ENGLISH TENSE/AGREEMENT MEASURES AS POTENTIAL DIAGNOSTIC MARKERS IN SPANISH-ENGLISH BILINGUALS WITH DEVELOPMENTAL LANGUAGE DISORDER

By

Juliana Ronderos

A Thesis submitted to the Department of Psychology,

University of Houston

In partial fulfillment of the requirements for the degree of

MASTER OF ARTS in Psychology

Chair of Committee: Arturo E. Hernandez. Ph.D.

Committee Member: Anny P. Castilla-Earls, Ph.D.

Committee Member: Monique T. Mills, Ph.D.

University of Houston

August 2020

Acknowledgments

I would like to thank my advisors, Dr. Anny Castilla-Earls and Dr. Arturo Hernandez, for providing guidance and insight on this thesis and my research on language development in bilinguals. I would also like to thank the research assistants in the Dual Lab in the Department of Communication Sciences and Disorders for collecting the data, and transcription and coding of language samples used for this research. Finally, I would like to thank my family for their immense patience, encouragement, and support throughout this project.

This research was supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) Grant R15DC013670-01A1 Evidence-Based Clinical Diagnosis of Specific Language Impairments in Spanish-Speaking Children and by the National Institutes of Health (NIH) Grant 1K23DC015835-01 Differentiating First Language Loss from Language Impairment in Bilingual Children awarded to Dr. Anny Castilla-Earls.

ABSTRACT

Considerable individual variability is characteristic of bilingual language development, including the development and mastery of morphosyntactic skills (Paradis, 2005; Paradis et al., 2008). To address this variability, best practices for assessment of language disorders in bilingual children recommend evaluating both languages to take into account any cultural and linguistic biases (Bedore & Peña, 2008; Kohnert, 2010). However, bilingual assessment is not always feasible. In the current study, we explored the potential clinical use of an English-only assessment approach using tense/agreement composite measures.

Participants included 93 Spanish-English typically developing bilingual children and 62 peers with language disorders. Measures of tense/agreement diversity, productivity, and accuracy (Hadley & Short, 2005; Bedore & Leonard, 1998) were calculated from language samples in English. We created a new measure, morphosyntactic development levels, inspired on the tense/agreement composites while also considering a child's relative language dominance and length of exposure to English. The morphosyntactic development levels reflected diversity and productivity of tense/agreement morphemes and then evaluated accuracy after sufficient levels of diversity and productivity were reached. All measures were evaluated with regards to their ability to predict group difference and their discriminant accuracy for clinical utility.

All measures demonstrated the ability to predict group differences. The morphosyntactic development levels provided the most informative results and classification accuracy values. These results suggest that English-only language measures may have informative

value in assessment of bilingual children's language when used in combination with parental report, language dominance, and exposure information.

TABLE OF CONTENTS

ACKNOWLEDGMENTS II
ABSTRACT III
LIST OF TABLES VI
LIST OF FIGURES VII
INTRODUCTION 1
Identification of DLD in Bilingual Children 3
Development of Tense/Agreement Morphemes in English 7
Tense/agreement in English in Studies with Bilingual Children 10
Context for the present study 11
METHOD 13
Participants 13
Measures 15
Procedures 20
Reliability 21
Data Analytic Strategy 22
RESULTS 24
Group Comparisons 24
Diagnostic Classification of Tense Agreement Measures 26
Diagnostic Classification of Morphosyntactic Development Levels 27
DISCUSSION 28
Tense/Agreement Morpheme Composite Measures 29
Morphosyntactic Development Levels 31
Future Research and limitations 32
CONCLUSION 34
REFERENCES 35

LIST OF TABLES

Table 1. Child and family demographics (*N*=155). 51
Table 2. Descriptive statistics and results of Independent sample T-tests (*N*=155)..... 52
Table 3. Sensitivity, specificity, and likelihood ratios by measure. 53

LIST OF FIGURES

Figure 1. Language Dominance Distribution 48

Figure 2. Mean productions by tense/agreement morpheme category and diagnostic classification 49

Figure 3. Summary of Morphosyntactic Development Levels..... 50

Figure 4. Positive and Negative Likelihood Ratios and 95% Confidence Intervals 54

Introduction

Bilingual children in the United States are a heterogeneous group with different levels of language dominance and language experiences, making assessment a difficult task. Best practices for assessment of language disorders in bilingual children recommend evaluating adequate proficiency in both languages as an approach that takes into account cultural and linguistic biases (Bedore & Peña, 2008; Kohnert, 2010). However, this is not always feasible. In the United States, English is still most frequently used in the assessment of bilinguals, as only 6% of speech-language pathologists (SLPs) and audiologists in the United States can provide assessment in languages in addition to English (American Speech-Language Hearing Association (ASHA), 2018). Meaningful English-only assessment tools would be valuable for monolingual SLPs with bilingual children in their caseload. In this study, we explored the potential clinical use of tense/agreement composite measures from English-only language samples of preschool and early elementary school Spanish-English bilinguals to be able to differentiate between children with typical language skills and those at risk of language disorders.

Developmental language disorder (DLD) is used to describe children with a significant deficit in language learning abilities not associated with hearing loss, intellectual disability, or another differentiating biomedical condition (Bishop et al., 2017; Bishop et al., 2016). The term DLD is the current consensus term for child language disorders (Bishop et al., 2017; Bishop et al., 2016). DLD encompasses the more narrow definition of specific language impairment (SLI), sometimes also referred to as primary language impairment (PLI), which refers to children whose difficulties are specific only to language and who otherwise present within average non-verbal intelligence and no other

neurological conditions (Leonard, 2014). The broader term DLD also includes children with language disorders with comorbid conditions and below-average non-verbal IQ scores (although they do not meet the criteria for intellectual disability). For this study, we will use the consensus term DLD while recognizing that many of our participants and the studies we cite meet the stricter criteria for SLI. The prevalence of DLD in preschool-aged children in the United States is approximately 7.4-7.5% (Tomblin et al., 1997; Norbury et al., 2016). Children with a history of DLD are less likely to be married, have close personal relationships, complete high school, obtain a university degree, and/or hold high-paying jobs compared to their typically developing peers (Johnson et al., 2010). Therefore, it is of critical importance that children with DLD are identified early to be eligible to receive intervention to improve their long-term social, academic, and professional outcomes.

During the preschool and early elementary years, language difficulties in children with DLD are highly apparent in the acquisition and mastery of grammar and morphosyntax (Leonard, 1998, 2014). The grammatical forms that are problematic for children with DLD, however, may vary by language. English monolingual children with DLD have difficulties with the use of tense/agreement morphemes (Rice & Wexler, 1996; Rice et al., 1998) and these are often used as clinical markers of DLD in English monolingual children. Spanish, however, is more inflected than English and requires agreement in number and gender for both noun and verb phrases providing additional areas of potential difficulty for children with DLD. In Spanish-speaking children, multiple areas of difficulty have been targeted in research, including articles and direct object

pronouns (Castilla-Earls et al., 2016), and tense markers (Grinstead et al., 2013) which are frequently used as markers of DLD in Spanish monolingual children.

Identification of DLD in Bilingual Children

The population of bilinguals in the United States consists of many cultural and linguistic backgrounds and a full spectrum of language abilities and exposure for both of their languages. Some bilinguals are simultaneous bilinguals (individuals who are exposed to both languages from birth until age 3) and some are dual language learners or sequential bilinguals (those primarily exposed to one language at home and, after significant progress towards acquiring this language, after age 3, usually start learning English in school). In the United States, the number of school-age children (ages 5-17) classified as dual language learners is steadily increasing and approximately 70% of them are Hispanics who speak Spanish at home in addition to at least some English at school (U.S. Census Bureau, 2015). Since the trajectory and the rate in which each language develops depends largely on the context, the degree, and the quality of exposure to each of the languages, this will affect language learning in terms of opportunity for frequent and rich input, availability of formal schooling, and attitudes and motivation toward speaking the languages (Paradis et al., 2011).

Several studies have compared the morphosyntactic abilities of bilingual children with and without DLD to those of monolingual children with and without DLD to further assess what characteristics of morphosyntactic development can differentiate between a language disorder and second language acquisition. Paradis and Crago (2000) compared morphosyntactic skills (tense/agreement and temporal context) in French between a group of English-speaking children learning French and French-monolingual children with and

without DLD. Their results show that French-monolingual children with DLD and children learning French demonstrated similarities in their use of morphosyntax, specifically as predicted by the (extended) optional infinitive framework (EOI; Rice & Wexler, 1996; Rice et al., 1998) where children omit obligatory tense/agreement morphemes. They found that the tense/agreement difficulties in children with DLD were also found in typical development at later ages in the acquisition of the second language. In a subsequent study, Paradis (2005) examined the language abilities of typically developing children from multiple linguistic backgrounds learning English as a second language. During the first 18 months of consistent exposure to English, the researchers assessed both spontaneous speech during play and elicited speech using the grammatical probes of Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001). The results confirmed the predictions that the accuracy rates and error patterns in this group were similar to same-age monolingual children with DLD. Further, the grammatical probes from the TEGI, which have been normed with monolingual children, consistently placed these English language learners within the criterion score range indicating DLD. These low scores for English language learners on norm-referenced assessments, the researchers posited, could be the result of the cultural and linguistic differences or a shift in dominance rather than that of an underlying DLD condition.

In a separate study to analyze the characteristic errors of children with DLD, Paradis et al. (2003) studied a group of French-English simultaneous bilingual children with DLD and compared them to age-matched French- and English-monolingual children with DLD. The children were compared in terms of their morphosyntactic abilities in language production, particularly concerning the EOI framework predictions. The results

of this study suggest that the patterns predicted under the EOI framework were realized for both bilingual and monolingual groups of children with DLD. All groups were less accurate for tense/agreement morphemes than for other morphemes not related to tense/agreement. Furthermore, the mean accuracy for bilingual and monolingual groups was similar, indicating that the bilingual children did not exhibit deeper deficits in these tense/agreement morphemes that their monolingual peers.

These studies raise important clinical implications for assessment of bilingual children since the errors produced by typically developing bilinguals and monolingual children with DLD generally result in lower than average scores on English standardized tests. The similarity in accuracy rates and error types which include clinical markers of DLD indicates that bilinguals are at risk for both over- and under-diagnosis for DLD (Genesee et al., 2004; Sullivan & Bal, 2013). Best practices in identification of DLD in bilingual children recommend assessment in both child's language (Bedore & Peña, 2008; Kohnert, 2010) as there may be cross-linguistic influences as the languages develop, which result in target-deviant structures that seem to be influenced by the child's other language (Genesee et al., 2004).

