This is an Accepted Manuscript of a book chapter published by Routledge in Typical and Atypical Language Development in Cultural and Linguistic Diversity on 29 September 2023, available online: http://www.routledge.com/9781003251194.

9. Grammatical Profiles of Mandarin-English Bilingual Children at Risk for Developmental Language Disorder

Li Sheng; Man Yang; Yao Du; Elizabeth Peña; Lisa Bedore

https://orcid.org/0000-0001-7092-6798
https://orcid.org/0000-0002-8077-8896
https://orcid.org/0000-0003-2858-2420
https://orcid.org/0000-0002-6119-7658
https://orcid.org/0000-0002-1973-3939

Abstract

Developmental language disorder (DLD) is under-identified in US bilingual children who speak Asian languages. To achieve a better understanding of the manifestations of DLD in one of the largest Asian American populations, the current study examines the grammatical profiles of Mandarin-English bilingual children at risk for DLD. Fifty-five 4- to 7-year-old bilingual children were administered an experimental bilingual screener that consisted of four subtests. Parents completed questionnaires about children's language use and proficiency. Four children were identified as at risk (AR) for DLD using a set of five indicators of risk status. A matched sample of eight children formed the typically developing (TD) comparison group. Welch's T-tests were conducted to compare the groups on 11 English grammatical structures, seven Mandarin grammatical structures, and sentence repetition scores in both languages. The AR group showed significant deficits in English items targeting verb morphology, noun morphology, complex sentences and sentence repetition. Mandarin items targeting passive voice, quantifiers, aspect, relative clause, and sentence repetition separated the two groups. As the first investigation of the grammatical profiles of Mandarin-English bilingual children at risk for DLD, the results highlighted several potential markers of DLD in this population that warrant further investigations.

Introduction

Children with Developmental Language Disorder (DLD, also known as Specific Language Impairment) have significant language learning problems in the absence of intellectual, social emotional, neurological, or sensory disorders (Bishop et al., 2017; Leonard, 2014). According to studies conducted in the United States, DLD is under-diagnosed in the general population (Hendricks et al., 2019; McGregor, 2020); in children of Asian descent, under-diagnosis is even more exacerbated (Cooc, 2019). Chinese-English bilingualism is common in multilingual societies such as Malaysia and Singapore. The number of Chinese-English bilingual children is increasing rapidly in English-speaking countries (Australian Bureau of Statistics, 2014; Statistics Canada, 2016; US Census Bureau, 2013). Thus, there is a pressing need to understand the manifestations of DLD in this population. The present study aims to further our knowledge about the grammatical profiles of bilingual Mandarin-English children at risk for DLD and move us closer towards identifying sensitive markers of DLD in this population. In the following sections, we summarize findings of

grammatical deficits in children with DLD who speak English and Chinese; discuss recent studies of grammatical development in Chinese-English bilinguals; and present our research questions.

English DLD

There is mounting evidence that low accuracy in grammatical morphemes that mark tense and subject-verb agreement are hallmarks of English DLD in preschool and early school-age years. Specifically, the regular past tense *-ed*, the third-person singular present tense *-s*, the auxiliary *do*, and the copula and auxiliary *be* are regarded as *clinical markers* of DLD in English because they reliably and robustly separate young children with and without DLD (e.g., Bedore & Leonard, 1998; Rice & Wexler, 1996). Children with DLD also show significant weaknesses in the use of the irregular past tense, possessive *'s*, prepositions, noun plurals, articles and aspects (e.g., Grela et al., 2004; Leonard, 1995; Leonard et al., 2007; Marshall & van der Lely, 2012; Oetting & Rice, 1993; Rice & Wexler, 1996). However, the differences between DLD and typically developing (TD) groups on these forms tend to be smaller and overlapping performance between groups is often observed. With regard to morphosyntax, children with DLD produce shorter utterances (e.g., Rice et al., 2010) and fewer sentential arguments (e.g., Grela & Leonard, 1997). They struggle with constructions that require insertions of "*do*", such as negation: "These men like pizza and these men *do not*"); constructions that involve syntactic movement, such as *Wh*-questions (e.g., "Who is the package from?"; Hewitt et al., 2005; Rice et al., 2009), passives (e.g., "The cat
is chased by the dog"; Leonard et al., 2006), and relative clauses (e.g., "The boy that the dog chased got away."; Frizelle & Fletcher, 2014; Hesketh, 2006).

