessment, we applied a contextually driven approach to provide detailed information about children's language capabilities (Ebert & Scott, 2014). Consequently, preschoolers' use of linguistic features in this context offered confirmatory evidence of the validity of the linguistic features produced in the more prompted, standardized assessment context. Children were prompted to provide a language sample using language-specific stimuli (e.g., castle, princesses). The English-speaking SLP encouraged spontaneous productions by making a statement such as, "Look what we see here. Let's see what we can do with these things," followed by prompts to facilitate interaction (e.g., following the child's lead, asking open-ended questions, commenting). Language samples were transcribed, consistent with the protocols in K. N. Washington et al. (2019).

Data Analysis

Adult Responses

To answer RQ 1, adult responses to the CELF Preschool-2 subtests were analyzed using content analysis, a qualitative method that utilizes systematic analysis to reveal the presence of themes, providing interpretation of meaning from the text (Neundorf, 2002; Weber, 1990). This approach employed foundational tenets of linguistic contrast analysis. However, since traditional linguistic contrast analysis classifies linguistic patterns as being errored relative to a monolingual standard (cf. Khansir & Pakdel, 2019), we used qualitative methodology to facilitate the interpretation of linguistic themes based on a nondeficit approach (i.e., language use as informed by adult speakers of the language community). Transcribed adult responses were first analyzed to establish themes in responses to CELF Preschool-2 items regarding linguistic features in the JC–English bilingual context. These themes were established by three native English speakers who received specific training concerning the linguistic features of JC (cf. K. N. Washington et al., 2019) and then confirmed by the second author, a bilingual JC–English speaker. A number of themes were established: (a) JC-influenced morphological structure (e.g., *sleep* for *sleeps*; Word Structure), (b) JC-influenced lexical variations (e.g., *drop* for *fell*, *gleaner* for *newspaper*; Word Structure and Expressive Vocabulary), and (c) use of functional description (e.g., *to see from afar* for *binoculars*; Expressive Vocabulary). Importantly, lexical variation and the use of functional description were identified as two separate linguistic themes as documented linguistic patterns evidenced in Jamaican culture were considered (Nero & Stevens, 2018).

Child Responses

After themes were established from adult responses, child responses to the CELF Preschool-2 subtests were coded for each of the three linguistic themes noted above to answer RQ 2. Thirty-five (20%) children's language samples were also analyzed to provide confirmatory evidence of the linguistic themes established via content analysis of adult responses. Productions in these language samples were coded using the themes identified in the adult CELF Preschool-2 responses to establish if these themes were also present in children's spontaneous language in English. For example, the production "dem are sleeping" was coded as containing JC-influenced morphological structure due to the use of the JC subject pronoun "dem" for "they" in the English context. This analysis provided confirmatory evidence of potential corresponding linguistic variation in children's communication across communicative contexts.

Reliability

Interrater agreement for coded themes in adult and child samples was established using a kappa coefficient to account for chance agreement between independent raters. The kappa coefficient is useful to document agreement of categorical data and uses the following kappa coefficient scale: less than chance agreement (< .01), slight agreement (.01–.20), fair agreement (.21–.40), moderate agreement (.41–.60), substantial agreement (.61–.80), and almost perfect agreement (.81–.99; Viera & Garrett, 2005). For adults, 100% of the coded productions (n = 242) were analyzed. Kappa statistics revealed "almost perfect agreement" between independent raters for all themes: (a) JC-influenced morphological structure, κ = .92, 95% confidence interval (CI) [.87, .97], p < .001; (b) JC-influenced lexical variation, κ = .86, 95% CI [.79, .93], p < .001; and (c) use of functional description, κ = .81, 95% CI [.64, .99], p < .006. For children, 20% of the CELF Preschool-2 coded productions (n = 394) were analyzed. Kappa statistics revealed "almost perfect agreement" between independent raters for all themes: (a) JC-influenced morphological structure, κ = .83, 95% CI [.77, .89], p < .001; (b) JC-influenced lexical variation, κ = .81, 95% CI [.72, .88], p < .001; and (c) use of functional description, κ = .89, 95% CI [.78, .99], p < .001. Lastly, 100% of coded productions (n = 382) in children's language samples were analyzed. Kappa statistics revealed "almost perfect agreement" between independent raters for JC-influenced morphological structure, κ = .93, 95% CI [.89, .97], p < .001; "substantial agreement" for JC-influenced lexical variation, κ = .71, 95% CI [.62, .81], p < .001; and "almost perfect agreement" for use of functional description, κ = .81, 95% CI [.64, .99], p < .006.

Results—Study I

RQ 1: Patterns of Linguistic Features in JC–English Adults

A content analysis was conducted to identify linguistic features used by JC–English bilinguals that may influence responses to items on the CELF Preschool-2. The presence of themes and the frequency of the themes' occurrence varied between the two subtests, which could be impacted by the nature of the subtests' target language domain.

Word Structure

Across all adult responses (n = 792) to the CELF Preschool-2 Word Structure subtest, 157 responses were identified as JC-influenced and were then thematically coded according to the linguistic features of the response. Coding of adult responses revealed the following: (a) JC-influenced morphological structure coded at 15.2% (n = 120/792) and (b) JC-influenced lexical variation coded at 4.7% (n = 37/792). Use of functional description was not a coded linguistic feature in the subtest. Findings and exemplars of adult responses are presented in Table 2.

Expressive Vocabulary

Across all adult responses (n = 660) to the CELF Preschool-2 Expressive Vocabulary subtest, 85 responses were identified as JC-influenced and coded according to the linguistic features of the response. Coding of adult responses revealed the following: (a) JC-influenced lexical variation coded at 11.4% (n = 75/660) and (b) use of functional description coded at 1.5% (n = 10/660). JC-influenced morphological structure was not a coded linguistic feature in this subtest (see Table 2).

RQ 2: Comparison of Patterns of Linguistic Features for JC–English DLLs and Adults

Coding of children's responses by way of the established linguistic themes from the content analysis of adult models revealed that JC–English DLLs demonstrate similar patterns of linguistic features as adults in both standardized and naturalistic assessment contexts.

Word Structure

Across all children's responses (n = 4,224) on the CELF Preschool-2 Word Structure subtest, 1,434 responses were identified and coded for the adult-informed linguistic themes. Coding of child responses revealed the following: (a) JC-influenced morphological structure coded at 26.4% (n = 1,115/4,224) and (b) JC-influenced lexical variation coded at 7.6% (n = 319/4,224). Consistent with adults, use of functional description was not a coded linguistic feature in this subtest. Findings and exemplars of children's responses are also presented in Table 2. A complete list of alternate responses for this subtest and percentage of use can be found in Supplemental Material S1.

Expressive Vocabulary

Across all children's responses (n = 3,520) on the CELF Preschool-2 Expressive Vocabulary subtest, 535 responses were identified and coded for the adult-informed linguistic themes. Coding of child responses revealed the following: (a) JC-influenced lexical variation coded at 12.3% (n = 434/3,520) and (b) use of functional description coded at 2.9% (n = 101/3,520). Consistent with adults, JC-influenced morphological structure was not a coded linguistic feature in this subtest (see Table 2). See Supplemental Material S2 for a complete list of alternate responses and percentage of use for this subtest.

Language Samples

The linguistic patterns identified in the standardized assessment context were also present in the subset

Table 2. Themes for adult and child responses to subtests and exemplars.

Subtest	Theme	CELF Preschool-2 responses *n* (%)		Exemplars *(target)*
		Adults	Children	
Word Structure	1: JC-influenced morphological structure	120 (15.2%)	1,115 (26.4%)	Climb; used to climb *(climbed)* Falls; fallen; falled *(fell)* Fast *(faster)* Fly; flying *(flies)* He is *(they are)* Her *(hers)* Him standing *(he is standing)* Horse; two horse *(horses)* King; king crown *(king's)* Sleep; sleeping *(sleeps)*
	2: JC-influenced lexical variations	37 (4.7%)	319 (7.6%)	Burst the bubble (*blew*) Dropped (*fell*) Is climbing; will climb (*will slide)* Musician; entertainer *(singer)* Pon the chair *(on the chair)*
	3: Use of functional description	0 (0%)	0 (0%)	None noted
Expressive Vocabulary	1: JC-influenced morphological structure	0 (0%)	0 (0%)	None noted
	2: JC-influenced lexical variations	75 (11.4%)	434 (12.3%)	Almanac *(calendar)* Animal doctor *(veterinarian)* Book; gleaner; news *(newspaper)* Camera; eye scope; spy glass *(telescope)* Congregation; crowd *(audience)* Foot; foot bottom; footstep *(footprint)* Logo; sticker *(stamp)* Medal *(trophy)* Stem *(branch)* Throwing out the milk *(pouring milk)*
	3: Use of functional description	10 (1.5%)	101 (2.9%)	Something to spy with *(binoculars*) Tells the month of the year *(calendar)* Test your weight *(scale)* Use to spy *(telescope)*

Note. The total quantity (*n*) of the Clinical Evaluation of Language Fundamentals Preschool–Second Edition (CELF Preschool-2; Wiig et al., 2006) coded responses per linguistic theme across participants is represented in this table. Percentages were derived from the proportion of the number of coded responses for the linguistic theme and the total number of all responses in the subtest. JC = Jamaican Creole.

of children's spontaneous language samples in English (n = 35) that were analyzed to provide confirmatory evidence of the linguistic themes. Across these 35 English language samples (n = 2,417 total productions), 382 JC-influenced productions were identified and coded for the themes distilled from the adult samples. Coding of child language sample productions revealed the following: (a) JC-influenced morphological structure coded at 13.1% (n = 316/2,417), (b) JC-influenced lexical variation coded at 2.3% (n = 56/2,417), and (c) use of functional description coded at 0.4% (n = 10/2,417).