This heterogeneous nature of language abilities of bilingual children makes assessment, diagnosis, and intervention in bilingual children with DLD a complex process. Language sample analysis is frequently endorsed as an approach that is resistant to cultural and linguistic biases (Gutierrez-Clellen et al., 2000; Heilmann, 2010; Heilmann et al., 2010). In assessment of bilingual children, language sample analysis measures in both languages have been used as an unbiased indicator in language assessments and continue to be the gold standard for identification of DLD in Spanish-English bilingual

children, especially when accompanied by parental or teacher language concern (Bedore & Peña, 2008; Restrepo, 1998). Once language samples are transcribed and coded, they can be analyzed for several measures of language development that can be judged against the dialectical background of the child. The measures derived from language samples which have been sensitive to identify DLD in both monolingual and bilingual children include the mean length of utterance in words (MLU_w) as a measure of morphosyntactic complexity (Gutierrez-Clellen et al., 2000) and the number of different words (NDW) as a measure of lexical diversity (Golberg et al., 2008). These measures are considered culturally sensitive and are recommended for language sample analysis with Spanish-English bilingual children (Rojas & Iglesias, 2009). These measures, however, do not fully capture the development of obligatory tense/agreement morphemes which has been established as a clinical marker for language disorders (Bedore & Leonard, 1998; Gutierrez-Clellen et al., 2008; Leonard, 2014; Rice & Wexler, 1996).

More recently, there has been increased research supporting the development of standardized assessments normed on a bilingual population to adequately identify DLD in Spanish-English bilinguals. The Bilingual English-Spanish Assessment (BESA; Peña et al., 2018) has been developed to provide a comprehensive picture of an early bilingual's language development, including morphosyntax, semantics, and phonology. Once administered in both languages, the BESA has been designed to assess the language of Spanish-English bilinguals in both the child's languages and to use the child's "best language" to determine if DLD is present. The morphosyntax subtests of the BESA, in particular, utilize cloze and sentence repetition tasks to target a variety of grammatical morphemes and sentence structures and have been shown to discriminate children with

DLD in English and Spanish. Classification accuracy for the BESA morphosyntax subtests, per the BESA manual (Peña et al., 2014, 2018), is as follows: sensitivity ranged 87-89% for English and 87-94% for Spanish; specificity ranged 81-88% for English and 81-88% for Spanish. To provide validation for the measures in the Spanish morphosyntax subtest of the BESA, Gutiérrez-Clellen et al. (2006) evaluated the discriminant accuracy of certain Spanish grammatical measures in Spanish-speaking children between the ages of 4-7 to include examples of elicited articles, clitics, subjunctive verbs, and complex syntax which have good discrimination accuracy for bilingual children in Spanish. These findings also led to age-based cutoff scores which are unique in the BESA to account for potential developmental changes in the manifestation of DLD (Gutiérrez-Clellen et al., 2006).

Development of Tense/Agreement Morphemes in English

Tense/agreement marking is an area of known difficulty for English-monolingual children with DLD (Leonard, 2014). Tense/agreement morphemes consist of five target morpheme categories, a combination of bound morphemes (the inflections for regular third person singular *-s* and regular past tense *-ed*) and free morphemes (the function words copula *to be*: *am, are, is, was* and *were*; the auxiliary forms of the verb *to be*: *am, are, is, was,* and *were*; and the auxiliary forms of the verb *to do*: *do, does,* and *did*). All of these have a collection of properties which require person or number agreement, or both. Tense/agreement composite measures have been used by researchers and clinicians to compare the language abilities of children because the production of these morphemes is vulnerable in children with DLD (Bedore & Leonard, 1998; Gutierrez-Clellen et al., 2008; Rice & Wexler, 1996; Rice et al., 1998).

One such composite is the finite verb morphology composite (FVMC; Bedore & Leonard, 1998) which is used to measure accuracy based on the percent of correct productions in obligatory contexts of selected morphemes (the subset of tense/agreement morphemes included in this composite has been slightly different depending on the researcher). Rice et al. (1998) examined morphosyntactic growth using an accuracy composite and demonstrated that children with DLD developed morphosyntactic tense/agreement morphemes in similar trajectories to younger children with typically developing language. The Rice/Wexler Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001) uses an accuracy composite as a screening tool to assess the risk of DLD. Accuracy composites of tense/agreement morphemes have been shown to have high levels of diagnostic accuracy (good sensitivity and specificity) in identifying DLD in school-age monolingual English-speaking children which makes it an important tool for clinical decision making (Bedore & Leonard, 1998; Gladfelter & Leonard, 2013; Rice & Wexler, 1996; Rice et al., 1998).

Recent studies on tense/agreement diversity and productivity provide support for the gradual morphosyntactic learning (GML) account proposed by Rispoli and Hadley (2011) for the development of a morphosyntactically defined system. The GML is based on a sentence production framework which takes a *developmental* perspective to account for tense/agreement diversity and growth. Under the GML framework, children produce novel sentences as part of their developmental progression to fully utilize language. Initially, children evolve expressively through direct activation and continue evolving towards grammatical encoding for more exact communication of person and time (Rispoli & Hadley, 2011). Recent studies by Hadley, Rispoli, and colleagues (Hadley & Holt,

2006; Hadley & Short, 2005; Rispoli et al., 2009; Rispoli et al., 2012) have used composite measures of tense/agreement measuring diversity and productivity as developmental precursors to accuracy composites. They argue that until children's language is sufficiently diverse and productive, accuracy composites may suffer from measurement error by overestimating the child's grammatical knowledge since it may be based only on limited diversity of sentences and obligatory contexts.

Two tense/agreement composite measures were developed by Hadley and colleagues (Hadley & Holt, 2006; Hadley & Short, 2005) as more developmentally appropriate for use in younger children because they target diversity and productivity. The tense marker total (TMT) serves as a measure of diversity as it awards children one point for the production of each surface form of the five target morpheme categories. The tense/agreement productivity (TAP) measures productivity within each of the five target morpheme categories by awarding one point (and a maximum of 5 per category) for each sufficiently different use. Both of these measures do not include productions reflecting direct activation which might be learned on a rote basis but focus on productions reflecting grammatical encoding. In their study of English monolinguals, Hadley and Short (2005) found evidence that scores on these composite measures of diversity and productivity of tense/agreement morphemes were able to reliably predict which of the children as young as 24-29 months were deemed at risk for DLD by the time they were 3-years old. In a series of subsequent studies, these measures provided evidence for gradual longitudinal tense/agreement morpheme growth in young children (Hadley & Holt, 2006; Rispoli et al., 2009; Rispoli et al., 2012). More recently, other researchers have explored the clinical use of the TMT, TAP, and FMVC in groups of older children with and without

DLD for diagnostic accuracy (Gladfelter & Leonard, 2013) and to track the growth trajectories of tense/agreement morphemes (Leonard et al., 2017).

Tense/agreement in English in Studies with Bilingual Children

It is imperative to understand the trajectory of tense/agreement morpheme marking in bilingual children to avoid over-diagnosing bilingual children as having DLD (Paradis, 2005). To validate the English morphosyntax subtest of the BESA, Gutierrez-Clellen and Simon-Cereijido (2007) evaluated the discriminant accuracy of English grammatical measures, including tense/agreement morphemes, in Latino English-speaking children between the ages of 4-7. The results supported previous research that grammatical measures in English focused primarily on verb morphology in addition to other grammatical forms that have good clinical validity. These grammatical measures tested in this study were able to identify children with fair sensitivity and good specificity with English-speaking children despite exposure or use of Spanish (Gutierrez-Clellen & Simon-Cereijido, 2007).

More recently, a few studies have investigated the clinical utility of tense/agreement measures of diversity, productivity, and accuracy using language samples of Spanish-English bilinguals. Gusewski and Rojas (2017) examined the trajectory of English tense-marking in preschool and early elementary school-aged typically developing Spanish-English bilingual children at four points over two years. They calculated the finite verb morphology composite (FVMC; Bedore & Leonard, 1998) and an adapted version of the FVMC which added auxiliary *do* forms and irregular past-tense forms to the original FVMC verb morpheme categories to measure growth using accuracy. Their results showed significant growth for both composites suggesting that after

sufficient time exposure or immersion in English as the L2 allows English L2 learners to achieve high tense-marking accuracy over time. Potapova et al. (2018) compared a group of typically developing Spanish-English bilinguals and a group of peers with low language skills using the TMT, TAP, and a tense/agreement accuracy composite measure. Across both groups, TMT and TAP were correlated with traditional language sample measures (MLU_w and NDW). Furthermore, they found that these measures predicted group differences between children with typical language and low language and also exhibited growth over time. The researchers concluded that these measures provide a practical means of assessing language skills in developing bilingual children with potential for clinical utility.

Context for the present study

The majority of research to date on bilingual children's use of English tense/agreement morphemes has been conducted with small samples sizes (for exceptions, see Gutiérrez-Clellen et al., 2006; Paradis et al., 2013) or has been in children with a variety of language backgrounds rather than specifically with Spanish-English bilinguals (for exceptions, see Gutiérrez-Clellen et al., 2006; Gutierrez-Clellen et al., 2008). Further, studies with Spanish-English bilingual children have not taken into account the continuum of language experiences that is characteristic of this group of bilinguals in the United States. In the present study, we aim to take into account bilingual children's different language dominance profiles and exposure to English in a larger sample of children. For the 155 children in our study, language samples were elicited through story retell and story generation tasks and were coded for tense/agreement composite measures of diversity, productivity, and accuracy. We calculated the tense marker total and the

tense/agreement productivity (TMT and TAP, respectively; Hadley & Holt, 2006; Hadley & Short, 2005), and a modified version of the finite verb morphology composite (FVMC; Bedore & Leonard, 1998) to reflect the productive accuracy of tense marking in obligatory contexts for the tense/agreement morphemes forms of interest. We investigated if English tense/agreement measures of diversity, productivity, and accuracy can discriminate between children with typical language development and children with DLD in Spanish-English bilinguals. Then, we took a novel approach in that we considered demographic characteristics, including age and language experience (dominance and exposure to English), to evaluate children's achievement levels of morphosyntactic development. We have created these morphosyntactic development levels to take into account tense/agreement accuracy only once a sufficiently diverse and productive tense/agreement system has been established.

The specific research questions for this study were:

1. Do measures of tense marking diversity, productivity, and accuracy in English discriminate children with DLD from children with typical language development in early Spanish-English bilingual children?
2. Can we combine these measures consider tense/agreement diversity, productivity, and accuracy to model levels of English morphosyntactic development that can be used along with demographic characteristics (such as age, language dominance, and length of English exposure) to discriminate children with DLD from children with typical language development in Spanish-English bilinguals?