The grammatical deficits in monolingual English-speaking children with DLD are best characterized as a profile difference (Leonard, 2014). While weaknesses are noted for most grammatical features under investigation, a subset of the features – those related to verb tense and agreement marking, are regarded as diagnostically most useful and hence referred to as clinical markers. Whether or not these markers can differentiate Mandarin-English bilingual children with and without DLD is an empirical question we attempt to address in the current study.

English grammatical development in Chinese-English bilinguals

Studies with TD Chinese-English bilinguals in North America show that English verb morphology is vulnerable in this population, and some children may show early plateau (Paradis et al., 2016) or failure to reach mastery after long-term immersion (Jia & Fuse, 2007). Similarly, TD Mandarin-English bilinguals in Singapore also showed late emergence in learning English tense-marking morphemes (Brebner et al., 2016). With regard to noun morphology, evidence of learning challenges is presented in a longitudinal study of Chinese-English sequential bilinguals who immigrated to the United States at different ages (Jia, 2003). In contrast to the early mastery of the noun plural –s in monolingual children, three of the 10 children did not master this form after five years of L2 immersion.

In the only published study on Chinese-English bilingual DLD, Ooi and Wong (2012, study 2) examined the utility of language sample analysis in predicting diagnostic status among bilingual

children in Malaysia, where a non-standard vernacular of English is spoken. Three measures were derived from English conversational samples: mean length of utterance in words, lexical diversity, and Index of Productive Syntax (IPSyn, Scarborough, 1990) – which evaluates the emergence of 56 morphosyntactic features. The DLD group showed shorter utterance length and lower IPsyn scores than TD peers. Discriminant functions indicated that only IPSyn scores differentiated DLD from TD children (*n*=9 per group, age range=3;8-5;11) at statistically significant levels. The overall classification accuracy was 77.8%, a value close to being clinically useful. Because the IPSyn offers an aggregate score of various morphosyntactic features, which specific grammatical features contributed the most to DLD discrimination is unclear.

There is evidence suggesting that English verb morphology may not be reliable for identifying DLD in bilingual children. Paradis and colleagues examined English verb morphology production in bilinguals with and without DLD (age range= 4-8 years old) from diverse cultural and linguistic backgrounds (Blom & Paradis, 2013, 2015; Paradis et al., 2013). Some of the participants' first languages (L1s) do not mark tense (i.e., Cantonese, Mandarin, and Vietnamese) whereas others spoke tense-marking L1s. As a whole, the DLD group showed significantly weaker performance in English verb morphology than TD peers. However, among children whose L1 do not mark tense on the verb, the TD and DLD groups did not differ (Paradis et al., 2013). Moreover, L1 development scores based on parent report using the Alberta Language and Development Questionnaire (ALDeQ, Paradis et al., 2010) contributed the most in discriminating DLD from TD whereas English verb morphology production only contributed weakly.

Chinese DLD

Aspect markers are regarded as true inflectional morphemes in Cantonese and Mandarin, two closely related Chinese languages with extremely sparse grammatical morphology. A recent systematic review found that aspect markers are the most studied grammatical element in the Chinese DLD literature (Sheng et al., under review). Compared to TD peers, Cantonese- (Fletcher et al., 2005) and Mandarin-speaking (Yu et al., 2019) children with DLD are less likely to produce perfective and imperfective aspect markers and show more restricted use and inefficient processing of aspect markers over a range of verb types. With regard to morphosyntax, Chinese children with DLD exhibit deficits in producing and comprehending passive sentences (Leonard et al., 2006) and relative clauses (He & Yu, 2013; Yu et al., 2017). Difficulties producing nominal classifiers, an obligatory component in noun phrases that use numerals (e.g., 两条狗 'two classifier dog') or demonstrative (e.g., 这条狗 'this classifier dog'), are also noted in Chinese children with DLD (Hao et al., 2018; Stokes & So, 1997).