Method—Study II

Study II involves the same child participants, materials, and procedures as described in Study I. The analysis and results described in Study I provided the foundation to inform scoring procedures developed in Study II. These scoring procedures that are unique to Study II will be further described below.

Procedure

CELF Preschool-2 Original Scoring Procedure

Scoring of children's responses on the CELF Preschool-2 subtests was achieved using the original CELF Preschool-2 standard English scoring procedure specified in the test manual (Wiig et al., 2006). Items from the Word Structure subtest were scored: "1" for a correct target response/structure and "0" for an incorrect response. Items from the Expressive Vocabulary subtest were scored: "2" for target response or appropriate substitute, "1" for a response related to the target response, and "0" for incorrect response. Importantly, the CELF Preschool-2 recommends modified scoring when used with linguistically diverse children; however, specific patterns of performance are not currently available in the user manual to inform the scoring

of JC–English speakers' language productions. Basal and ceiling rules consistent with the user manual were applied. Item scores within each subtest were summed to form a raw score for each subtest. Raw scores were converted to subtest scaled scores to compare performance to the standardized sample. According to the CELF Preschool-2 manual, the majority of children (i.e., 68.3%) should score within 1 *SD* above or below the mean (i.e., subtest scaled scores of 7–13, language classification ranging from mild deficit to high average), while very few children (i.e., 15.9%) should score greater than 1 *SD* below the mean (i.e., subtest scaled scores 0–6, classification of language disordered; Wiig et al., 2006, p. 69).

Adapted CELF Preschool-2 Scoring Procedure

An adapted scoring procedure for the Word Structure and Expressive Vocabulary subtests of the CELF Preschool-2 was developed. This procedure was informed by the responses of JC-English–speaking adults to subtest items in Study I. These themes and corresponding examples were confirmed by Professor Hubert Devonish, the former Chair of the Jamaican Language Unit (i.e., a language planning agency with expertise in JC linguistic structure and language practices), to ensure accuracy in the rescoring procedures. For example, for Word Structure, if a preschooler produced a + root verb (e.g., "a sliip") for the present progressive structure (i.e., English target of "is sleeping"), it was rescored as a correct production due to the JC morphosyntactic influence (i.e., the present tense continuative aspect) on the English production. For Expressive Vocabulary, the response "gleaner" would be considered correct for the target "newspaper," considering lexical differences of JC. Children's responses were examined across all subtest items to conduct rescoring procedures. Adapted raw scores were calculated adhering to basal and ceiling rules. Specifically, the start item remained (basal), but a new ceiling was established based on our adapted scoring procedures. For example, we established a secondary ceiling due to the number of consecutive errors (i.e., consistent with the scoring manual, but took into account items that were previously considered to be errors but were now considered to be accurate) or the ceiling equaled the last item on the subtest. Adapted raw scores using the adapted scoring procedure were then converted to the adapted subtest scaled score for interpretation (Wiig et al., 2006).

Reliability

Interrater reliability for scoring of the CELF Preschool-2 subtests was established using a kappa coefficient (Viera & Garret, 2005) to account for chance agreement between independent raters. Twenty percent of samples ($n = 70$) from each subtest were analyzed. For original scoring, kappa statistics revealed "almost perfect agreement" between independent raters for Word Structure, $\kappa = .85$, 95% CI [.75, .95], $p < .001$, and Expressive Vocabulary, $\kappa = .93$, 95% CI [.86, 1.0], $p < .001$. For adapted scoring, kappa statistics also revealed "almost perfect agreement" between independent raters for Word Structure, $\kappa = .92$, 95% CI [.84, 1.0], $p < .001$, and Expressive Vocabulary, $\kappa = .85$, 95% CI [.75, .95], $p < .001$.

Data Analysis

All data were analyzed using SPSS Version 25. To answer RQ 3, children's CELF Preschool-2 subtest original scaled scores, obtained using the standard English scoring procedure, were statistically compared with the CELF Preschool-2 normative sample using one-sample *t* tests. The one-sample *t* test allows for statistical comparison when the mean and standard deviation of a normative population are known and used for comparison to a collected sample in which the mean and standard deviation are initially unknown (Altman, 1991). A Kolmogorov–Smirnov test of normality with Lilliefors correction was used to examine the distribution of original scores for each subtest. To answer RQ 4, children's performance on the CELF Preschool-2 subtests was rescored using the adapted JC scoring procedure informed by adult models. We used multiple statistical approaches to address this RQ. First, we conducted one-sample *t* tests to compare adapted scaled scores that considered JC with the CELF Preschool-2 normative sample scaled scores. A Kolmogorov–Smirnov test of normality with Lilliefors correction was also used to examine the distribution of adapted scores for each subtest. Classification accuracy statistics were then used to inform the diagnostic accuracy of adapted scoring procedures. Lastly, we conducted one-way analyses of variance (ANOVAs) for each subtest to statistically compare JC–English DLLs' original and adapted CELF Preschool-2 scaled scores.

Results—Study II

RQ 3: Comparison of JC–English DLLs' Scores to CELF Preschool-2 Normative Data

We addressed RQ 3 by comparing the original subtest scaled scores (i.e., those using the standard English scoring procedure) from our sample of JC–English DLLs to the CELF Preschool-2 normative sample means. To provide the most appropriate comparison, we used data reported in the CELF Preschool-2 manual that identified children's subtest performance by age (Wiig et al., 2006, p. 116). Table 3 provides descriptive statistics for our sample and the CELF Preschool-2 normative sample.

CELF Preschool-2 Word Structure Subtest

Word Structure original subtest scaled scores for 4-year-old ($M = 7.3$, $SD = 2.6$) and 5-year-old ($M = 6.9$, $SD = 3.1$) participants were more than 1 *SD* below the mean of the CELF Preschool-2 normative sample of 4-year-olds ($M = 10.6$, $SD = 3.2$) and 5-year-olds ($M = 10.1$, $SD = 2.8$). One-sample *t* tests indicated significant differences between JC–English DLLs' subtest scaled scores and the

Table 3. Study sample and the Clinical Evaluation of Language Fundamentals Preschool–Second Edition (CELF Preschool-2) normative sample scaled scores on CELF Preschool-2 Word Structure and Expressive Vocabulary subtests.

Subtest	Age group	Normative sample	JC–English original scoring	JC–English adapted scoring
		M (*SD*)	*M* (*SD*)	*M* (*SD*)
Word Structure	4-year-olds	10.6 (3.2)	7.3 (2.6)*	11.9 (3.04)*
	5-year-olds	10.1 (2.8)	6.9 (3.1)*	11.6 (3.3)*
	All children	10.6 (2.6)	7.1 (2.8)*	11.7 (3.1)**
Expressive Vocabulary	4-year-olds	10.8 (2.8)	8.6 (2.3)*	10.9 (2.4)
	5-year-olds	10.6 (3.3)	7.9 (2.6)*	10.1 (2.3)
	All children	10.9 (2.7)	8.3 (2.4)*	10.6 (2.4)**

Note. The data reported for the CELF Preschool-2 normative sample are a subset of the sample that described children's performance by age (Wiig et al., 2006, p. 116). This was used as the point of reference as the CELF Preschool-2's complete normative sample included children aged 3;0–6;11 (years;months). JC = Jamaican Creole.

*Statistically significant difference between JC–English dual language learners and normative sample means, $p < .001$.
**Statistically significant difference between original and adapted score means, $p < .001$.