Method

Participants

The participants for this study were 155 Spanish-English bilingual children (86 boys, 67 girls) between the ages of 4;0 and 7;5 (M = 65.90, SD = 11.05 in months) recruited from schools and speech-language clinics in Buffalo, NY and Houston, TX. At the time of recruitment, forty-nine of these children were receiving speech-language services either in their school or in a speech-language clinic. The children came from different socio-economic and linguistic backgrounds and displayed different language dominance profiles and lengths of exposure to English, reflecting a wide spectrum of Spanish-English bilinguals in the United States.

Children included in the study met the following inclusionary criteria: (a) Spanish and English were the only languages spoken in the home or at school. (b) Children were between 4;0 and 7;11 years old. (c) All children had average non-verbal cognitive abilities as evidenced by a standard score equal or greater than 75 on the Matrices subtest of the Kaufman Brief Intelligence Test, Second Edition (KBIT-II; Kaufman & Kaufman, 2004). (d) Finally, all children had to pass a hearing screening test. All children were administered an otoacoustic emission screening at 1000 to 4000 Hz and any participants who failed the screening received a pure tone screening test at 25dB HL at 1000, 2000, and 4000 Hz on an alternate day.

Diagnosis classification.

The Bilingual English-Spanish Assessment (BESA; Peña et al., 2018) is a normed-referenced test developed for the identification of Spanish-English bilingual children with DLD which has been normed with Spanish-English bilinguals ages 4;0 to 6;11 in the

United States. The Bilingual English-Spanish Assessment – Middle Extension (BESA-ME; Peña et al., 2008) is an experimental measure available for researchers, to assess the language abilities of children in middle elementary, ages 7;0 to 9;11. The BESA and BESA-ME consist of several subtests testing different aspects of language, including semantics, and morphosyntax in English and Spanish. The morphosyntax subtests of the BESA and BESA-ME have been shown to have good discrimination accuracy to identify DLD in Spanish-English bilingual children. Classification accuracy for the BESA morphosyntax subtests per the BESA manual (Peña et al., 2018) is as follows: sensitivity ranges 87%-89% for English and 78-91% for Spanish; specificity ranges 81%-88% for English and 81%-88% for Spanish. Preliminary data for the BESA-ME (from the field test version; Peña et al., 2016) indicated that sensitivity ranges were between 63%-71% for English morphosyntax and between 80%-91% for Spanish morphosyntax. Specificity ranges were 85%-89% for English morphosyntax and 92%-97% for Spanish morphosyntax.

In this study, we used the morphosyntax subtest of the BESA and BESA-ME to classify children to either the DLD or the typical development group. The morphosyntax subtest was administered in both Spanish and English. To target a variety of grammatical morphemes and sentence structures, these tests utilized cloze and sentence repetition tasks that have been shown to discriminate children with DLD in either language. Once administered in both languages, we used the age cutoffs in the BESA and BESA-ME manuals (Peña et al., 2008; Peña et al., 2018) with the “best language” score to identify children as DLD when this score was equal to or below the age cutoff score. Using this grouping procedure yielded a group of 62 children in the DLD group (43 boys, 19 girls)

and 93 children in the TD group (43 boys, 48 girls). Table 1 summarizes child and family demographics for the children in the two groups.

Language Experience: length of English exposure and language dominance.

To further understand the linguistic characteristics of these children, we recorded measures of length of English exposure and language dominance. As part of a language use parental questionnaire, parents were asked about the age at which children were first exposed to English. Using this information, we calculated a length of English exposure in months (M=34.69, SD=18.14). Likewise, to have a continuous measure of language dominance, we considered the relative difference between the child's scores in Spanish and English for the BESA and BESA-ME morphosyntax subtests. Figure 1 shows the language dominance distribution for the 155 participants, more positive scores in this continuum indicate higher morphosyntax abilities in Spanish, and more negative scores indicate higher morphosyntax abilities in English. This relative difference in morphosyntax abilities between the two languages was utilized as a proxy for language dominance for this study.

Measures

Transcript preparation and broad language sample measures.

We gathered English language samples using story retell and story generation elicitation tasks using Frog picture books (Mayer, 1967, 1973, 1974; Mayer & Mayer, 1975) following SALT elicitation protocols for these stories (Miller & Iglesias, 2017). Transcription of these English language samples was performed by research assistants who were native speakers of English who had not been involved in the assessment sessions. All transcripts were coded using SALT conventions (Miller & Iglesias, 2017).

Broad language sample measures were collected, including: (a) number of utterances (NU); (b) mean length of utterance in words (MLU_w); (c) number of total words (NTW); and (d) number of different words (NDW). The measures of NU, MLU_w, NTW, and NDW were automatically generated by the SALT software (Miller & Iglesias, 2017).

Tense/agreement Composite Measures.

The language samples collected in the English language skills session were used to compute three composite measures for tense/agreement, indicating diversity (tense marker total; TMT), productivity (tense/agreement productivity; TAP) and a measure for tense/agreement accuracy (a modified version of the finite verb morphology composite; FVMC_m). These composites were computed from non-repetitive, spontaneous, and intelligible utterances with *explicit subjects* as detailed in the following sections.

Tense Marker Total (TMT). The TMT is a measure of tense/agreement morpheme diversity computed using methodology based on Hadley and Short (2005). The TMT was computed by marking fifteen forms of tense/agreement morphemes: (a) third-person singular regular present tense *-s*; (b) regular past tense *-ed*; (c) five forms of the copula verb *to be* (i.e. *am, are, is, was, were*); (d) five forms of the auxiliary verb *to be* (i.e. *am, are, is, was, were*); and (e) three forms of the auxiliary verb *to do* (i.e. *do, does, did*). The calculation of the TMT measure awarded 1 point for the initial occurrence of each of these tense/agreement morpheme forms. The range of possible scores for the TMT was between 0 and 15. Certain rules applied for counting tense/agreement morphology diversity. In general, the TMT recorded only correct production of tense/agreement morpheme forms, however, overregularization of regular past tense *-ed* (e.g. *catch/ed, go/ed*) was counted towards this measure. Uncontracted forms used with pronouns (e.g.

she is going, we are happy) and contracted forms used with nouns (e.g. *the frog's jumping, the boy's sad*) were counted towards the TMT as well as most productions with negative contractions (e.g. *the lady wasn't happy, he doesn't want, she didn't go*). However, tense morphemes contracted to a pronoun (e.g. *she's going, he's sad*) and the specific negative contracted forms *don't* and *ain't* were excluded from this measure to avoid counting measures potentially learned on a rote basis.

Tense/agreement productivity (TAP). The TAP score is a measure of tense/agreement morpheme productivity also computed using methodology in Hadley and Short (2005). This measure was originally called productivity score (Hadley & Short, 2005) and later renamed TAP (Hadley & Holt, 2006). The fifteen forms of tense/agreement morphemes used for the TMT calculation were categorized into 5 categories for purposes of calculating the TAP: (a) third-person singular regular present tense *-s*; (b) regular past tense *-ed*; (c) copula *to be*; (d) auxiliary verb *to be*; and (e) auxiliary verb *to do*. Each of these categories could be awarded up to 5 points for "sufficiently different" uses of the morphemes. The range of scores possible for the TAP measure calculation was from 0 to 25. For third-person singular present tense *-s* and regular past tense *-ed*, "sufficiently different" meant the use of distinct lexical verbs could be awarded additional points in each category (e.g. *kick/3s, jump/3s* and *push/ed, laugh/ed*). For copula *be*, auxiliary *be*, and auxiliary *do* verbs, "sufficiently different" meant use of different subject + verb combinations (e.g. *the boy is walking*, vs. *the frog is walking; he is walking*, vs. *he was walking*). Certain rules applied for which forms were counted towards the TAP measure of productivity. Similar to the TMT calculation, the TAP measure included the overregularization of regular past tense *-ed*, but all other errors

were excluded. Contracted forms of copula and auxiliary *to be* were counted except when contracted to pronominal forms (e.g. *the frog's jumping* and *the boy's sad* were counted, but *she's screaming* and *they're happy* were not). Likewise, most verbs with negative contractions were included except in the case of the specific negative forms *don't* and *ain't* which were not counted (e.g. *he doesn't want* counted; while *they don't want* did not count).

Tense/agreement accuracy (FVMC_m). We computed a modified version of the finite verb morphology composite (FVMC_m) as a measure of accuracy to reflect percent of correct usage in obligatory contexts for the target tense/agreement morphemes (Bedore & Leonard, 1998; Rice & Wexler, 1996, 2001; Rice et al., 1998). The FVMC_m was constructed to reflect accuracy for all tense/agreement morpheme categories included in the TMT and TAP measures. Thus, the FVMC_m composite measure for accuracy was based on correct production of the following verb forms: (a) third-person singular regular present tense *-s*; (b) regular past tense *-ed*; (c) copula *to be*; and (d) auxiliary verb *to be* and (e) auxiliary verb *to do*. All obligatory contexts for these forms were included in the calculation of the FVMC_m. The FVMC_m was calculated by dividing the number of correct productions by the number of obligatory contexts. This number was then multiplied by 100 to generate a percentage.

Experimental Measure: Morphosyntactic Development Levels

In addition to calculating the composite tense/agreement morpheme composite measures, we wanted to reflect children's development of English morphosyntax considering their age and their language experience. To do this, we combined the tense/agreement composite measures into development attainment categories we

designated as morphosyntactic development levels. The composite tense/agreement composite measures tend to be highly dependent on each other; for example, tense/agreement diversity (as measured by TMT) usually precedes productivity (as measured by TAP). Accuracy as measured by the FVMC_m, on the other hand, can be subject to measurement error depending on the diversity and number of obligatory contexts and can lead to overestimation of a child's grammatical knowledge (Hadley & Short, 2005; Rispoli et al., 2009; Rispoli et al., 2012).