While group-level differences across multiple features are observed, it is unclear to what extent these features reliably separate children with and without DLD on a child-by-child basis, i.e., constitute diagnostic markers in Chinese. In a study of Cantonese-speaking children with DLD, Stokes et al. (2006) found that as in English (e.g., Conti-Ramsden et al., 2001; Redmond et al., 2019), a quick sentence repetition task resulted in good classification accuracy (87%). Similarly, Wang et al. (2022) reported classification accuracy of 94% when using a Mandarin sentence repetition task to identify preschoolers with DLD. The sentence repetition task is a potential clinical

marker across many languages because it can be designed to target vulnerable linguistic features in the respective languages.

Chinese grammatical development in Chinese-English bilinguals

To the best of our knowledge, no studies have examined markers of DLD in the L1 of Chinese-English bilinguals. Only a handful of studies have examined the L1 grammatical development of Chinese-English bilinguals growing up in an English-speaking environment (Hao et al., 2019; Jia & Paradis, 2015; 2020; Kan, 2019; Wei & Lee, 2001). Mandarin-English bilinguals in the US showed very low production of Mandarin passive sentences, relative clauses, quantifiers (e.g., 所有, 都 – all, 每 – every), classifiers, and reduplicated verbs in narrative samples (Hao et al., 2019). Similarly, Cantonese-English bilingual children in the UK and the US showed low accuracy in elicited production of classifiers (Kan, 2019) and restricted use of classifiers and quantifiers in conversational and narrative samples (Wei & Lee, 2001). Finally, Mandarin-English bilingual children in Canada showed more restricted use of classifiers (Jia & Paradis, 2015) and lower accuracy of relative clause production (Jia & Paradis, 2020, study 1) than Mandarin monolinguals in China. Overall, these patterns suggest incomplete learning of the L1 in Chinese bilinguals growing up in an English-dominant environment.

Summary and Questions

Different trajectories of English morphological development are reported in Chinese-English bilinguals relative to English monolinguals and relative to bilinguals whose L1 marks verb tenses (Blom & Paradis, 2013, 2015; Brebner et al., 2016; Jia, 2003; Jia & Fuse, 2007; Paradis et al., 2013, 2016). Thus, we may expect that the English grammatical features that can differentiate DLD from TD in Chinese-English bilinguals would be different from those in English monolinguals. The literature on monolingual Chinese DLD has identified aspect markers, passives, relative clause, classifier, and the sentence repetition task as potential markers. Many of these features are vulnerable and subject to incomplete acquisition in TD Chinese-English bilinguals. But the extent to which these features differentiate TD from DLD in bilinguals has not been investigated. To the best of our knowledge, only one published study has specifically focused on Chinese-English bilingual DLD (Ooi & Wong, 2012). However, only English was assessed in this study and the aggregated measure did not tell us which specific grammatical feature(s) held promise in differentiating diagnostic groups.

To address these gaps, the current study aims to delineate the grammatical profiles of Mandarin-English bilingual children with and without risk for DLD and identify potential markers of impairment for this population. We collected information about children's language ability and language use from their primary caregivers. In the absence of tools to assess Mandarin-English bilingual children, we devised our own tool, the *Bilingual English Mandarin Oral Screener* (BEMOS), an experimental screener consisting of morphosyntax and semantics subtests in English and Mandarin. Following common approaches (Bedore et al., 2018), we identified a subset of the children as having risk for DLD and selected TD peers who were matched on age and language use. To delineate the DLD group's grammatical profile, we compared their performance to TD peers on

each of the English and Mandarin grammatical features from the morphosyntax subtests of the screener to answer the following research questions:
1) What English grammatical features differentiated bilingual children at risk (AR) for DLD from their TD peers?
2) What Mandarin grammatical features differentiated the AR group from their TD peers?
In addition, we conducted an item analysis to identify a set of items that yielded substantial differences between DLD and TD groups ($\geq$ 30%, Allen & Yen, 2002) and entered them into a discriminant function to answer the following question:
3) What is the classification accuracy of the English and Mandarin grammatical items that show differences of 30% or more between AR and TD groups?

Method

Participants

Participants were recruited from a major southern city in the USA, through word of mouth, announcements at weekend Chinese schools, and mailing lists for Chinese parents. Fifty-five Mandarin-English bilingual children ages 4 to 7 years participated. Two children were referred by a certified bilingual Speech-Language Pathologist as suspected of having language delay. No other children had any existing concerns regarding language development. All children were reported to have normal vision and hearing and no prior diagnosis of speech, language, or neurological impairment. Questionnaires and interviews regarding current language use, language history, parental education and occupation were conducted at the beginning of the first testing session. All children were from Mandarin-speaking homes and had been exposed to both English and Mandarin for at least two years at the time of testing. Language use was evaluated using the *Bilingual Input Output Survey* (BIOS, Peña et al., 2018) that queried the child's input and output at school and at home on a typical weekday and typical weekend on an hourly basis. All children had at least 10% daily exposure to each language at the time of testing.