CELF Preschool-2 normative sample for 4-year-olds, $t(103) = -12.94$, $p < .001$, $d = 1.3$, and 5-year-olds, $t(69) = -8.68$, $p < .001$, $d = 1.1$ (Cohen, 1988). The 95% CI for the difference in means ranged from −3.79 to −2.78 (4-year-olds) and from −3.97 to −2.51 (5-year-olds), suggesting that these differences did not occur by chance. Using the CELF Preschool-2 manual scoring guidelines, children's performance was classified as follows: high average ($n = 8$, 4.5%), average ($n = 72$, 40.9%), mild deficit ($n = 10$, 5.7%), and language disordered ($n = 86$, 48.9%). Application of a one-sample Kolmogorov–Smirnov test of normality with Lilliefors correction suggested that the distribution of original scores did not approximate the normal curve, K-S(176) = .097, $p < .001$, and was negatively skewed (i.e., due to the distribution of children's original scores among the diagnostic categories). This observation is in line with previous research indicating a skewed distribution of scores when using traditional standardized assessments with linguistically diverse children (i.e., 53.5%–84% of children were classified as language disordered; Barragan et al., 2018; Pearce & Williams, 2013; J. A. Washington & Craig, 1992).

Children who scored within the average and high average range on this subtest displayed the following characteristics: More were from dual-income ($n = 55$, 68.8%) than single-income ($n = 21$, 26.3%) homes, an equivalent number were male ($n = 40$, 50%) and female ($n = 40$, 50%), and half scored within the above average to very superior range on the PTONI ($n = 40$, 50%). Children who scored within the mild deficit and language disordered range on this subtest displayed the following characteristics: More were from dual-income ($n = 66$, 68.9%) than single-income ($n = 29$, 30.2%) homes, more were female ($n = 53$, 55.2%) than male ($n = 43$, 44.8%), and relatively few scored within the above average to very superior range on the PTONI ($n = 20$, 20.8%).

CELF Preschool-2 Expressive Vocabulary Subtest

Expressive Vocabulary subtest scaled scores for 4-year-old ($M = 8.6$, $SD = 2.3$) and 5-year-old ($M = 7.9$, $SD = 2.6$) participants were slightly less than 1 *SD* below the mean of the CELF Preschool-2 normative sample of 4-year-olds ($M = 10.8$, $SD = 2.8$) and 5-year-olds ($M = 10.6$, $SD = 3.3$). One-sample *t* tests indicated significant differences between JC–English DLLs' subtest scaled scores and the CELF Preschool-2 normative sample for 4-year-olds, $t(103) = -10.18$, $p < .001$, $d = 0.9$, and 5-year-olds, $t(69) = -8.89$, $p < .001$, $d = 1.1$ (Cohen, 1988). The 95% CI for the difference in means ranged from −2.69 to −1.81 (4-year-olds) and from −3.34 to −2.11 (5-year-olds), suggesting that differences did not occur by chance. Using the CELF Preschool-2 manual scoring guidelines, children's performance was classified as follows: high average ($n = 16$, 9.1%), average ($n = 97$, 55.1%), mild deficit ($n = 23$, 13.1%), and language disordered ($n = 40$, 22.7%). Application of a one-sample Kolmogorov–Smirnov test of normality with Lilliefors correction suggested that the distribution of original scores did not approximate the normal curve, K-S(176) = .071, $p = .03$, and was negatively skewed (i.e., due to the distribution of children's original scores among the diagnostic categories). This finding is in line with the aforementioned research documenting an increased percentage of bilingual and bidialectal children being classified as language disordered when using traditional standardized scoring (cf. Barragan et al., 2018; Hendricks & Adlof, 2017; Pearce & Williams, 2013; J. A. Washington & Craig, 1992).

Children who scored within the average and high average range on this subtest displayed the following characteristics: More were from dual-income ($n = 82$, 73.9%) than single-income ($n = 22$, 19.8%) homes, slightly more were male ($n = 56$, 50.5%) than female ($n = 55$, 49.5%), and slightly less than half scored within the above average to very superior range on the PTONI ($n = 46$, 41.4%). Children who scored within the mild deficit and language disordered range on this subtest displayed the following characteristics: More were from dual-income ($n = 42$, 64.6%) than single-income ($n = 22$, 33.8%) homes, more were female ($n = 39$, 60%) than male ($n = 26$, 40%), and relatively few scored

within the above average to very superior range on the PTONI (n = 15, 23.1%).

RQ 4: Adapted Scoring of the CELF Preschool-2 Considering JC Linguistic Features

We addressed RQ 4 using multiple statistical approaches. We first compared the adapted subtest scaled scores (i.e., using the adapted scoring procedure) from our sample of JC–English DLLs to the CELF Preschool-2 normative sample means (see Table 3). We then used classification accuracy statistics to inform the diagnostic accuracy of adapted scoring procedures. Lastly, one-way ANOVAs were used to determine if statistically significant differences were present between JC–English DLLs' original and adapted CELF Preschool-2 subtest scaled scores.

Comparison of JC–English DLLs' Adapted Scores to CELF Preschool-2 Normative Data

CELF Preschool-2 Word Structure Subtest

On this subtest, 22 out of 24 (91.7%) test items received adapted scoring (see Supplemental Material S1). Word Structure adapted subtest scaled scores for 4-year-old (M = 11.9, SD = 3.04) and 5-year-old (M = 11.6, SD = 3.3) participants were higher than the CELF Preschool-2 normative sample of 4-year-olds (M = 10.6, SD = 3.2) and 5-year-olds (M = 10.1, SD = 2.8). One-sample t tests indicated significant differences between JC–English DLLs' adapted subtest scaled scores and the CELF Preschool-2 normative sample for 4-year-olds, $t(103) = 4.24$, $p < .001$, $d = 0.4$, and 5-year-olds, $t(69) = 3.83$, $p < .001$, $d = 0.5$ (Cohen, 1988). The 95% CI for the difference in means ranged from 0.67 to 1.84 (4-year-olds) and from 0.71 to 2.25 (5-year-olds), suggesting that these differences did not occur by chance. Using the CELF Preschool-2 manual scoring guidelines, children's performance was classified as high average (n = 79, 44.8%), average (n = 80, 45.5%), mild deficit (n = 4, 2.3%), and language disordered (n = 13, 7.4%). Application of a one-sample Kolmogorov–Smirnov test of normality with Lilliefors correction suggested that the distribution of adapted scores did not approximate the normal curve, K-S(176) = .152, $p < .001$, and was positively skewed (i.e., due to the distribution of children's adapted scores among the diagnostic categories). This observation is in line with previous research that suggests adapted scoring may not fully account for other factors, such as SES, that can impact children's performance (cf. Gross et al., 2014) and may also be impacted by the features of the language pairing (cf. Hemsley et al., 2010). The consideration of the language pairing and contextual factors that inform this finding in this study will be expanded upon in the Discussion section.

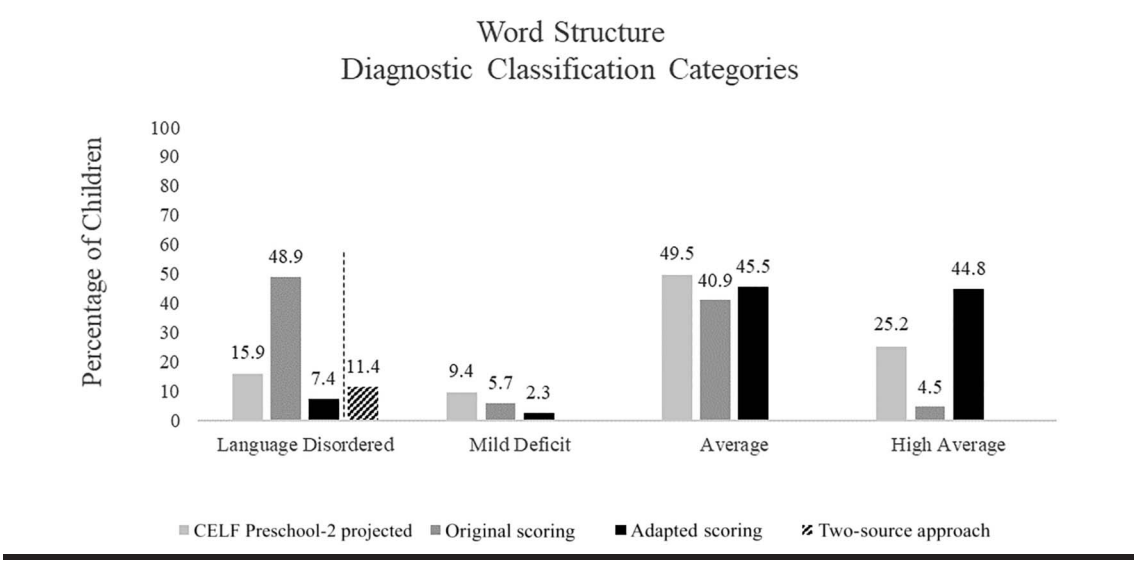
To address potential concerns related to underdiagnosis of disorder using adapted scoring (cf. Oetting et al., 2019), classification accuracy was informed using the gold-standard approach of "reporting of concerns from two sources" (i.e., parent and teacher or SLP; Restrepo, 1998). Adapted scoring for this subtest yielded a classification accuracy rate of 93.8% with an adequate level (.85) of sensitivity (i.e., a true positive) and a high level (.94) of specificity (i.e., a true negative; Plante & Vance, 1994). Ultimately, application of this gold-standard approach to identify true positive and true negative classification revealed a more accurate representation of the percentage of preschoolers categorized as being language disordered (i.e., 11.4% vs. 7.4% using adapted scoring). Figure 1 provides an illustration of diagnostic classification as informed by children's performance based on CELF Preschool-2 manual projections, original scoring, and adapted scoring across classification categories. For the language disordered category, we also illustrate classification as informed by the "two-source approach," which was used to address concerns related to underdiagnosis of language disorder based on adapted scoring alone.