The purpose of the morphosyntactic development levels was to take into account children's stage of English morphosyntactic development from initial diversity through productivity, evaluating accuracy only once enough diversity and productivity had been attained. Every new level would imply that the child met the requirements for the previous level in addition to the requirements of the current level. An ordinal scale, using levels 0 – 5, was created to combine characteristics of tense/agreement composite measures. Level 0 was set as the *base level*. Children at base level had not developed the diversity of tense agreement characteristics in English to advance to level 1. We observed that in our language sample set, the most widely used tense/agreement morpheme category was the copula *to be*, followed by a block of the regular past tense *-ed* and the auxiliary verb *to be* (see Figure 2 for details). For this reason, these three verb tense categories were selected for advancement to level 1 where children demonstrated *initial diversity* of these morpheme forms. Children had to produce verb forms in at least 2 of these 3 selected initial diversity verb tense categories to advance to level 1. Level 2 was deemed *productive diversity*; to advance to level 2, children had to meet the criteria for level 1 and to produce at least 2 "sufficiently different" correct productions of initially diverse

tense/agreement morpheme categories (we used the definition of "sufficiently different" from the TAP calculation). In level 3, we observed *increased productive diversity*; children had to meet the criteria for level 2 and had to add at least one more tense/agreement morpheme category where they were productive (producing at least 2 "sufficiently different" correct productions). In level 4, we started to look at *initial accurate productivity* where we required that children meet all criteria for level 3 and also demonstrate an overall accuracy of at least 50% in all obligatory contexts for all 5 tense/agreement morpheme categories. Finally, level 5 was where children demonstrated *accurate productivity*; children met criteria for level 4 and demonstrated 80% overall accuracy in all obligatory contexts for all 5 tense/agreement morpheme categories. This approach to translating performance into achievement levels was derived from approaches used to determine achievement levels in cognitive processing tasks such as the auditory serial memory task to assess short-term memory for non-linguistic auditory information (see Yim, 2006 for a detailed description of the task; and Ebert, 2011; Ebert et al., 2014 for details of the achievement level calculations). The overall range of morphosyntactic development levels for our sample is shown in Figure 3.

Procedures

Children were recruited for the study by distributing information about language screening for bilingual children, letters to bilingual early-childhood education centers, and in collaboration with independent school districts. When working with specific schools in the area, parent consent forms and language use questionnaires were sent home with each child by the school's speech-language pathologist (SLPs) and / or the classroom teachers. The language use questionnaire contained information about family demographics and

language use in the home and school. Children who met the eligibility criteria and for whom we received completed parental consent forms and language use questionnaires were included in the study.

Assessment sessions

Multiple sessions were planned for each participant to conduct the battery of assessments. Language skills in Spanish and English were assessed in different sessions conducted by research assistants who were native speakers of either Spanish or English and who were trained to administer all standardized assessments, to elicit language samples, and maintain child engagement. After children provided assent, we collected non-verbal IQ measures (KBIT-II; Kaufman & Kaufman, 2004), hearing screening, language samples, and a language assessment measure using the morphosyntax subtest of the BESA and the BESA-ME (Peña et al., 2008; Peña et al., 2018) in both Spanish and English. Language samples were elicited using story retell and story generation with a set of Frog picture books (Mayer, 1967, 1973, 1974; Mayer & Mayer, 1975) using the SALT elicitation protocols for these books. They were recorded using electronic recording equipment which would provide good quality audio and minimize disruptions to language sample gathering. Session length was determined by child cooperation and engagement and each session was capped at a maximum of 60 minutes.

Reliability

As part of this study, we took precautions to ensure that at each step of transcription, coding, and scoring, data reliability was maintained. All research assistants involved in transcription and coding were blind to the group status of the individual participants. A team of bilingual Spanish-English research assistants received training in

transcription and coding per SALT conventions (Miller & Iglesias, 2017). Once English language samples were transcribed, another research assistant reviewed all the transcriptions for completeness and marked any discrepancies they found. A third research assistant reviewed language samples to make a final determination on the transcripts.

After all English transcripts were finalized, a group of 2 research assistants, native speakers of English, were trained by the first author on the identification of the target tense/agreement morphemes and obligatory contexts. Different codes were used to identify all initial, subsequent, correct/incorrect, and omitted obligatory uses of the target tense/agreement morphemes. The language samples were all given to both research assistants for coding independently. Then the coding was compared to determine differences on each transcript. Any differences in coding were reviewed by both research assistants to make a final determination. Once all transcripts were coded, they were reviewed by the first author for completeness. SALT research version 18 (Miller & Iglesias, 2017) was used to review all utterances with target tense/agreement morphemes and to derive the raw data for the calculations of the TMT, TAP, and FVMC_m measures.

Data Analytic Strategy

We first conducted a series of independent t-test analyses to compare the group means to address the first research question, whether measures of English tense marking diversity, productivity, and accuracy (i.e. TMT, TAP, or FVMC_m) predicted diagnostic classification. A logistic multivariate regression model was used with diagnostic classification (TD or DLD) as the dependent variable using each of the measures separately to determine whether each measure accurately predicted diagnostic classification. We controlled for demographic characteristics, including chronological age,

language dominance, and length of exposure to English to determine the best model. The resulting classification tables from the logistic regression were entered into the MedCalc online calculator (https://www.medcalc.org/calc/diagnostic_test.php) to calculate sensitivity, specificity, and positive and negative likelihood ratios (LR+ and LR-, respectively) for diagnostic accuracy. Interpretation of sensitivity and specificity was done based on the guidelines set by Plante and Vance (1994) for acceptable diagnostic values: values less than .80 are considered unacceptable for diagnostic purposes, values between .80-.89 are considered fair, and values of .90 and above are considered of good diagnostic value. To interpret likelihood ratios, we used the guidelines set by Dollaghan (2007). Using these guidelines, a positive likelihood ratio indicates the level of confidence that LR+ values equal to 3 or greater were considered moderately positive and LR+ values equal to 10 or greater were considered largely positive. LR- values equal to 0.30 or less were considered moderately negative and LR- values equal to 0.10 or less were considered largely negative. Confidence intervals of likelihood ratios provided an additional test of whether the calculated LRs fall within these informative ranges.

The second research question asked whether we can use these measures of tense diversity, productivity, and tense/agreement accuracy to create levels of English morphosyntactic development and use these along with demographic characteristics to discriminate bilingual children with DLD and TD. To address this question, we ran a series of logistic multivariate regression models to determine the predictive ability of the morphosyntactic development levels in addition to demographic factors (age, language dominance, length of English exposure) in discriminating children with DLD and TD. We

also calculated sensitivity, specificity, and likelihood ratios for diagnostic accuracy information to determine the clinical utility of this measure.

Results

Group Comparisons

The groups were compared in terms of child characteristics and all language measures were compared to look at group differences. Table 2 shows group means, independent t-test comparisons, and effect sizes between the two groups for demographic characteristics, normed-reference test results, language sample standard measures, and tense/agreement measures. For demographic characteristics, we included age, language dominance, and length of English exposure. Two significant differences between the groups concerning child characteristics were of interest. The first one was age; the DLD group's average age was 8.5 months younger than the TD group's average, $t(147.9)=5.257$, $p<.001$, $d=0.824$. Secondly, this study recruited children with a variety of language dominance profiles and wide-ranging exposure to English. Although there was no significant difference for the DLD and TD groups in their group means of the continuous measure for language dominance (which is based on Spanish and English BESA / BESA-ME morphosyntax subtest standard scores, $t(126.6)=-0.663$, $p=.508$, $d=-0.093$), there was a significant difference in terms of length of English exposure. Children in the DLD group had been exposed to English on average 7 months less than the children in the TD group, $t(135.9)=2.415$, $p=.023$, $d=0.398$. These differences in demographic characteristics between the two groups were of interest for our analyses and were added as covariates to the regression models to control for potential bias.

Results for normed-reference tests included the Matrices subtest of the KBIT- II (Kaufman & Kaufman, 2004) and the BESA/BESA-ME morphosyntax subtest standard scores in Spanish, English and "Best Language" (Peña et al., 2008; Peña et al., 2018). There was no difference in non-verbal IQ between the groups based on the mean scores on the KBIT-II. As expected, since this was the measure used for diagnostic classification, children in the DLD and TD groups had significantly different BESA/BESA-ME "Best Language" scores, $t(153)=20.180$, $p<.001$, $d=3.309$. However, it is important to note that both BESA/BESA-ME Spanish and English scores were also significantly different between both groups despite the varied language dominance profiles in both groups.

The groups were also compared on standard measures derived from the English language samples from the story retell and story generation tasks. The total number of utterances in the language sample was the only measure that did not significantly differ between the DLD and TD groups, $t(114.3)=-0.629$, $p=.515$, $d=-0.107$. The group means for all of the standard language sample measures (MLUw, NTW, NDW, and the number of obligatory contexts for the target verb forms) were significantly different for the DLD and TD groups. Further, the mean scores for all of the tense/agreement composite measures were also significantly different between the groups (TMT: $t(153)=6.101$, $p<.001$, $d=1.000$; TAP: $t(150.7)=7.302$, $p<.001$, $d=1.132$; FCMV_m: $t(147)=6.409$, $p<.001$, $d=1.077$); Morphosyntactic Development Levels: $t(152.3)=8.227$, $p<.001$, $d=1.233$). Detailed results for descriptive statistics and group comparisons on all of these measures are summarized in Table 2.

Diagnostic Classification of Tense Agreement Measures

For research question 1, logistic regression analysis was employed to predict the probability that a participant would be classified in either the DLD or TD group given each of the tense/agreement measures. The best model using the diversity of tense/agreement morphemes to predict diagnosis classification, used the TMT score, chronological age (centered at 48 months for ease of interpretation), and language dominance as predictor variables and was statistically significant, $\chi^2(3, N=155) = 53.670$, $p<.001$. This model using TMT for diversity correctly classified 76.1% of all children. The best model using the productivity of tense/agreement morphemes to predict diagnosis classification, used the TAP score, chronological age (centered at 48 months), and language dominance as predictor variables and was also statistically significant, $\chi^2(3, N=155)=60.890$, $p<.001$. This model using TAP for productivity correctly classified 80% of all children. Finally, the best model for accuracy of tense/agreement morphemes predicting diagnosis classification, used the FVMC_m measure, chronological age (centered at 48 months), and language dominance as predictors and was statistically significant as well, $\chi^2(3, N=149)=51.857$, $p<.001$. This model using the FVMC_m for accuracy of tense/agreement morphemes was able to correctly classify 79.2% of all children.

Table 3 shows the resulting sensitivity, specificity, and likelihood ratios for the tense/agreement composite measures for classification accuracy. All of these tense/agreement measures fell short of the .80 guideline for acceptable sensitivity. The range of sensitivity values for these composites suggests that between 50-82% of the children with DLD were correctly identified, which is not in the acceptable range.

However, for all three tense/agreement composite measures, specificity values exceed the .80 criteria for adequate diagnostic accuracy. The 95% confidence interval for these values indicates that between 76-92% of the TD group was accurately classified using these measures.