Determination of risk status. Consistent with the literature on bilingual language assessment (Bedore et al., 2018; Castilla-Earls et al., 2020), we used multiple sources of information to gather convergent evidence to identify children at risk for DLD. A set of five indicators were used including parental rating of child language proficiency, and the child's performance on the Mandarin morphosyntax (MMS), Mandarin semantics (MSM), English morphosyntax (EMS), and English semantics (ESM) subtests of the BEMOS. The parental rating measure, *Inventory To Assess Language Knowledge* (ITALK, Peña et al., 2018), consisted of 5 questions about children's language comprehension, speech production, grammar, vocabulary, and sentence production in English and in Mandarin. Parents rated each area using a 0-5 point scale. For example, for sentence production a score of 0 indicates inability to produce sentences, a score of 1 indicates that the child uses 1-2 word phrases in the target language, and a 5 indicates that the child uses 5+ word phrases or sentences in the target language. Responses to the 5 questions were averaged for each language for each child. Mandarin and English average scores were compared and the higher score was retained. Then means and standard deviations (SDs) were calculated respectively for younger (4-5 year-olds; n= 25) and older children (6-7 year-olds; n= 30). Scores in the better language that were more than

1SD below the mean for the child's age group were flagged as indicating a possible risk for DLD.

To establish cut-off scores to identify cases of risk on each subtest of the BEMOS, means and SDs for each subtest were first calculated for subsamples of the 55 children who completed the BEMOS. The subsamples consisted of children who were balanced bilinguals (those who use the language 40-60% of their daily life) or dominant (those who use the language 60-90% of their daily life) in the target language based on daily use reported on the BIOS. Specifically, for the MMS and MSM subtests, means and SDs were calculated from children who used Mandarin 40-90% of the time (n=13 for the younger group; n=13 for the older group). For the EMS and ESM, means and SDs were calculated from children who used English at 40-90% of the time (n=20 for the younger group; n=29 for the older group). For each of the four subtests, cases were identified as "no risk" if the child's score was above -1SD of the age group mean. Scores below -1SD were flagged as at risk.

Across the 5 indicators (parent rating of proficiency and 4 BEMOS subtests), if children fell in the risk range on 4 (n=1) or 5 (n=1) indicators, they were identified as having risk for DLD. If a child had 3 risk indicators and 2 of the 3 came from the child's stronger language (i.e., the language having a higher proficiency rating), then the child was also identified as having risk for DLD (n=2). Table 1 presents the detailed profile of each of the four AR children. We then reviewed the information for all the children who had zero risk indicator (n=22) and matched them to a child in the AR group if they were within 4 months in age and 15% in language use. Eight children met these criteria and were included in the TD comparison group. Each AR child had at least one but sometimes up to four TD matches. One of the two children suspected of having a language delay from the bilingual Speech-Language Pathologist's referral were in the AR group. As shown in Table 1, the AR and TD groups were matched on age, maternal education level, and language use, all ps>10. The TD group had higher English and Mandarin proficiency ratings, though the difference was significant only for English proficiency [*Welch's* $t(9)=1.98$, $p=.04$]. The mothers and fathers of all 12 children selected for the current study spoke Mandarin as their native language and use Mandarin with their children from birth. Systematic exposure to English began at various time points but all children had at least two years of English exposure at the time of testing. Table 1 presents the detailed profile of each of the four AR children and the eight TD controls.

Table 1. Descriptive information, subtest scores, and Welch's t test results of the at risk (AR) and typically developing (TD) groups.