CELF Preschool-2 Expressive Vocabulary Subtest

On this subtest, 16 out of 20 (80%) test items received adapted scoring (see Supplemental Material S2). Expressive Vocabulary adapted subtest scaled scores for 4-year-old (M = 10.9, SD = 2.4) and 5-year-old (M = 10.1, SD = 2.3) participants were comparable to the CELF Preschool-2 normative sample of 4-year-olds (M = 10.8, SD = 2.8) and 5-year-olds (M = 10.6, SD = 3.3). One sample t tests indicated no significant differences between JC–English DLLs' adapted subtest scaled scores and the CELF Preschool-2 normative sample for 4-year-olds, $t(103) = 0.49$, $p = .625$, $d = 0.1$, and 5-year-olds, $t(69) = -1.91$, $p = .06$, $d = 0.2$ (Cohen, 1988). The 95% CI for the difference in means ranged from −0.35 to 0.58 (4-year-olds) and from −1.08 to 0.22 (5-year-olds), suggesting that these findings did not occur by chance. Using the CELF Preschool-2 manual scoring guidelines, children's performance was classified as high average (n = 57, 32.4%), average (n = 102, 57.9%), mild deficit (n = 9, 5.1%), and language disordered (n = 8, 4.6%). Application of a one-sample Kolmogorov–Smirnov test of normality with Lilliefors correction suggested that the distribution of adapted scores approximated the normal curve, K-S(176) = .068, $p = .05$. This is in line with previous research documenting a change in the distribution of bilingual and bidialectal children's scores that more closely approximated to expectations of the normative curve when using adapted scoring compared to traditional standardized scoring (cf. Craig et al., 2004; Gross et al., 2014).

Similar to the approach used for the Word Structure subtest, concerns related to underdiagnosis using adapted scoring were addressed using the aforementioned gold-standard approach (cf. Restrepo, 1998) to inform classification accuracy. Adapted scoring for this subtest yielded a classification accuracy rate of 92.1%, with an adequate level (.88) of sensitivity (i.e., a true positive) and a high level (.92) of specificity (i.e., a true negative; Plante & Vance, 1994). Application of this gold-standard approach to identify true positive and true negative classification revealed a more accurate representation of the percentage of preschoolers categorized as being language disordered (i.e., 11.4% vs. 4.6% using adapted scoring). Figure 2 provides an illustration

Figure 1. Children's diagnostic classification categories on the Clinical Evaluation of Language Fundamentals Preschool–Second Edition (CELF Preschool-2) Word Structure subtest. The "two-source approach" is illustrated for the language disorders category: adapted scoring sensitivity (i.e., true positive) = .85; adapted scoring specificity (i.e., true negative) = .94.



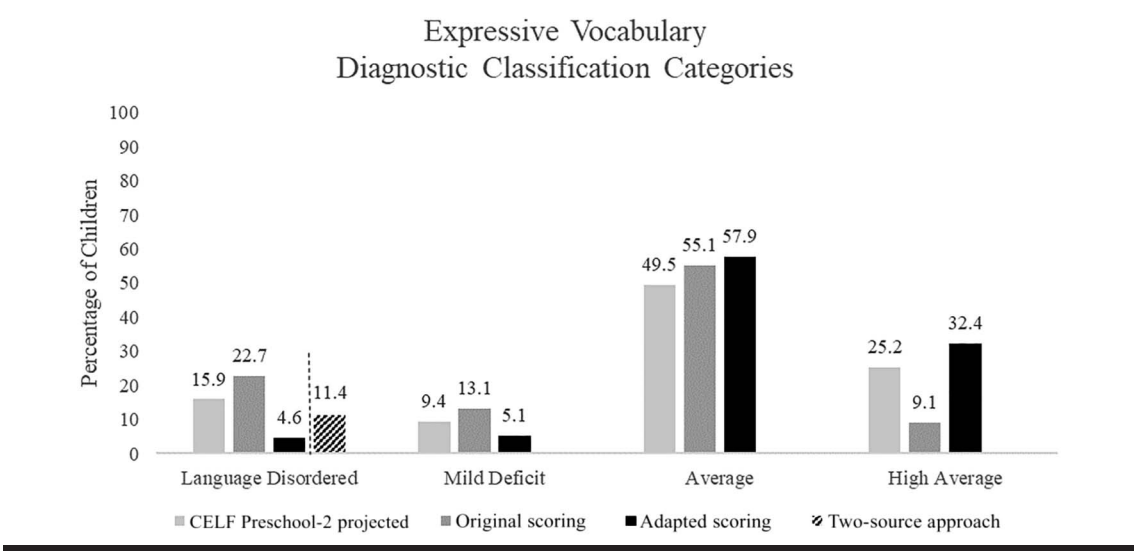
of diagnostic classification as informed by children's performance based on CELF Preschool-2 manual projections, original scoring, and adapted scoring across classification categories. For the language disordered category, we also illustrate classification as informed by the "two-source approach," which was used to address concerns related to underdiagnosis of language disorder based on adapted scoring alone.

Comparison of CELF Preschool-2 Original and Adapted Scores for JC–English DLLs

Word Structure Subtest

A one-way ANOVA was employed to account for differences in original subtest scaled scores using the standard English scoring procedure and adapted subtest scaled scores using the adapted scoring procedure. Levene's test

Figure 2. Children's diagnostic classification categories on the Clinical Evaluation of Language Fundamentals Preschool–Second Edition (CELF Preschool-2) Expressive Vocabulary subtest. The "two-source approach" is illustrated for the language disorders category: adapted scoring sensitivity (i.e., true positive) = .88; adapted scoring specificity (i.e., true negative) = .92.



of homogeneity of variances revealed that the assumption of homogeneity of variances was violated ($p = .17$); thus, Welch's ANOVA was used to protect against Type I error. Original subtest scaled scores were significantly different from adapted subtest scaled scores, Welch's $F(1, 346.29) = 179.24$, $p < .001$, $\omega^2 = 0.38$, indicating that the mean of original subtest scaled scores ($M = 7.1$, $SD = 2.8$) was significantly lower than the mean of adapted subtest scaled scores ($M = 11.7$, $SD = 3.1$), with a large effect size (Kirk, 1996).

Expressive Vocabulary Subtest

A one-way ANOVA was employed to account for differences in original subtest scaled scores using the standard English scoring procedure and adapted subtest scaled scores using the adapted scoring procedure. There was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .78$). Original subtest scaled scores were significantly different from adapted subtest scaled scores, $F(1, 350) = 80.25$, $p < .001$, $\omega^2 = 0.18$, indicating that the mean of original subtest scaled scores ($M = 8.3$, $SD = 2.4$) was significantly lower than the mean of adapted subtest scaled scores ($M = 10.6$, $SD = 2.4$), with a large effect size (Kirk, 1996). See Table 3 for JC–English DLLs' performance compared to the CELF Preschool-2 normative sample across scoring procedures.