Positive and negative likelihood ratios (LR+ and LR-, respectively) and their 95% confidence intervals were calculated for the tense/agreement composite measures. LR+ values indicate the ability of a positive test result to confirm a particular diagnosis (i.e. increased probability that a positive test confirms a diagnosis of DLD); LR+ values between 3-10 are considered moderately to very informative (Dollaghan, 2007). Panel (A) in

Figure *4* shows the forest plots for these LR+ values and confidence intervals. All three tense agreement measures had LR+ values over the moderately informative threshold (these values are summarized in Table 3). LR- values indicate the ability of a negative test result to confirm the absence of a diagnosis (i.e. decreased probability of having a diagnosis of DLD given a negative test result, confirming classification in the TD group) with LR- values between 0-0.3 considered moderately to very informative (Dollaghan, 2007). Panel (B) in

Figure *4* shows the forest plots for these LR- values and confidence intervals. All three tense agreement measures fell short of the threshold for moderately informative negative tests (values and confidence intervals are summarized in Table 3).

Diagnostic Classification of Morphosyntactic Development Levels

For research question 2, we used logistic regression analysis to examine if our experimental measure for morphosyntactic development levels could be used to

discriminate children with DLD from children with typical language development in Spanish-English bilinguals. For the best model to predict diagnostic accuracy (DLD vs. TD), the predictor variables were morphosyntactic development levels (as a categorical value), chronological age (centered at 48 months for interpretation purposes), and language dominance. The logistic regression model was statistically significant, $\chi^2(7, N=155) = 81.700$, $p<.001$. This model using morphosyntactic development levels correctly classified 85.2% of all children. To evaluate these measures for diagnostic accuracy, we also calculated sensitivity, specificity, and likelihood ratios, these results are summarized in Table 3. Morphosyntactic development levels, when controlling for age and language dominance, had acceptable sensitivity (.871) and specificity (.839) which indicate that 87.1% of children were correctly identified as DLD and 83.9% correctly classified as TD. The LR+ and LR- for the morphological levels were both within the moderate to highly informative range (LR+ = 5.40; LR- = 0.15).

Discussion

The purpose of this study was to analyze the discriminant accuracy of tense/agreement morpheme composite measures of diversity, productivity, and accuracy (TMT, TAP, and FVMC_m, respectively) from English language samples to identify children with DLD in Spanish-English bilingual children in the United States. Compared to previous studies assessing tense/agreement morpheme production which have generally consisted of small sample sizes (for exceptions, see Gutiérrez-Clellen et al., 2006; Paradis et al., 2013), we examined a group of 155 early-elementary Spanish-English bilingual children representing a wide range of language dominance profiles and length of exposure to English. We wanted to assess the potential clinical utility of these measures in English,

as tense/agreement morphemes are an area of difficulty for both English-monolingual children and bilingual children with DLD. Further, inspired by these measures, we created a new measure of morphosyntactic development levels which considered a developmental perspective of English acquisition and the broad range of language experience (language dominance and exposure to English) that we found in our sample. While we recognize that best practices recommend assessment of bilingual children in both of their languages (Bedore & Peña, 2008; Kohnert, 2010), meaningful measures using English tense/agreement morphemes to discriminate bilingual children with DLD would provide important information in clinical practice for monolingual SLPs who have bilingual children in their caseloads.

Tense/Agreement Morpheme Composite Measures

In this study, we considered three tense/agreement composite measures that were derived from language samples elicited through story retell and story generation tasks. Variations of the FVMC have been used for assessing language using accuracy in clinical and research for monolinguals (Bedore & Leonard, 1998; Gladfelter & Leonard, 2013; Rice et al., 1998) and with bilingual children (Gutierrez-Clellen et al., 2008) with high discriminant accuracy. However, there are some limitations of using accuracy as a measure to evaluate the language of children in the early development of English tense/agreement systems (Hadley et al., 2018; Rispoli et al., 2009). Composite measures of accuracy may overestimate a child's grammatical abilities if they measure accuracy in total obligatory contexts in an immature system where sentence diversity is only just developing. Since bilinguals in early elementary years are likely to be in emerging stages of English tense/agreement development, we also examined TMT and TAP measures

which assess diversity and productivity in developing systems. These two measures, TMT and TAP, were developed to reduce the potential for measurement weaknesses and have shown clinical utility in identifying monolingual children at risk for DLD (Gladfelter & Leonard, 2013; Hadley & Holt, 2006; Hadley & Short, 2005; Rispoli et al., 2009; Rispoli et al., 2012).

Our results suggest that the three tense/agreement composite measures used in the study –TMT, TAP, and FVMC_m– all significantly differentiate between bilingual children with and without DLD. This supports previous findings that show that measures of diversity and productivity (Potapova et al., 2018) and measures of accuracy (Gutierrez-Clellen et al., 2008) differentiate between typical and atypical language development in Spanish-English bilinguals in the early elementary years. All of these results also replicate findings with English-monolinguals where children with DLD perform significantly worse than children with typical language skills in tense/agreement morpheme composite measures (Bedore & Leonard, 1998; Gladfelter & Leonard, 2013; Hadley & Holt, 2006; Hadley & Short, 2005; Rice et al., 1998; Rispoli et al., 2009; Rispoli et al., 2012).

We evaluated the results of the discriminant accuracy calculations against guidelines set by Plante and Vance (1994) where measures with sensitivity and specificity values above 90% are considered to have good discrimination and values between 80-89% are considered acceptable. Likelihood ratios were evaluated using guidelines from Dollaghan (2007) where LR+ values equal to 3 or greater were considered moderately positive, LR+ values equal to 10 or greater were considered largely positive, LR- values equal to 0.30 or less were considered moderately negative, and LR- values equal to 0.10 or less were considered largely negative. In our study, all three composite measures of

tense/agreement displayed acceptable specificity and LR+s in the moderate to highly informative range, but sensitivity and LR- values failed to meet the threshold for acceptable and informative values. This is in contrast to previous results with English-monolingual children where these measures have shown both sensitivity and specificity values in the acceptable to good range and likelihood ratios in the moderate to highly informative range (Bedore & Leonard, 1998; Gladfelter & Leonard, 2013; Gutiérrez-Clellen et al., 2006; Gutierrez-Clellen & Simon-Cereijido, 2007; Gutierrez-Clellen et al., 2008). This is not surprising, as these measures were developed to measure the growth of tense/agreement morpheme systems in English-monolingual children. However, despite our sample being vastly different from the studies obtaining acceptable discriminant accuracy, these results suggest that even across a broad range of ages, language dominance profiles, and lengths of exposure to English, these measures could provide useful information for screening Spanish-English bilingual children. High-scoring children in these measures have a high probability of being correctly categorized as having typical developing language skills.

Morphosyntactic Development Levels

The composite tense/agreement composite measures are highly related to one another as they measure different aspects of the same developing language skill. These measures provide important information with regards to English morphosyntactic development and demonstrate good diagnostic accuracy in detecting DLD in English-monolingual children from toddlers to early elementary (Gladfelter & Leonard, 2013; Hadley & Short, 2005). Further, as we were able to replicate in this study, these measures also predict group differences in Spanish-English bilinguals (Gutierrez-Clellen et al.,

2008; Potapova et al., 2018). However, these measures do not have adequate sensitivity to be able to detect DLD, in other words, under-identifying bilingual children with DLD. We believe this is because they do not take into account English morphosyntactic development within the context of the child's individual language experience, including exposure and language dominance.

Inspired by these tense/agreement composite measures, we created levels of English morphosyntactic development combining attainment thresholds of diversity, productivity, and finally, accuracy once a diverse and productive system had been established in English. We wanted to take advantage of the wide range of ages, dominance profiles, and language exposure profiles in our sample to provide insight into the developmental trajectory of tense/agreement production in Spanish-English bilingual children in the early elementary years. Our results indicate that measuring children's level of morphosyntactic development along with information on age and language dominance significantly differentiates Spanish-English bilingual children with DLD from their typically developing peers. Both sensitivity and specificity values for this measure show fair discriminant accuracy per guidelines in Plante and Vance (1994). Further, the positive and negative likelihood ratios (and their respective 95% confidence intervals) for the morphosyntactic development levels were fully within the ranges considered moderately to highly informative (Dollaghan, 2007). This may provide meaningful information for English-speaking SLPs with potential clinical utility for identifying Spanish-English bilingual children with language disorders when assessment of both languages is not possible.

Future Research and limitations

We acknowledge, that best practices stipulate that, when possible, we should evaluate both languages in young Spanish-English bilinguals (Bedore & Peña, 2008; Kohnert, 2010). As we are aware, this is not always possible. We are encouraged by our results and the potential clinical utility of using tense/agreement composite measures in English to evaluate the development of bilingual children's morphosyntactic systems. Our large sample of children consisted of a wide range of ages, language dominance profiles, and length of exposure to English. In our analysis, we accounted for differences in age and language experience by adding these as covariates to our regression models. However, we recognize that having more comparable groups (i.e., DLD and TD) specifically concerning age distribution and language dominance would allow us to perform a more in-depth analysis of the classification accuracy of these measures. In our analysis, we opted to take advantage of the larger sample size and diverse group, which we felt better represents the heterogeneity of the population of Spanish-English bilingual children in the United States, and this still yielded promising results.

In this study, we used story retell and story generation tasks to elicit children's language samples. In our samples, we found that the copula verb *to be* was the most used tense/agreement morpheme category while third-person singular present tense -*s* was the morpheme least used followed by the forms of the auxiliary verb *to do*. In contrast, previous studies of English monolingual children using spontaneous language samples found that the sequence of morphosyntactic development follows the pattern: copula verb *to be* > [third-person singular present tense -*s*, regular past tense -*ed*, the auxiliary verb *to do*] > auxiliary verb *to be* (Gladfelter & Leonard, 2013; Rispoli & Hadley, 2011; Rispoli

et al., 2009; Rispoli et al., 2012). These differences may be related to the number of opportunities available in these stories for eliciting obligatory contexts for some of these tense/agreement morpheme categories. Future studies using other elicitation tasks for language sample analysis may be necessary to replicate these results.

Conclusion

Our results indicate that English morphosyntactic development for early elementary Spanish-English bilinguals can be meaningfully described by assessing bilingual children's tense/agreement morpheme diversity and productivity in combination with accuracy in addition to language experience (dominance and exposure). Further, these levels of morphosyntactic development have the potential for clinical use with acceptable discriminant accuracy in screening children with typical language who may still be developing their morphosyntactic systems in English. Specifically, when we combined the tense/agreement morpheme composite measures into attainment levels of morphosyntactic development, we found that we were able to identify children with DLD in this group of heterogeneous bilinguals with fair values of classification accuracy. These are promising indications that English-only measures may have informative value in assessment of bilingual children's language when used in combination with parental report, language dominance, and language exposure information (Bedore & Peña, 2008; Bedore et al., 2018; Gutierrez-Clellen et al., 2008).