	Age[a]	Sex	MatEd[b]	English use[c]	English proficiency[d]	Mandarin proficiency[d]	EMS[e]	MMS[e]	ESM[e]	MSM[e]	EMS1[f]	MMS1[f]
AR1	50	F	21	68%	2.75	3.25	.33	.10	.24	.53	.21	0
AR2	62	M	19	41%	2.50	3.50	.42	.47	.71	.76	.25	.43
AR3	62	F	18	21%	3.50	3.25	.40	.45	.76	.82	.18	.32
AR4	74	M	22	46%	4.00	4.75	.40	.33	.71	.65	.18	11
Mean	**62.00**		**20.00**	**44%**	**3.19**	**3.69**	**.39**	**.34**	**.60**	**.69**	**.21**	**.21 (.20)**
(SD)	**(9.80)**		**(1.83)**	**(19%)**	**(.69)**	**(.72)**	**(.04)**	**(.17)**	**(.25)**	**(.13)**	**(.03)**	
TD1	51	F	18	61%	4.00	3.50	.54	.63	.76	.71	.46	.71
TD2A	59	F	15	45%	4.50	4.50	.77	.51	1.00	.59	.79	.54
TD2B	61	F	22	52%	4.50	4.25	.73	.67	.94	.65	.75	.75
TD3	65	M	16	30%	2.33	4.25	.63	.65	.82	.59	.68	.75
TD4A	75	M	16	49%	4.25	4.50	.81	.71	.88	.76	.79	.71
TD4B	75	M	17	59%	4.33	4.75	.79	.78	.88	.88	.81	.93
TD4C	77	M	24	51%	3.75	4.50	.73	.82	1.00	.82	.71	.96
TD4D	78	M	22	57%	4.75	4.50	1.00	.76	1.00	.82	1.00	.93
Mean	**67.63**		**18.75**	**51%**	**4.05**	**4.34**	**.75**	**.69**	**.91**	**.73**	**.74**	**.79 (.15)**
(SD)	**(10.04)**		**(3.41)**	**(10%)**	**(.76)**	**(.38)**	**(.14)**	**(.10)**	**(.09)**	**(.11)**	**(.15)**	

Note. AR1 and AR3 each had one TD match. AR2 had two TD matches (TD2A and TD2B). AR4 had four TD matches (TD4A-TD4D). Risk indicators for each AR child were underlined. a. Age is reported in months. b. MatEd= maternal education reported in years. c. Language use is measured through the *Bilingual Input Output Survey* (BIOS, Peña et al., 2018). d. Language proficiency rating is based on parent reports using a 0-5 point scale on the *Inventory To Assess Language Knowledge* (ITALK, Peña et al., 2018). e. Proportion correct responses on the English morphosyntax (EMS), Mandarin morphosyntax (MMS), English semantics (ESM), and Mandarin semantics (MSM) subtests of the *Bilingual English Mandarin Oral Screener* (BEMOS, Sheng et al., in development). F. Proportion correct responses for the 28 most differentiating English items and the 28 most differentiating Mandarin items.

Materials

The BEMOS EMS and ESM subtest items were selected from the English subtests of the *Bilingual English Spanish Oral Screener* (BESOS, Peña et al., 2008) and *Bilingual English Spanish Assessment-Middle Extension* (BESA-ME, Peña et al., 2010), two established screeners used for identifying Spanish-English bilingual children with DLD (Lugo-Neris et al., 2015; Peña et al., 2020). The BEMOS MSM subtest items were adapted from the Spanish semantics subtest of the BESOS because the stimuli were considered to be culturally appropriate for children growing up in the United States. The BEMOS MMS contained five items adapted from the Spanish morphosyntax subtest of the BESOS that assessed passive and prepositional phrases and 46 items developed by the first author. The morphosyntax subtests consist of two task types: cloze and sentence repetition. Table 2 and Table 3 present sample English and Mandarin cloze items.

Table 2. Sample English morphosyntax items with targeted response in bold

Target	Model	Prompt
3rd personal singular	Everyday these dogs drink water.	And here this dog does it too. What does he do everyday? Everyday the dog (**drinks** water).
Auxiliary+negation	These men have mustaches.	And these men (**don't** have moustaches).
Passive	The table was painted by the boy.	And here, the chair (**is/was painted** by the girl)
Past tense *–ed*	Today he is walking his dog.	And yesterday, he did it too. What did he do yesterday? Yesterday he (**walked** the dog).
Auxiliary + *ing*	Mary and Jeremy want to watch TV.	They are doing it now. What are they doing here. They (**are watching** TV).
Copula	In this pond there is one fish.	And in this pond there (**are** two fish).
Possessive *'s*	Look, this is a doctor and this is the doctor's watch.	This is a lady, and this is the (**lady's** watch).
Prepositional phrase	Now we are going to say where the cats are. These cats are in the jar.	Here the cats are (**on** the plate).
Plural	This girl has an apple.	And here, what does she have? The girl/she (has many **apples**).
Question inversion	This girl is watching her friend open a birthday present.	Tell me what she says. (**What is it?**)
Relative clause	Here the lady is playing the violin that is blue.	Here the lady is (playing **the violin that is red**).