Discussion

The growing presence of DLLs on SLPs' caseloads necessitates a corresponding change in assessment practices, which have been historically benchmarked by studies based on native and monolingual speakers of the ambient language (Guiberson, 2020; Guiberson & Ferris, 2019; Pearce & Williams, 2013; Skahan et al., 2007). This change is needed to reduce potential misdiagnosis of disorder in children who do not share the same linguistic and cultural profile of the SLP and to improve adapted scoring procedures that may also result in underidentification (Hendricks & Adlof, 2017; Oetting et al., 2019). In the current study, we addressed this concern by conducting the first investigation documenting the appropriateness of standardized assessments for JC–English DLLs, by way of adult models from the same linguistic community, to inform assessment procedures. A total of 176 JC-English–speaking DLLs and 33 adult JC–English speakers were included across two studies. First, we identified patterns of linguistic features in JC-English–speaking adults' responses and compared them to JC-English–speaking DLLs' responses (Study I). Second, the data from Study I were used to develop an adapted scoring procedure that considered JC linguistic features that inform the distinction between difference and disorder in JC-English–speaking DLLs' expressive language (Study II). With these data, we offer a critical contribution to the body of literature to improve SLPs' cultural competence and responsivity for working with the JC–English DLL child and to make tools available to guide practice (i.e., adapted scoring procedures). As such, this study is responsive to the critical need for innovative ideas and research practices to inform SLPs' clinical practice for DLLs (cf. Guiberson, 2020, for a discussion).

Linguistic Patterns of JC–English Speakers in the English Assessment Context

Study I addressed two RQs to provide information about linguistic patterns of JC–English speakers. We made use of adult models from the same linguistic community as the JC–English DLLs to analyze linguistic features using qualitative methods. Our content analysis revealed that, in the English context, adult speakers produced JC-influenced linguistic features that were also observed in JC–English DLLs' productions in both standardized and naturalistic settings. We interpret this pattern of linguistic consistency to mean that the patterns of morphosyntactic and lexical use in children are a representation of the JC language structure rather than being developmental in nature. Stated differently, we found that JC linguistic features used by adults and children in the English context consisted of morphological, lexical, and descriptive variations (see Table 2). Using qualitative analysis, we found that these variations were rule-governed and systematic forms of cross-linguistic influence, meeting the definition of *difference*, not *disorder*, used in SLPs' educational and clinical practices (ASHA, n.d.; Paradis et al., 2011). Although JC and English have many structural (e.g., subject–verb–object sequences) and lexical (i.e., JC comprises approximately 90% English vocabulary) similarities consistent with the historical relationship between the languages (Deuber, 2009; Devonish & Harry, 2008; Trudgill, 2011), differences in JC morphosyntactic rules and lexical varieties impacted JC–English speakers' productions in the English context (cf. K. N. Washington et al., 2019).

These differences are of clinical relevance for both assessment and treatment. For example, morphosyntax and vocabulary were recently reported as the most often targeted goals in treatment, with an overwhelming percentage of goals (approximately 80.5%) targeting verb tense use (Selin et al., 2019). Specifically, present progressive, regular past tense, and irregular past tense verbs as well as plural nouns are reported as the most common grammatical forms targeted by SLPs in early educational settings (Finestack & Satterlund, 2018). These current practice patterns pose significant risk to JC–English DLLs for both misdiagnosis and selection of inappropriate treatment targets. This is because the absence of bound morphemes and differences in tense marking in JC influences language use in the English context for adults and children (e.g., *sleep* vs. *sleeps* [third-person singular], *climb* vs. *climbed* [regular past tense], *fall* vs. *fell* [irregular past tense], *horse* vs. *horses* [plural]). Percentage use of the linguistic themes reflected in Supplemental Materials S1 and S2 further document this risk of misdiagnosis (e.g., 73% of children's responses exhibited zero marking of the plural). Furthermore, though adults and children yielded similar linguistic patterns, the higher percentage of use in children's patterns, specifically regarding morphological structure, likely

reflects developmental changes of importance to be considered in the assessment context. As a result, cautious interpretation of the absence of these forms must be applied in assessment and when identifying appropriate treatment targets for JC–English DLLs, similar to that of other language pairings (Paradis et al., 2011) or nonmainstream English varieties (Lee & Oetting, 2014) so that SLPs can appropriately detail patterns of language difference and disorder.

Utility of Adult Models to Inform the English Assessment Context

Study II involved the analyses of child responses using adult models provided in Study I. Our findings document the utility and ecological validity of adult models from the same linguistic community to inform the interpretation of language assessment results for children who use more than one language on a daily basis. By comparing child responses to adult models in Study I, we were able to identify productions on the CELF Preschool-2 that were not reflective of developmental errors but rather were indicative of typical forms of cross-linguistic influence in the English context. Importantly, the vast majority of test items (Word Structure: 22/24, 91.7%; Expressive Vocabulary: 16/20, 80%) received adapted scoring, documenting that the consideration of cross-linguistic influence is as appropriate for preschool populations as it is for school-age children (Hendricks & Adlof, 2017). Without the application of the information from the adult models, these forms would have been scored as incorrect. Our findings are also consistent with research documenting the inherent cultural and linguistic bias associated with standardized assessments (Barragan et al., 2018; Pearce & Williams, 2013). Specifically, content bias is often involved in standardized assessment with the assumption that children have undergone similar life experiences that shape conceptual knowledge (De Lamo White & Jin, 2011). By way of our content analysis, we identified potential instances of content bias for JC–English speakers who typically come from Jamaica and not from other countries. For example, in assessing Expressive Vocabulary, 54.5% (n = 18) adults produced a JC linguistic item such as *spying glass*, reflecting JC-influenced lexical variation, for the target item "telescope." Children also produced similar JC linguistic items (n = 49, 27.8%; e.g., *camera*, *eye scope*) for this target (see Supplemental Material S2). These findings document that adults and children may have distinct experiences in the Jamaican context that do not align with concepts tested in standardized English assessments developed in the United States. Including adult models to inform child responses is not only a recommendation of the IEPMCS but is also in line with long-standing practice guidelines emphasizing the need for converging evidence and multiple methods of assessment due to the linguistic variability of DLL populations (ASHA, n.d.; McLeod et al., 2017). Though these recommendations exist, to our knowledge, this study is the first to empirically document this approach of using local adult models in an investigation of DLL preschoolers' language use.

Standardized Assessment Performance of JC–English DLLs

Study II sought to determine differences between children's performance using original and adapted scoring procedures for JC–English DLLs. Importantly, we found that being a simultaneous bilingual (i.e., a DLL proficient in both JC and English) did not prevent these children from being misdiagnosed. Our results indicated a statistically significant difference between JC–English DLLs' scores based on the standard English scoring procedure compared to the CELF Preschool-2 normative sample. Using the standard English scoring procedure, JC–English DLLs were classified as being language disordered at significantly higher rates compared to that expected for the standardized English sample (cf. Wiig et al., 2006; see Figures 1 and 2). This statistical significance of the performance rates (informed by mean score differences) suggested that using standard English scoring procedures places JC–English DLLs at risk of overdiagnosis of DLD. Our results align with previous studies that also showed overdiagnosis of disorder in linguistically diverse populations using standardized measures in school-age children (e.g., Barragan et al., 2018; Gross et al., 2014; Pearce & Williams, 2013). Our findings extend this concern to preschool-age children.

The information gathered above sets the stage for using adapted scoring procedures informed by adult models. Results revealed a statistically significant change in children's subtest scores using adapted scoring procedures. The mean scores for JC–English DLLs improved from 7.1 (SD = 2.8) to 11.7 (SD = 3.1) on the Word Structure subtest and from 8.3 (SD = 2.4) to 10.6 (SD = 2.4) on the Expressive Vocabulary subtest. JC–English DLLs' mean scores using adapted scoring were also comparable to the CELF Preschool-2 normative sample, with higher mean scores for Word Structure (see Table 3). Of importance to our study are (a) that the number of children classified as language disordered substantially decreased (i.e., by 37.5% for Word Structure and by 11.3% for Expressive Vocabulary) on both subtests when adapted scoring procedures were applied (see Figures 1 and 2) and (b) that children who were classified as language disordered using the adapted scoring approach exhibited similar linguistic profiles to those of monolingual speakers with DLD (Leonard, 2014), suggestive of greater sensitivity (i.e., a true positive: those children suspected of having DLD are actually diagnosed with DLD) in diagnostic classification critical to guiding clinical decision making (McLeod et al., 2013). In this study, adapted scoring that made use of adult models evidenced high classification accuracy (i.e., greater than 90%) for both Word Structure and Expressive Vocabulary subtests, with adequate sensitivity (Word Structure: .85; Expressive Vocabulary: .88) and high specificity (Word Structure: .94; Expressive Vocabulary: .92) levels (Plante & Vance, 1994). Important to note is that these sensitivity and specificity levels were higher in

comparison to previous works using adapted scoring not based on adult models from the same linguistic community (e.g., .64–.81 sensitivity and .64–.77 specificity; Hendricks & Adlof, 2017; Oetting et al., 2019). The apparent diagnostic accuracy associated with using adult models suggests a feasible approach to addressing concerns about the potential underidentification of disorder when using modified/adapted scoring procedures. As a consequence, consideration of the multiple sources of information (e.g., Restrepo's [1998] recommendation to use parent + teacher or SLP report) to contextualize assessment results remains a requisite component to inform decision making using adapted approaches. Furthermore, in the descriptive categories used to explain derived scores, the percentage of children identified in the high average range also increased for both subtests when the adapted scoring procedures were applied. These shifts in diagnostic classification describe an important clinical construct referred to as "clinically meaningful change" (Bain & Dollaghan, 1991). Although this construct is often used to describe treatment outcomes, it was considered in our results to characterize the substantial shifts in categorically qualifying preschoolers' language abilities.