References

American Speech-Language Hearing Association (ASHA). (2018). *Demographic Profile of ASHA Members Providing Bilingual Services*. https://www.asha.org/uploadedFiles/Demographic-Profile-Bilingual-Spanish-Service-Members.pdf

Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: a discriminant function analysis. *Journal Of Speech, Language, And Hearing Research: JSLHR, 41*(5), 1185-1192. https://doi.org/10.1044/jslhr.4105.1185

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *The International Journal of Bilingual Education and Bilingualism, 11*, 1-29.

Bedore, L. M., Pena, E. D., Anaya, J. B., Nieto, R., Lugo-Neris, M. J., & Baron, A. (2018). Understanding Disorder Within Variation: Production of English Grammatical Forms by English Language Learners. *Lang Speech Hear Serv Sch, 49*(2), 277-291. https://doi.org/10.1044/2017_LSHSS-17-0027

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2 consortium. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary

Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58*(10), 1068-1080. https://doi.org/10.1111/jcpp.12721

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & Consortium, C. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *Plos One, 11*(7).

Castilla-Earls, A. P., Restrepo, M. A., Pérez-Leroux, A. T., Gray, S., Holmes, P., Gail, D., & Chen, Z. (2016). Interactions between bilingual effects and language impairment: Exploring grammatical markers in Spanish-speaking bilingual children. *Applied Psycholinguistics, 37*(5), 1147-1173. https://doi.org/10.1017/S0142716415000521

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Brookes Pub. http://www.loc.gov/catdir/toc/ecip0716/2007017769.html

Ebert, K. D. (2011). *A Comparison Between Nonlinguistic Cognitive Processing Treatment and Traditional Language Treatment for Bilingual Children with Primary Language Impairment* (Publication Number 3466931) [Ph.D., University of Minnesota]. ProQuest Dissertations & Theses Global. Ann Arbor. http://search.proquest.com.ezproxy.lib.uh.edu/docview/885874442?accountid=7107

https://libraries.uh.edu/resolver?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+Global&atitle=&title=A+Comparison+Between+Nonlinguistic+Cognitive+Processing+Treatment+and+Traditional+Language+Treatment+for+Bilingual+Children+with+Primary+Language+Impairment&issn=&date=2011-01-01&volume=&issue=&spage=&au=Ebert%2C+Kerry+Danahy&isbn=978-1-124-80838-3&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Ebert, K. D., Kohnert, K., Pham, G., Rentmeester Disher, J., & Payesteh, B. (2014). Three Treatments for Bilingual Children With Primary Language Impairment: Examining Cross-Linguistic and Cross-Domain Effects. *Journal of Speech, Language and Hearing Research, 57*(1), 172-186. https://doi.org/10.1044/1092-4388(2013/12-0388)

Genesee, F., Paradis, J., & Crago, M. B. (2004). *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning*. Brookes Publishing Company. http://ezproxy.lib.uh.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED491779&site=ehost-live

http://www.brookespublishing.com/store/books/genesee-6865/index.htm

Gladfelter, A., & Leonard, L. B. (2013). Alternative Tense and Agreement Morpheme Measures for Assessing Grammatical Deficits During the Preschool Period [Article]. *Journal of Speech, Language & Hearing Research, 56*(2), 542-552. https://doi.org/10.1044/1092-4388(2012/12-0100)

Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics, 29*(1), 41-65. https://doi.org/10.1017/S014271640808003X

Grinstead, J., Baron, A., Vega-Mendoza, M., De la Mora, J., Cantú-Sánchez, M., & Flores, B. (2013). Tense marking and spontaneous speech measures in Spanish specific language impairment: a discriminant function analysis. *Journal Of Speech, Language, And Hearing Research: JSLHR, 56*(1), 352-363. https://doi.org/10.1044/1092-4388(2012/11-0289)

Gusewski, S., & Rojas, R. (2017). Tense Marking in the English Narrative Retells of Dual Language Preschoolers [Article]. *Language, Speech & Hearing Services in Schools, 48*(3), 183-196. https://doi.org/10.1044/2017_LSHSS-16-0093

Gutierrez-Clellen, V. F., Restrepo, M. A., Bedore, L. M., Peña, E., & Anderson, R. (2000). Language Sample Analysis in Spanish-Speaking Children: Methodological Considerations [Article]. *Language, Speech & Hearing Services in Schools, 31*(1), 88-98.

http://ezproxy.lib.uh.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=3284088&site=ehost-live

Gutiérrez-Clellen, V. F., Restrepo, M. A., & Simón-Cereijido, G. (2006). Evaluating the Discriminant Accuracy of a Grammatical Measure With Spanish-Speaking Children [Article]. *Journal of Speech, Language & Hearing Research, 49*(6), 1209-1223. https://doi.org/10.1044/1092-4388(2006/087)

Gutierrez-Clellen, V. F., & Simon-Cereijido, G. (2007). The discriminant accuracy of a grammatical measure with Latino English-speaking children. *J Speech Lang Hear Res, 50*(4), 968-981. https://doi.org/10.1044/1092-4388(2007/068)

Gutierrez-Clellen, V. F., Simon-Cereijido, G., & Wagner, C. (2008). Bilingual children with language impairment: A comparison with monolinguals and second language learners. *Appl Psycholinguist, 29*(1), 3-19. https://doi.org/10.1017/S0142716408080016

Hadley, P. A., & Holt, J. K. (2006). Individual Differences in the Onset of Tense Marking: A Growth-Curve Analysis [Article]. *Journal of Speech, Language & Hearing Research, 49*(5), 984-1000. https://doi.org/10.1044/1092-4388(2006/071)

Hadley, P. A., McKenna, M. M., & Rispoli, M. (2018). Sentence Diversity in Early Language Development: Recommendations for Target Selection and Progress

Monitoring [Article]. *American Journal of Speech-Language Pathology, 27*(2), 553-565. https://doi.org/10.1044/2017_AJSLP-17-0098

Hadley, P. A., & Short, H. (2005). The Onset of Tense Marking in Children at Risk for Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 48*(6), 1344-1362. http://ezproxy.lib.uh.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ742729&site=ehost-live http://dx.doi.org/10.1044/1092-4388(2005/094)

Heilmann, J. J. (2010). Myths and Realities of Language Sample Analysis. *Perspectives on Language Learning and Education, 17*(1), 4-8. https://doi.org/10.1044/lle17.1.4

Heilmann, J. J., Miller, J. F., & Nockerts, A. (2010). Using Language Sample Databases [Article]. *Language, Speech & Hearing Services in Schools, 41*(1), 84-95. https://doi.org/10.1044/0161-1461(2009/08-0075)

Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (2010). Twenty-Year Follow-Up of Children With and Without Speech-Language Impairments: Family, Educational, Occupational, and Quality of Life Outcomes [Article]. *American Journal of Speech-Language Pathology, 19*(1), 51-65. https://doi.org/10.1044/1058-0360(2009/08-0083)

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test* (Second Edition ed.). Pearson, Inc.

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

Leonard, L. B. (1998). *Children with Specific Language Impairment*. The MIT Press. https://books.google.com/books?id=aWK1QgAACAAJ

Leonard, L. B. (2014). *Children with specific language impairment* (Second edition ed.). The MIT Press.

Leonard, L. B., Haebig, E., Deevy, P., & Brown, B. (2017). Tracking the Growth of Tense and Agreement in Children With Specific Language Impairment: Differences Between Measures of Accuracy, Diversity, and Productivity. *Journal of speech, language, and hearing research : JSLHR, 60*(12), 3590-3600. https://doi.org/10.1044/2017_JSLHR-L-16-0427

Mayer, M. (1967). *A boy, a dog and a frog*. Penguin Books.

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

Mayer, M. (1974). *Frog goes to dinner*. Penguin Books.

Mayer, M., & Mayer, M. (1975). *One Frog Too Many*. Penguin Books.

Miller, J., & Iglesias, A. (2017). *Systematic Analysis of Language Transcripts (SALT), Research Version 2018 [Computer Software].* In SALT Software, LLC.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of Child Psychology and Psychiatry, 57*(11), 1247-1257. https://doi.org/10.1111/jcpp.12573

Paradis, J. (2005). Clinical Forum: Grammatical Morphology in Children Learning English as a Second Language: Implications of Similarities With Specific Language Impairment. *Language, Speech, and Hearing Services in Schools, 36*(3), 172-187. https://doi.org/10.1044/0161-1461(2005/019)

Paradis, J., & Crago, M. (2000). Tense and Temporality: A Comparison Between Children Learning a Second Language and Children with SLI [Article]. *Journal of Speech, Language & Hearing Research, 43*(4), 834. https://doi.org/10.1044/jslhr.4304.834

Paradis, J., Crago, M., Genesee, F., & Rice, M. (2003). French-English Bilingual Children With SLI: How Do They Compare With Their Monolingual Peers? [Article]. *Journal of Speech, Language & Hearing Research, 46*(1), 113. https://doi.org/10.1044/1092-4388(2003/009)

Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning. Second Edition*. Brookes Publishing Company. http://ezproxy.lib.uh.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED517294&site=ehost-live http://www.brookespublishing.com/store/books/paradis-70588/index.htm

Paradis, J., Rice, M. L., Crago, M., & Marquis, J. (2008). The Acquisition of Tense in English: Distinguishing child second language from first language and specific language impairment. *Appl Psycholinguist, 29*(4), 689-722. https://doi.org/10.1017/s0142716408080296

Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating Children With Language Impairment Among English-Language Learners From Diverse First-Language Backgrounds [Article]. *Journal of Speech, Language, and Hearing Research, 56*(3), 971-981. https://doi.org/10.1044/1092-4388(2012/12-0050)

Peña, E. D., Bedore, L. M., Gutierrez-Clellen, V. F., Iglesia, A., & Goldstein, B. A. (2008). *Bilingual English-Spanish Assessment - Middle Extension Experimental Test Version (BESA-ME).* Unpublished manuscript.

Peña, E. D., Bedore, L. M., Gutierrez-Clellen, V. F., Iglesia, A., & Goldstein, B. A. (2016). *Bilingual English-Spanish Assessment - Middle Extension Field Test Version (BESA-ME).* Unpublished manuscript.

Peña, E. D., Gutierrez-Clellen, V. F., Iglesias, A., Goldstein, B., & Bedore, L. M. (2014). *Bilingual English-Spanish Assessment (BESA)*. A-R Clinical Publications.