The cloze task elicits particular words, phrases, or clauses from the child to complete a sentence.

The sentence repetition task requires the child to repeat a sentence verbatim. The MMS subtest contains 27 cloze items that cover 7 Mandarin grammatical features: passive, possessive, prepositional phrase, classifier, quantifier, aspect marker, and relative clause. All cloze items are accompanied by pictorial prompts to facilitate elicitation of the target structure. The Mandarin sentence repetition task contains 3 sentences segmented into 24 individual words or phrases for the purpose of scoring. The EMS subtest contains 36 cloze items that cover 11 grammatical features: third-person singular, auxiliary and negation, passive, past tense, auxiliary + progressive-ing, copula, possessive, prepositional phrase, plural noun, question inversion, and relative clause. The English sentence repetition task contains 3 sentences. Twelve key words or phrases from these sentences were selected for scoring because these elements were found to be the most discriminating in the development of the BESOS.

Table 3. Sample Mandarin morphosyntax items with targeted response in bold

Target	Model	Prompt
Passive	这扇窗户被男孩打碎了。 This window was broken by the boy.	那么这扇窗户呢？(被女孩打碎了）。 And here, this window (**was broken** by the girl.)
Possessive *'s*	看这儿有一个叔叔，这个是叔叔的钱包。 Look, this is an uncle and this is the uncle's wallet.	这儿有一个阿姨，这个是(**阿姨的**钱包)。 This is an auntie, and this is the (**auntie's** wallet).
Prepositional phrase	看这幅图，这个小女孩坐在椅子上。 Look here, the girl is sitting on the chair.	那再看这幅图，这个小女孩站在哪里呢？她(站**在**椅子**后面**)。 Now here, where is the girl standing? She is standing (**behind** the chair).
Classifier	这里有三朵花。 Here, there are three duo (classifier) flowers.	那么这里呢？这里有(三**棵**树) And here, there are three **ke** (classifier) trees.
Quantifier	看这里，每一只小猫都在喝水。 Here, every cat is drinking water.	那么这里呢？(**每一只**小猴**都**在吃香蕉)。 And here, **every** monkey is eating bananas.
Aspect	看这里，叔叔喝了一杯水。 Here, the uncle has drunk a glass of water.	那么这里呢？叔叔(**在喝**一杯水)。 And here, the uncle (is drinking a glass of water).
Relative clause	看这里，小猫在追背红书包的小女孩。 Here, the cat is chasing the girl who carries a red backpack.	那么这里呢？小猫(在追**背蓝书包的小女孩**）。 And here, the cat (is chasing **the girl who carries a blue backpack**).

The MSM subtest contains 17 items, 10 of which are expressive and 7 are receptive. Expressive items require the child to make a verbal response whereas receptive items require the child to point to the correct picture(s) from a picture array. The ESM subtest contains 17 test items, 12 of which are expressive and 5 are receptive. All the receptive and some of the expressive items are accompanied by pictorial prompts.

Procedures

The study procedures were approved by the Institution Review Board of the University of Texas. The primary caregiver of the child signed an informed consent, and was interviewed using a Mandarin-English bilingual version of the BIOS and ITALK. Children were tested individually at home, a university laboratory, or a place of their convenience. The two languages were tested on separate days, between two days and a week apart. Language of the first testing session (English, Mandarin) and subtest (morphosyntax, semantics) were counterbalanced across children. The tests were administered by native speakers of the respective languages. Test administration took between 30 to 45 minutes for each language. All picture stimuli were presented on an iPad.