Our analyses revealed a tendency for overidentification of language disorder when considering performance for Word Structure (i.e., morphosyntax; 37.5%) in comparison to Expressive Vocabulary (i.e., lexical items; 11.3%). We speculate that this may be due to lexical changes in language occurring more slowly than syntactic changes over time, resulting in more differences being present between JC and English morphosyntactic rules (Deuber, 2009; Trudgill, 2011; K. N. Washington et al., 2019). As such, morphosyntactic patterns that overlap with indicators of DLD could largely influence overdiagnosis rates for JC–English DLLs in the English context. Another influential factor could be the sensitivity of the language domains measured in documenting diagnostic indicators of DLD. Children with DLD often exhibit relative strengths in vocabulary skills (i.e., a compensatory strategy) in comparison to weaker grammatical skills (Ullman, 2016; Ullman et al., 2020; Yarian et al., 2021); thus, measures of morphosyntax may be more sensitive in capturing the indicators DLD.

All findings discussed thus far describe linguistic considerations to explain JC–English DLLs' language use in the English context. We conclude with findings that illustrate other potential influences on DLLs' performance on standardized assessments. Previous research has focused on other factors that influence bilingual children's performance on standardized measures, such as SES (cf. Barragan et al., 2018; Gross et al., 2014); however, recent work by Finneran et al. (2019) identified that misdiagnosis often occurs due to cultural and linguistic bias that is not explained by income alone. In our study, children were from a variety of socioeconomic backgrounds (as indexed by maternal education), and no distinct patterns of performance were identified unique to children of low SES; however, it is possible that higher SES and nonverbal IQ could have positively impacted preschoolers who scored within the average range using the standard English scoring procedure. We make this observation to offer the explanation that, using adapted scoring, our sample of JC–English DLLs also had a greater proportion classified as high average compared to projections, which could have been impacted by the larger representation of higher familial education levels in our sample compared to the CELF Preschool-2 normative sample. Though differences between original and adapted scores document cross-linguistic features as a primary influence on children's performance in this study, other factors (e.g., SES, school experience) should continue to be considered when examining DLLs' performance on language measures (Gross et al., 2014; Pearce & Williams, 2013).

Limitations and Future Directions

As with any research, this study was not without limitations. This study primarily included children who were simultaneous JC–English DLLs attending preschools in Kingston, Jamaica. This meant that Jamaican children in the United States were not included in our sample. Importantly, however, unlike other linguistically diverse populations, JC–English speakers in the United States (and in other migrant countries) come from Jamaica and not from other countries, offering some ecological validity to recruiting children from Jamaica to inform performance practices for working with this linguistic populace in the U.S. context. That said, future research examining patterns of Jamaican children of varying geographical backgrounds (in Jamaica and in migrant countries) and bilingual typologies (i.e., timing of dual language learning) may yield further insights into additional linguistic patterns for JC speakers. For example, examining patterns of performance in JC–English DLLs in the United States will be of importance as it is known that other contextual factors (e.g., schooling experience) shape dual language development. Other studies have documented differences in performance on standardized measures relating to children's bilingual typologies, an important consideration for future research to explore (Barragan et al., 2018; Gross et al., 2014). We also acknowledge that the appropriateness of standardized assessments was assessed using only two subtests designed for children in the United States, rather than an entire assessment battery. While a more comprehensive battery of assessment would have been more ideal, these data were collected as part of a larger study, the Jamaican Creole Language Project, that included a time-intensive protocol. Therefore, we strategically selected the two subtests involving the language domains of morphosyntax and vocabulary as these areas inform the profile of DLD for DLLs (Blom & Paradis, 2015), are the most commonly measured in clinical practice (Selin et al., 2019) and research (Wright Karem et al., 2019), and are highly targeted in treatment (Finestack & Satterlund, 2018). Further to this point was that we did not report on children's performance in both languages spoken but instead focused our investigation on assessment in English, where misdiagnosis is likely to occur and which represents the language spoken by most U.S. SLPs. It is also acknowledged that both best practice recommendations and the CELF

Preschool-2 manual emphasize the need for converging evidence using multiple measures in the diagnosis of language disorder. As such, future research should consider Jamaican children's performance in both spoken languages to inform linguistic patterns in each language (Gross et al., 2014; Wright Karem et al., 2019). This study also demonstrated how local sampling of adult responses to standardized stimuli in a specific geographical region (i.e., Kingston, Jamaica) can facilitate cross-linguistic interpretation. However, additional sampling of adult models in response to various language tasks is needed to provide increased understanding of the variability in JC–English language use and to expand upon assessment methods to distinguish between language difference and DLD. Exploration of local sampling with other language pairings and nonmainstream linguistic varieties may continue to document applicable comprehensive assessment approaches and considerations for linguistically diverse populations.

Clinical Implications and Conclusions

The findings from this study contribute knowledge about the important role of using adult models from the same linguistic community to inform adapted scoring procedures in contexts where assessment tools are not normed on the target population. These findings also increase the cautionary tale that JC–English DLLs are at risk of being overdiagnosed with DLD in standardized assessments using standard English scoring procedures. Our study demonstrates that, by using adapted scoring informed by adult models, diagnostic accuracy improves in classifying JC–English DLLs' language abilities. The application of mixed methods (i.e., qualitative and quantitative approaches) in this study also highlights the importance of multifaceted analysis to foster a deeper understanding of the linguistic profiles of DLL children. Use of adapted scoring provides a feasible and ecologically valid approach, supported by expert best practice recommendations (McLeod et al., 2017), to understand and assess the expressive language profiles of DLLs at risk for misdiagnosis. Specifically, our findings augment knowledge regarding JC linguistic patterns (see Supplemental Materials S1 and S2) that not only support SLPs' understanding of the clinical profile for this population to avoid misdiagnosis but also impact the entire service delivery continuum in considering appropriate treatment targets (Finestack & Satterlund, 2018; Selin et al., 2019). This study responds to the need for innovative research practices to understand the complexity of DLLs and inform clinical practice (Guiberson, 2020). More broadly, by investigating an understudied language pairing, we expand our theoretical understanding of bilingualism and model an approach applicable to an array of diverse linguistic communities that extend beyond the more traditionally studied bilingual paradigm (e.g., Spanish–English). Clinically, adult models could be included as part of a comprehensive assessment battery (e.g., dynamic assessment [Lewis et al., 2010] and contrastive phonological analysis [Telford Rose et al., 2020]), to address a long-standing and unmet need to reduce cultural and linguistic bias associated with assessment informed only by monolingual English benchmarks. In summary, our data show promising evidence for using adult models as an effective means to enhance cultural competence and responsivity among monolingual English-speaking SLPs serving clients who are culturally and linguistically diverse.

Author Contributions

Rachel Wright Karem: Conceptualization (Supporting), Data curation (Equal), Formal analysis (Lead), Investigation (Equal), Methodology (Equal), Resources (Supporting), Visualization (Equal), Writing – original draft (Lead), Writing – review & editing (Lead). **Karla N. Washington:** Conceptualization (Lead), Data curation (Equal), Formal analysis (Supporting), Funding acquisition (Lead), Investigation (Equal), Methodology (Equal), Project administration (Lead), Resources (Lead), Supervision (Lead), Visualization (Equal), Writing – original draft (Supporting), Writing – review & editing (Supporting).