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

Plante, E., & Vance, R. (1994). Selection of preschool language tests: a data-based approach [Article]. *Language, Speech & Hearing Services in Schools, 25*(1), 15. http://ezproxy.lib.uh.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=9409142484&site=ehost-live

Potapova, I., Kelly, S., Combiths, P. N., & Pruitt-Lord, S. L. (2018). Evaluating English Morpheme Accuracy, Diversity, and Productivity Measures in Language Samples of Developing Bilinguals. *Language, Speech, and Hearing Services in Schools, 49*(2), 260-276. https://doi.org/10.1044/2017_LSHSS-17-0026

Restrepo, M. A. (1998). Identifiers of Predominantly Spanish-Speaking Children With Language Impairment [Article]. *Journal of Speech, Language & Hearing Research, 41*(6), 1398-1411. https://doi.org/10.1016/0008-8749(92)90224-d

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal Of Speech And Hearing Research, 39*(6), 1239-1257. https://doi.org/10.1044/jshr.3906.1239

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

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over Time: The Longitudinal Course of Tense Acquisition in Children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 41*(6), 1412-1431. https://doi.org/10.1044/jslhr.4106.1412

Rispoli, M., & Hadley, P. A. (2011). Toward a theory of gradual morphosyntactic learning. In I. Arnon & E. V. Clark (Eds.), *Experience, Variation and Generalization : Learning a first language* (pp. 15-34). John Benjamins Publishing Company. http://ebookcentral.proquest.com/lib/uh/detail.action?docID=730698

Rispoli, M., Hadley, P. A., & Holt, J. K. (2009). The growth of tense productivity. *Journal Of Speech, Language, And Hearing Research: JSLHR, 52*(4), 930-944. https://doi.org/10.1044/1092-4388(2009/08-0079)

Rispoli, M., Hadley, P. A., Holt, J. K., Smith, A., & Loeb, D. (2012). Sequence and System in the Acquisition of Tense and Agreement. *Journal of Speech, Language, and Hearing Research, 55*(4), 1007-1021. https://doi.org/10.1044/1092-4388(2011/10-0272)

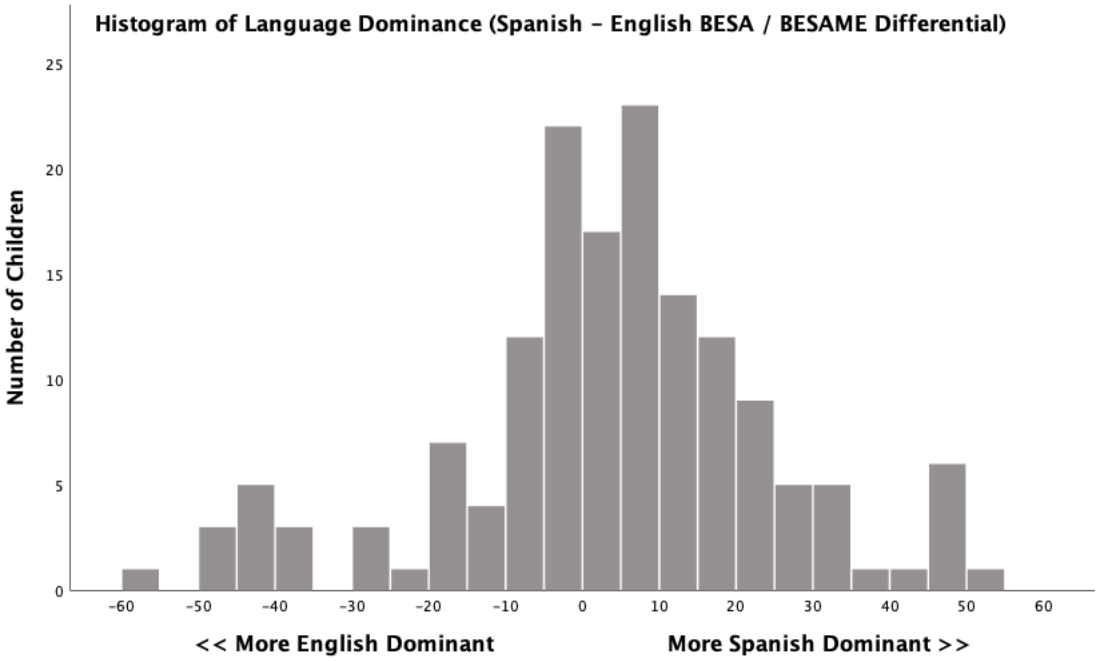
Rojas, R., & Iglesias, A. (2009). Making a Case for Language Sampling [Article]. *ASHA Leader, 14*(3), 10-13. https://doi.org/10.1044/leader.FTR1.14032009.10

Sullivan, A., & Bal, A. (2013). Disproportionality in special education: Effects of individual and school variables on disability risk. *Exceptional Children, 79*(4), 475-494.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of speech, language, and hearing research : JSLHR, 40*(6), 1245-1260. https://doi.org/10.1044/jslhr.4006.1245

U.S. Census Bureau. (2015). *American Community Survey. Detailed Languages Spoken at Home and Ability to Speak English for the Population 5 Years and Over: 2009-2013.* https://www.census.gov/data/tables/2013/demo/2009-2013-lang-tables.html

Yim, D. (2006). *Auditory and visual pattern learning and language skills in children and adults* (Publication Number 3240493) [Ph.D., University of Minnesota]. ProQuest Dissertations & Theses Global. Ann Arbor. http://search.proquest.com.ezproxy.lib.uh.edu/docview/305309105?accountid=7107 https://libraries.uh.edu/resolver?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+Global&atitle=&title=Auditory+and+visual+pattern+learning+and+language+skills+in+children+and+adults&issn=&date=2006-01-01&volume=&issue=&spage=&au=Yim%2C+Dongsun&isbn=978-0-542-95185-5&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/



Note: The difference in BESA/BESA-ME morphosyntax scores (BESA Standard Score in Spanish – BESA Standard Score in English) is being used as a proxy for language dominance in a continuum. Scores on the positive side of the spectrum indicate more Spanish language dominance and on the negative side indicate more English language dominance.

Figure 1. Language Dominance Distribution

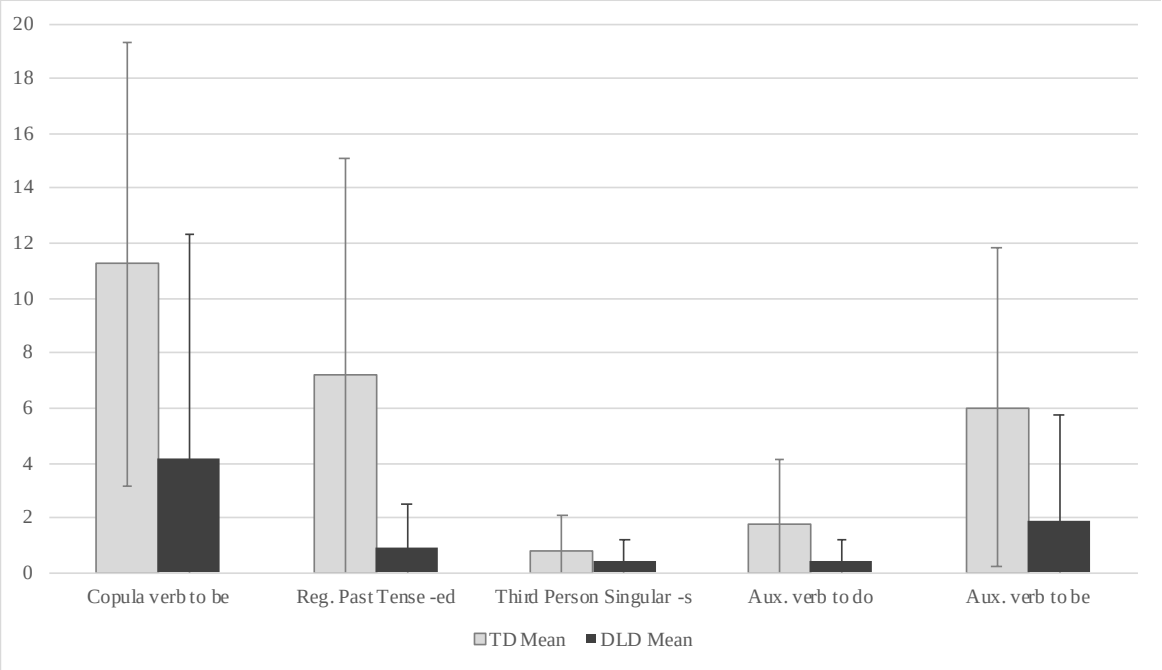


Figure 2. Mean productions by tense/agreement morpheme category and diagnostic classification

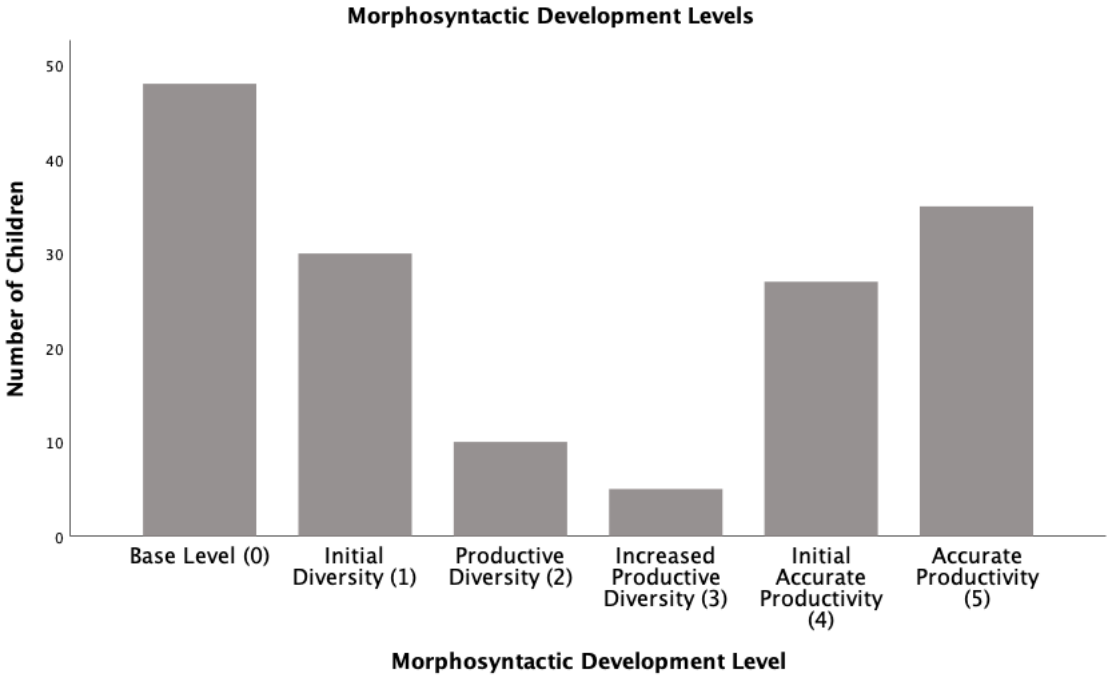


Figure 3. Summary of Morphosyntactic Development Levels

Table 1. Child and family demographics (*N*=155).