The semantics subtests began with general instructions to orient the child: "*We are going to look at some pictures and I will ask you some questions about them. Are you ready? Let's begin*". Next, the examiner read the prompt question from a scoring form and presented the child with the accompanying picture, if applicable. Questions were repeated once if the child failed to respond. Generic feedback was provided such as "*You are doing a good job*" or "*You are working hard*". For receptive items, the examiner circled the child's response on the scoring form. For expressive items, the examiner either marked one of several possible responses provided on the scoring form or if the child's response was not on the scoring form, wrote down the child's response verbatim. The session was audio-recorded for reliability check. If a child code-switched, the examiner encouraged the child to answer in the target language. If a child was unable or unwilling to answer in the target language, the examiner did not persist.

The morphosyntax subtests began with the cloze task using the following instruction: "*I will show you two pictures. I will tell you something about one picture, and you tell me something about the other picture*." For each target grammatical structure, there were two practice items, followed by two to six test items. If the child did not correctly answer the practice items, the examiner provided corrective feedback and asked the child to repeat back the correct answer. Only generic feedback was provided for the test items. The sentence repetition task was presented with the following instruction: "*Listen, I'm going to say a sentence. When I am done, you copy me. Just say what I say. Don't start until I finish. Ready? Listen.*" Two practice sentences were presented to familiarize children with the task. Prompts such as a tap on the table to remind the child to repeat the sentence were used as needed. Every test sentence was read only once after the examiner confirmed that the child was paying attention.

Scoring

All items were scored as either 1 or 0. The receptive semantic items were by nature binary. The

expressive semantic items could be either binary (i.e., requiring a particular response) or open-ended. Specifically, there were several category generation items that measured children's semantic depth (e.g., *Tell me all the clothes you can think of*). For these items, children had to provide 3 correct answers in the target language to earn a score of 1. The cloze items targeted particular morphemes, phrases, or clauses and children had to provide the target answer to earn credit. For sentence repetition, children were given credit if they produced the key words or phrases regardless of word order. The total possible scores for the EMS, MMS, ESM, and MMS were respectively 48, 51, 17, and 17.

Table 4. Descriptive information of the morphosyntax subtests

EMS	No. of items	No. of items with D ≥ .3	MMS	No. of items	No. of items with D ≥ .3
3rd person singular	4	3	Passive	3	3
Auxiliary+negation	5	3	Possessive *'s*	3	0
Passive	3	0	Prepositional phrase	2	0
Past tense *–ed*	2	1	Classifier	6	0
Auxiliary + *ing*	3	2	Quantifier	5	2
Copula	3	2	Aspect	4	3
Possessive *'s*	3	3	Relative clause	4	4
Prepositional phrase	2	0	Sentence repetition*	24	16
Plural	3	1			
Question inversion	4	4			
Relative clause	4	3			
Sentence repetition*	12	6			
Total	48	28	Total	51	28

Note. EMS = English morphosyntax subtest. MMS = Mandarin morphosyntax subtest. Items are arranged in the order they appear in the BEMOS. *Both English and Mandarin sentence repetition include three sentences. In English, the length of the three sentences are respectively 10, 11, and 13 words, or 14, 15, and 17 syllables. We scored the repetition of 12 keywords from the English sentences based on the recommendations of the BESOS. In Mandarin, the length of the three sentences are respectively 10, 17, and 18 characters (syllables). We scored the repetition of 24 Mandarin words and phrases from the sentences given the exploratory nature of the study.

Reliability

The reliability of the BEMOS was evaluated using the Cronbach test in SPSS. The internal consistency of the entire test was high (145 items; α=.97). Inter-rater reliability was calculated using Cohen's kappa for each of the 4 subtests. The first round of scoring for the entire sample of 55 children was done by two native speakers of Mandarin and two native speakers of English. The two sets of coders worked collaboratively and reached consensus for their coding. A highly proficient bilingual then randomly selected 20% of the sample and independently scored all testing items for these children. The inter-rater reliability (Kappa) was 0.94 (p<.001) for MMS, 0.90 (p<.001) for MSM, 0.93 (p<.001) for EMS, and 0.90 (p<.001) for ESM. According to Landis and Koch (1977a), a Kappa statistic between 0.81 and 1 can be considered almost perfect; therefore, our inter-rater

scoring agreement was very high.