Acknowledgments

The first author was a recipient of the U.S. Department of Education Office of Special Education Programs Preparation of Special Education and Early Intervention grant that funded her doctoral studies. The second author is a lead faculty member on this grant with salary support provided. The second author also receives salary support from the National Institute on Deafness and Other Communication Disorders (R21DC018170-01A1). The research presented was supported by an endowment to the Jamaican Creole Language Project. The authors would like to acknowledge the contributions of dissertation committee members (Nancy Creaghead, Kathryn Crowe, Amy Hobek, Lesley Raisor Becker, and Jennifer Vannest); Hubert Devonish and the Jamaica Language Unit; Laura and Richard Kretschmer; and research assistants Melanie Basinger, Corrine Deutenberg, Lauren Mikhail, and Cecilia Schwartz, who assisted in data collection and analysis and made this study possible. The authors would also like to extend their gratitude to the educators, families, adults, and children who graciously participated in this study.

References

Abu El Adas, S., Washington, K. N., Sosa, A., Harel, D., & McAllister, T. (2020). Variability across repeated productions in bilingual children speaking Jamaican Creole and English. *International Journal of Speech-Language Pathology, 22*(6), 648–659.

Altman, D. G. (1991). *Practical statistics for medical research*. Chapman & Hall.

American Speech-Language-Hearing Association. (n.d.). *Bilingual service delivery* [Practice portal]. https://www.asha.org/Practice-Portal/Professional-Issues/Bilingual-Service-Delivery/

Bain, B. A., & Dollaghan, C. A. (1991). The notion of clinically significant change. *Language, Speech, and Hearing Services in Schools, 22*(4), 264–270. https://doi.org/10.1044/0161-1461.2204.264

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. (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. https://doi.org/10.2167/beb392.0

Bishop, D. V., & Edmundson, A. (1987). Language-impaired 4-year-olds. *Journal of Speech and Hearing Disorders, 52*(2), 156–173. https://doi.org/10.1044/jshd.5202.156

Blom, E., & Paradis, J. (2015). Sources of individual differences in the acquisition of tense inflection by English second language learners with and without specific language impairment. *Applied Psycholinguistics, 36*(4), 953–976. https://doi.org/10.1017/S014271641300057X

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly, 3*(3), 229–242. https://doi.org/10.1207/s15434311laq0303_1

Castilla-Earls, A., 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,* 1147–1173. https://doi.org/10.1017/S0142716415000521

Castro, D. C., Garcia, E. E., & Markos, A. M. (2013). *Dual language learners: Research informing policy*. The University of North Carolina.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Craig, H. K., Thompson, C. A., Washington, J. A., & Potter, S. L. (2004). Performance of elementary-grade African American students on the Gray Oral Reading Tests. *Language, Speech, and Hearing Services in Schools, 35*(2), 141–154. https://doi.org/10.1044/0161-1461(2004/015)

De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language & Communication Disorders, 46*(6), 613–627. https://doi.org/10.1111/j.1460-6984.2011.00049.x

Deuber, D. (2009). 'The English we speaking': Morphological and syntactic variation in educated Jamaican speech. *Journal of Pidgin and Creole Languages, 24*(1), 1–52. https://doi.org/10.1075/jpcl.24.1.02deu

Devonish, H., & Harry, O. G. (2008). Jamaican Creole and Jamaican English: Phonology. In E. W. Schneider (Ed.), *Varieties of English, Vol 2: The Americas and the Caribbean* (pp. 256–289). Mouton de Gruyter.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2009). *Diagnostic Evaluation of Articulation and Phonology, U.S. Edition (DEAP)*. Pearson.

Dragoo, K. (2017). *The Individuals with Disabilities Education Act (IDEA), Part B: Key statutory and regulatory provisions*. https://www.everycrsreport.com/files/20190829_R41833_93da9c16f7893c0204c37090f2af7f6cc5e47fbe.pdf

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–Third Edition: Manual*. AGS.

Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). *Test de Vocabulario en Imágenes Peabody (TVIP)*. AGS.

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

Ehrler, D., & McGhee, R. (2008). *Primary Test of Nonverbal Intelligence*. Pro-Ed.

Finestack, L., & Satterlund, K. (2018). Current practice of child grammar intervention: A survey of speech-language pathologists. *American Journal of Speech-Language Pathology, 27*(4), 1329–1351. https://doi.org/10.1044/2018_AJSLP-17-0168

Finneran, D. A., Heilmann, J. J., Moyle, M. J., & Chen, S. (2019). An examination of cultural-linguistic influences on PPVT-4 performance in African American and Hispanic preschoolers from low-income communities. *Clinical Linguistics & Phonetics, 34*(3), 242–255. https://doi.org/10.1080/02699206.2019.1628811

Glascoe, F. P. (2000). *Parents' evaluation of developmental status: Authorized Australian version*. Centre for Community Child Health.

Gorman, B. K., Bingham, G. E., Fiestas, C. E., & Terry, N. P. (2016). Assessing the narrative abilities of Spanish-speaking preschool children: A Spanish adaptation of the Narrative Assessment Protocol. *Early Childhood Research Quarterly, 36,* 307–317. https://doi.org/10.1016/j.ecresq.2015.12.025

Greenwood, C. R., Carta, J. J., Kelley, E. S., Guerrero, G., Young Kong, N., Atwater, J., & Goldstein, H. (2016). Systematic replication of the effects of a supplementary, technology-assisted, storybook intervention for preschool children with weak vocabulary and comprehension skills. *The Elementary School Journal, 116*(4), 574–599. https://doi.org/10.1086/686223

Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology, 23*(4), 574–586. https://doi.org/10.1044/2014_AJSLP-13-0026

Guiberson, M. (2020). Introduction to the forum: Innovations in clinical practice for dual language learners, Part 2. *American Journal of Speech-Language Pathology, 29*(3), 1113–1115. https://doi.org.10/1044/2020_AJSLP-20-00153

Guiberson, M., & Ferris, K. P. (2019). Early language interventions for young dual language learners: A scoping review. *American Journal of Speech-Language Pathology, 28*(3), 945–963. https://doi.org/10.1044/2019_AJSLP-IDLL-18-0251

Hemsley, G., Holm, A., & Dodd, B. (2010). Patterns in diversity: Lexical learning in Samoan–English bilingual children. *International Journal of Speech-Language Pathology, 12*(4), 362–374. https://doi.org/10.3109/17549501003721064

Hendricks, A. E., & Adlof, S. M. (2017). Language assessment with children who speak nonmainstream dialects: Examining the effects of scoring modifications in norm-referenced assessment. *Language, Speech, and Hearing Services in Schools, 48*(3), 168–182. https://doi.org/10.1044/2017_LSHSS-16-0060

Individuals with Disabilities Education Act, 20 U.S.C § 1412. (2006).

International Expert Panel on Multilingual Children's Speech. (2012). *Multilingual children with speech sounder disorders: Position paper*. Research Institute for Professional Practice, Learning and Education, Charles Sturt University.

Justice, L. M., Bowles, R., Pence, K., & Gosse, C. (2010). A scalable tool for assessing children's language abilities within a narrative context: The NAP (Narrative Assessment Protocol). *Early Childhood Research Quarterly, 25*(2), 218–234. https://doi.org/10.1016/j.ecresq.2009.11.002

Khansir, A. A., & Pakdel, F. (2019). Contrastive analysis hypothesis and second language learning. *Journal of ELT Research, 4*(1), 35–43. https://doi.org/10.22236/JER_Vol4Issue1pp35-43

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56,* 746–759. https://doi.org/10.1177/0013164496056005002

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders, 43*(6), 456–473. https://doi.org/10.1016/j.jcomdis.2010.02.002

Lee, R., & Oetting, J. B. (2014). Zero marking of past tense in child African American English. *SIG 1 Perspectives on Language Learning and Education, 21*(4), 173–181. https://doi.org/10.1044/lle21.4.173

León, M., Washington, K. N., Fritz, K., León, M. A., Basinger, M., & Crowe, K. (2021). Intelligibility in Context Scale: Sensitivity and specificity in the Jamaican Context. *Clinical Linguistics & Phonetics, 35*(2), 171–154. https://doi.org/10.1080/02699206.2020.1766574

Leonard, L. B. (2014). Specific language impairment across languages. *Child Development Perspectives, 8*(1), 1–5. https://doi.org/10.1111/cdep.12053

Lewis, N., Castilleja, N., Moore, B. J., & Rodriguez, B. (2010). Assessment 360°: A panoramic framework for assessing English language learners. *SIG 14 Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse (CLD) Populations, 17*(2), 37–56. https://doi.org/10.1044/cds17.2.37

Lockart, R., & McLeod, S. (2013). Factors that enhance English-speaking speech-language pathologists' transcription of Cantonese-speaking children's consonants. *American Journal of Speech-Language Pathology, 22*(3), 523–539. https://doi.org/10.1044/1058-0360(2012/12-0009)

Masso, S., McLeod, S., Cronin, A., & Phạm, B. (2019). Transcription of Vietnamese adults' and children's consonants by English-speaking speech-language pathologists. *Folia Phoniatrica et Logopaedica, 72*(2), 92–107. https://doi.org/10.1159/000500245

McLeod, S., Harrison, L. J., McAllister, L., & McCormack, J. (2013). Speech sound disorders in a community study of preschool children. *American Journal of Speech-Language Pathology, 22*(3), 503–522. https://doi.org/10.1044/1058-0360(2012/11-0123)

McLeod, S., Harrison, L. J., & McCormack, J. (2012a). *The Intelligibility in Context Scale*. Charles Sturt University.