	DLD (*N*=62)	TD (*N*=93)
	n (%)	*n* (%)
Child Gender		
Female	19 (30.6%)	48 (51.6%)
Male	43 (69.4%)	43 (46.2%)
Receiving Speech-Language Services Now		
Yes	36 (58.1%)	13 (14%)
No	7 (11.3%)	45 (48.4%)
No response	19 (30.6%)	35 (37.7%)
Maternal Education		
Highschool	39 (62.9%)	41 (44.1%)
Some College	15 (24.2%)	5 (5.4%)
Associate degree	4 (6.5%)	9 (9.7%)
Bachelor's degree	6 (9.7%)	16 (17.2%)
Graduate degree	9 (14.5%	15 (16.1%)
No Response	2 (3.2%)	7 (7.5%)
Receiving Free / Reduced Lunch		
Yes	53 (85.5%)	63 (67.7%)
No	6 (9.7%)	23 (24.7%)
No response	3 (3.8%)	7 (7.5%)

Note: TD = Typically developing language skills; DLD = Developmental language disorder.

Table 2. Descriptive statistics and results of Independent sample T-tests (*N*=155)

	DLD (*n*=62)		TD (*n*=93)					
	M	*SD*	*M*	*SD*	*t*	*df*	*p*	*d*
Child Age (mths)	60.8	8.9	69.3	11.1	5.257[a]	147.9	<.001	0.824
Language Dominance	4.0	9.8	2.0	26.0	-0.663[a]	126.6	.508	-0.093
English Exposure (mths)	30.6	14.8	37.7	19.8	2.415[a]	135.9	.023	0.398
Normed-referenced results								
KBIT-II	101.6	12.6	101.2	13.0	-0.186	153	.853	-0.030
BESA/BESA-ME Spanish	70.5	8.5	91.7	16.4	10.530[a]	145.6	<.001	1.537
BESA/BESA-ME English	66.5	9.1	89.7	17.6	10.719[a]	145.2	<.001	1.563
BESA/BESA-ME Best Lang	72.5	7.9	101.1	9.1	20.180	153	<.001	3.309
Language sample measures								
No. of utterances	88.8	35.5	85.3	29.6	-0.629	114.3	.515	-0.107
MLUw	3.8	1.3	5.5	1.6	7.076	153	<.001	1.160
NTW	265.6	157.8	418.8	176.5	5.518	153	<.001	0.905
NDW	82.6	35.6	112.6	42.4	4.602	153	<.001	0.755
No. Obligatory Contexts	18.0	16.6	36.6	19.7	6.118	153	<.001	1.003
Tense/agreement composite measures								
TMT	2.7	2.3	5.2	2.7	6.101	153	<.001	1.000
TAP	4.1	3.9	9.4	5.2	7.302[a]	150.7	<.001	1.132
FVMC_m	0.37	0.28	0.67	0.27	6.409	147	<.001	1.077
Morphosyntactic Development Levels	0.95	1.25	3.11	2.01	8.227[a]	152.3	<.001	1.233

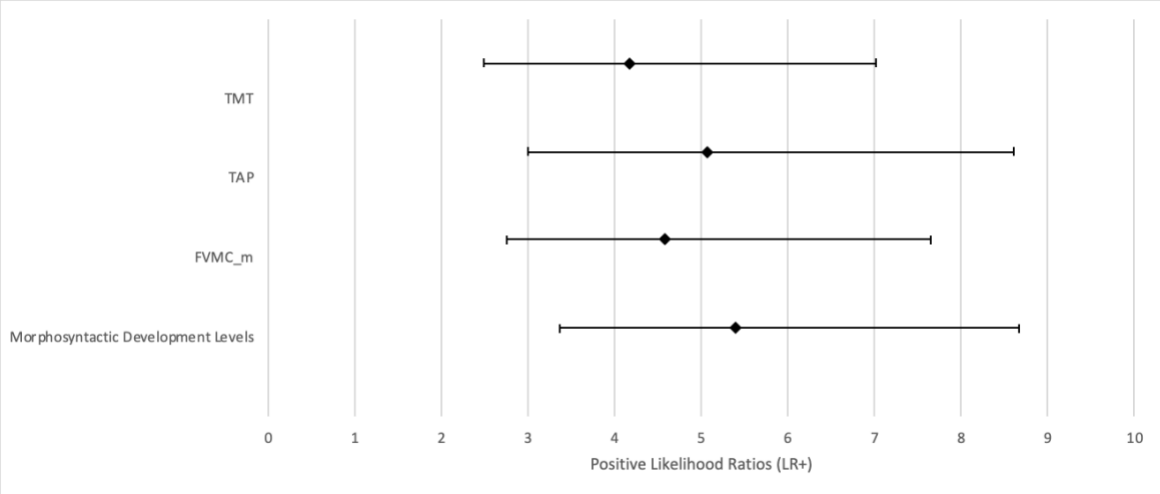
Note: The table reports group means and standard deviations for each variable as well as results of independent sample t-tests comparing the group means. [a] indicates that the t-test was significant for Levene's test for equality of variances, so equal variances were not assumed in the independent sample t-test calculations. Cohen's *d* is reported as the effect size for the comparison. Child Age and Length of English exposure are measured in months. DLD = Developmental language disorder; TD = Typically developing language skills; KBIT-II = Non-verbal subtest of the Kaufman Brief Intelligence Test, Second Edition; BESA = Bilingual English-Spanish Assessment; BESA-ME = Bilingual English-Spanish Assessment, Middle Elementary; MLUw = Mean length of utterance in words; NTW = Number of total words; NDW = Number of different words; TMT = Tense marker total; TAP = Tense agreement productivity; FVMC_m = modified version of finite verb morphology composite.

Table 3. Sensitivity, specificity, and likelihood ratios by measure.

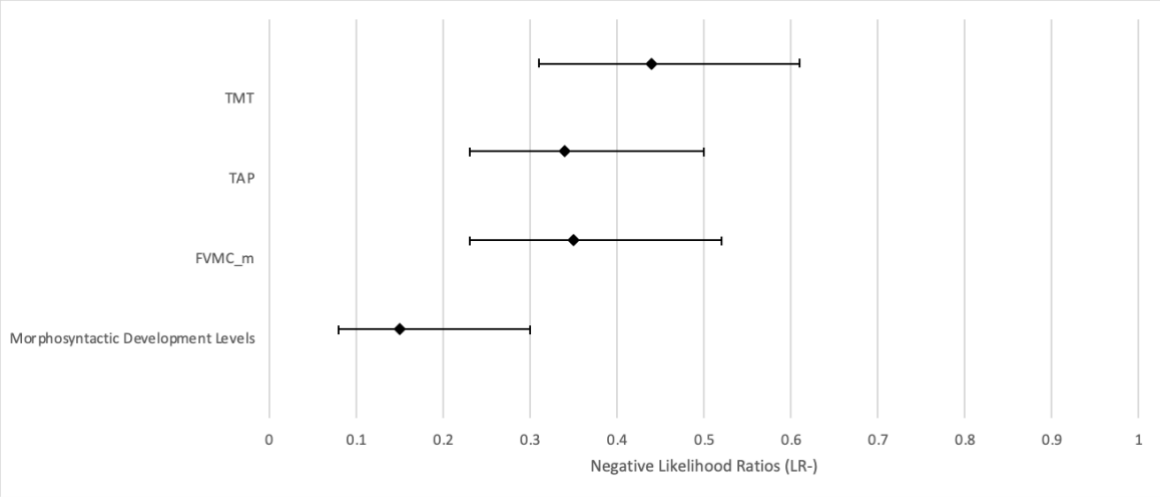
	Sensitivity [95% CI]	Specificity [95% CI]	LR+ [95% CI]	LR- [95% CI]	PPV [95% CI]	NPV [95% CI]	Accuracy [95% CI]
Tense/Agreement Composite Measures							
TMT	.629	**.850**	4.18	0.44	.253	.966	.833
	[.497, .748]	[.760, .915]	[2.49, 7.02]	[0.31, 0.61]	[.168, .363]	[.953, .975]	[.765, .888]
TAP	.710	**.860**	5.08	0.34	.292	.973	.849
	[.581, .818]	[.773, .923]	[2.99, 8.61]	[0.23, 0.50]	[.195, .411]	[.961, .982]	[.783, .901]
FVMC_m	.707	**.846**	4.59	0.35	.271	.973	.836
	[.573, .819]	[.755, .913]	[2.76, 7.65]	[0.23, 0.52]	[.183, .383]	[.959, .982]	[.766, .891]
Morphosyntactic Development Levels	**.871**	**.839**	5.40	0.15	.305	.988	.841
	[.762, .943]	[.748, .907]	[3.36, 8.67]	[0.08, 0.30]	[.214, .413]	[.977, .994]	[.774, .895]

Note: The table reports results for diagnostic accuracy calculations, including 95% confidence intervals (95% CI). PPV, NPV, and Accuracy calculations are dependent on disorder prevalence which has been taken as 7.5% for these calculations (Norbury et al., 2016). Bold text indicates sensitivity and specificity values that are have reached the criterion for adequate diagnostic accuracy of 0.80 or higher (Plante & Vance, 1994). LR+ = Positive likelihood ratio; LR- = Negative likelihood ratio; PPV = Positive predictive value; NPV = Negative predictive value; TMT = Tense marker total; TAP = Tense agreement productivity; FVMC_m = modified version of finite verb morphology composite.

(A) Positive Likelihood Ratios



(B) Negative Likelihood Ratios



Note: Markers indicate the likelihood ratio (LR) values and horizontal error bars indicate the 95% confidence interval for the values. For positive likelihood ratios (LR+) shown in (A), values further to the right are considered more informative. For negative likelihood ratios (LR-) shown in (B), values further to the left are considered more informative. Shaded areas in both forest plots indicate LRs considered moderately to highly informative (Dollaghan, 2007). TMT = Tense marker total; TAP = Tense agreement productivity; FVMC_m = modified version of finite verb morphology composite.

Figure 4. Positive and Negative Likelihood Ratios and 95% Confidence Intervals