McLeod, S., Harrison, L. J., & McCormack, J. (2012b). *Mezha fi Omoch ada Piipl kyan Andastan di Pikni: Jamiekan* [Intelligibility in Context Scale: Jamaican Creole] (K. N. Washington & H. Devonish, Trans.). Charles Sturt University.

McLeod, S., Verdon, S., & International Expert Panel on Multilingual Children's Speech. (2017). Tutorial: Speech assessment for multilingual children who do not speak the same language(s) as the speech-language pathologist. *American Journal of Speech-Language Pathology, 26*(3), 691–708. https://doi.org/10.1044/2017_AJSLP-15-0161

Morgan, G. P., Restrepo, M. A., & Auza, A. (2013). Comparison of Spanish morphology in monolingual and Spanish–English bilingual children with and without language impairment. *Bilingualism: Language and Cognition, 16*(3), 578–596. https://doi.org/10.1017/S1366728912000697

Morgan, P. L., Scheffner Hammer, C., Farkas, G., Hillemeier, M. M., Maczuga, S., Cook, M., & Morrano, S. (2016). Who receives speech/language services by 5 years of age in the United States? *American Journal of Speech-Language Pathology, 25*(2), 183–199. https://doi.org/10.1044/2015_AJSLP-14-0201

Morren, R. C., & Morren, D. M. (2007). Are the goals and objectives of Jamaica's Bilingual Education Project being met? [Working paper]. *SIL International,* 1–10. https://www.sil.org/system/files/reapdata/79/57/09/79570925941905031645561824600563786634/silewp2007_009.pdf

Nayeb, L., Lagerberg, D., Sarkadi, A., Salameh, E. K., & Eriksson, M. (2021). Identifying language disorder in bilingual children aged 2.5 years requires screening in both languages. *Acta Paediatrica, 110*(1), 265–272. https://doi.org/10.1111/apa.15343

Nero, S., & Stevens, L. (2018). Analyzing students' writing in a Jamaican Creole–speaking context: An ecological and systemic functional approach. *Linguistics and Education, 43,* 13–24. https://doi.org/10.1016/j.linged.2017.12.002

Neundorf, K. (2002). *The content analysis guidebook*. Sage.

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. https://doi.org/10.1044/2018_LSHSS-CLSLD-17-0156

Oetting, J. B., Berry, J. R., Gregory, K. D., Rivière, A. M., & McDonald, J. (2019). Specific language impairment in African American English and Southern White English: Measures of tense and agreement with dialect-informed probes and strategic scoring. *Journal of Speech, Language, and Hearing Research, 62*(9), 3443–3461. https://doi.org/10.1044/2019_JSLHR-L-19-0089

Paradis, J. (2016). The development of English as a second language with and without specific language impairment: Clinical implications. *Journal of Speech, Language, and Hearing Research, 59*(1), 171–182. https://doi.org/10.1044/2015_JSLHR-L-15-0008

Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning* (2nd ed.). Brookes.

Pearce, W., & Williams, C. (2013). The cultural appropriateness and diagnostic usefulness of standardized language assessments for Indigenous Australian children. *International Journal of Speech-Language Pathology, 15*(4), 429–440. https://doi.org/10.3109/17549507.2012.762043

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

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 41*(6), 1398–1411. https://doi.org/10.1044/jslhr.4106.1398

Selin, C. M., Rice, M. L., Girolamo, T., & Wang, C. J. (2019). Speech-language pathologists' clinical decision making for children with specific language impairment. *Language, Speech, and Hearing Services in Schools, 50*(2), 283–307. https://doi.org/10.1044/2018_LSHSS-18-0017

Semel, E., Wiig, E., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4)*. Pearson.

Semel, E., Wiig, E., & Secord, W. A. (2006a). *Clinical Evaluation of Language Fundamentals–Fourth Edition, Australian Standardised Edition (CELF-4 Australian)*. PsychCorp.

Semel, E., Wiig, E., & Secord, W. A. (2006b). *Clinical Evaluation of Language Fundamentals–Fourth Edition, Spanish (CELF-4 Spanish)*. Pearson.

Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology, 16*(3), 246–259. https://doi.org/10.1044/1058-0360(2007/029)

Telford Rose, S. L., Payne, K. T., De Lisser, T. N., Harris, O. L., & Elie, M. (2020). A comparative phonological analysis of Guyanese Creole and Standard American English: A guide for speech-language pathologists. *Perspectives of the ASHA Special Interest Groups, 5*(6), 1813–1819. https://doi.org/10.1044/2020_PERSP-20-00173

Terrell, S. L., Arensberg, K., & Rosa, M. (1992). Parent–Child Comparative Analysis. *Language, Speech, and Hearing Services in Schools, 23*(1), 34–42. https://doi.org/10.1044/0161-1461.2301.34

Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford University Press.

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S. A. Small (Eds.), *Neurobiology of language* (pp. 953–968). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00076-6

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The neurocognition of developmental disorders of language. *Annual Review of Psychology, 71,* 389–417. https://doi.org/10.1146/annurev-psych-122216-011555

U.S. Census Bureau. (2018). *2017 American Community Survey single-year estimates*https://www.census.gov/newsroom/press-kits/2018/acs-1year.html

Viera, A. J., & Garret, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*(5), 360–363. https://fammedarchives.blob.core.windows.net/imagesandpdfs/pdfs/FamilyMedicineVol37Issue5Viera360.pdf

Washington, J. A. (1996). Issues in assessing the language abilities of African American children. In A. Kahmi, K. Pollack & J. Harris (Eds.), *Communication development and disorders in African American children: Research, assessment, and intervention* (pp. 35–54). Brookes.

Washington, J. A., & Craig, H. K. (1992). Performances of low-income, African American preschool and kindergarten children on the Peabody Picture Vocabulary Test–Revised. *Language, Speech, and Hearing Services in Schools, 23*(4), 329–333. https://doi.org/10.1044/0161-1461.2304.329

Washington, K. N. (2012). Translation to practice: Typical bidialectal speech acquisition in Jamaica. In S. McLeod & B. A. Goldstein (Eds.), *Multilingual aspects of speech sound disorders in children* (pp. 101–105). Multilingual Matters. https://doi.org/10.21832/9781847695147-015

Washington, K. N., Fritz, K., Crowe, K., Kelly, B., & Wright Karem, R. (2019). Bilingual preschoolers' spontaneous productions: Considering Jamaican Creole and English. *Language, Speech, and Hearing Services in Schools, 50*(2), 179–195. https://doi.org/10.1044/2018_LSHSS-18-0072

Washington, K. N., McDonald, M., McLeod, S., Crowe, K., & Devonish, D. (2017). Validation of the Intelligibility in Context Scale for Jamaican Creole–speaking preschoolers. *American Journal of Speech-Language Pathology, 26*(3), 750–761. https://doi.org/10.1044/2016_AJSLP-15-0103

Weber, R. (1990). *Basic content analysis* (2nd ed.). Sage. https://doi.org/10.4135/9781412983488

Wiig, E. H., Secord, W. A., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals Preschool*. Harcourt Assessment.

Wiig, E. H., Secord, W. A., & Semel, E. (2006). *Clinical Evaluation of Language Fundamentals Preschool–Second Edition*. Harcourt Assessment.

Wright Karem, R., Washington, K. N., Crowe, K., Jenkins, A., Leon, M., Kokotek, L., Raisor-Becker, L., & Westby, C. (2019). Current methods of evaluating the language abilities of multilingual preschoolers: A scoping review using the International Classification of Functioning, Disability and Health–Children and Youth Version. *Language, Speech, and Hearing Services in Schools, 50*(3), 434–451. https://doi.org/10.1044/2019_LSHSS-18-0128

Yarian, M., Washington, K. N., Spencer, C. E., Vannest, J., & Crowe, K. (2021). Exploring predictors of expressive grammar across different assessment tasks in preschoolers with or without DLD. *Communication Disorders Quarterly, 42*(2), 111–121. https://doi.org/10.1177/1525740119868238