Analyses

To compare performance between groups, we conducted Welch's t-tests. Welch's t-test is an adaptation of the Student's t-test and is more reliable when the two samples have unequal variances and unequal sample sizes (Moser & Stevens, 1992; Welch, 1947). Between-group comparisons were made for each of the four BEMOS subtests. To delineate the grammatical profiles of AR children, we calculated the group means for each of the seven Mandarin grammatical forms and the Mandarin sentence repetition task and compared the mean scores between groups using Welch's t-test. Similarly, we compared the group means for each of the eleven English grammatical structures and the English sentence repetition task. Because of the small sample and the exploratory nature of the study, we reported one-tailed p values for the Welch's t-tests. Next, we conducted an item analysis to examine how each of the 99 morphosyntax items on the BEMOS MMS and EMS subtests differentiated the AR from TD group. Group means for each item was calculated, and the mean of the AR group was subtracted from that of the TD group to obtain a difference score (D). All items that had a D value of .3 or higher (Allen & Yen, 2002) were kept in subsequent discriminant functions.

Results

Table 1 presents the scores on the BEMOS subtests. Welch's t-tests showed that the two groups differed significantly in MMS [*Welch's* $t(6)=3.85$, $p<.01$], EMS [*Welch's* $t(10)=7.03$, $p<0.001$], and ESM [*Welch's* $t(4)= 2.41$, $p=0.04$], but not in MSM, $p>0.1$. Figure 1 shows the mean performance by AR and TD groups on the 11 English grammatical features and the English sentence repetition task. Of these, seven grammatical features showed group differences that were significant or approached significance: third person singular [*Welch's* $t(8)=2.28$, $p=0.03$], auxiliary negation [*Welch's* $t(7)=2.51$, $p=.02$], auxiliary –ing [*Welch's* $t(7)=1.45$, $p=.09$], possessive [*Welch's* $t(6)=2.28$, $p=.03$]; plural [*Welch's* $t(11)=2.05$, $p=.03$]; question inversion [*Welch's* $t(5)=3.58$, $p=.01$], and relative clause [*Welch's* $t(7)=4.78$, $p<.01$]. The English sentence repetition task also elicited significantly different group performance [*Welch's* $t(4)=2.88$, $p=.02$].[1]

<Figure 1 here>

Figure 1. Proportion correct on the English morphosyntax subtest as a function of group and grammatical structure. *$p<.05$; ~$p<.10$.

Figure 2 shows the group mean performance on the seven Mandarin grammatical features and the Mandarin sentence repetition task. Of these, three features showed group differences that were significant or approached significance including passive [*Welch's* $t(3)=2.14$, $p=.06$], quantifier [*Welch's* $t(10)=3.49$, $p<.01$], and relative clause [*Welch's* $t(7)=9.17$, $p<.01$]. The Mandarin sentence repetition task also elicited significantly different group performance [*Welch's* $t(8)=2.88$, $p=.01$].

[1] Note that the group differences pertained to the AR and TD status, not different age groups, because the two groups were matched on age. The two age levels were only used in calculating cutoff scores to determine if a child's score was at risk or not at risk.

<Figure 2 here>

Figure 2. Proportion correct on the Mandarin morphosyntax subtest as a function of group and grammatical structure. *p<.05; ~p<.10.

Item Analysis

Using a D value of 0.3 as the cutoff, 28 of the 48 EMS test items and 28 of the 51 MMS test items yielded considerable group differences. Detailed information on how many items under each grammatical feature achieved a D value $\geq$ 0.3 is presented in Table 4.

Discriminant Functions

To examine the classification accuracy of the English and Mandarin grammatical items that showed a D value of 0.3 or more between AR and TD groups, we entered children's total scores for the 28 English morphosyntax items and the 28 Mandarin morphosyntax items into two separate discriminant functions, respectively, for each language, in SPSS to predict the risk status of each child. Individual child's scores on the subtests of items are presented in the last two columns of Table 1 (EMS1, MMS1). As seen in Table 1, there is no overlap in performance in either English or Mandarin between groups. In English, the chi-square test was significant (Wilk's λ=.17, χ^2=16.72, canonical correlation=.91, p<.001). The total raw scores classified 92% of the cases correctly with 100% sensitivity (i.e., correct classification of children with risk) and 87.5% specificity (i.e., correct classification of TD children). One TD child was misclassified as AR[2]. In Mandarin, the chi-square test was significant (Wilk's λ=3.27, χ^2=14.00, canonical correlation=.88, p<.001). The total raw scores classified 100% of the cases correctly.

Discussion

Bilingual children with DLD are under-diagnosed and underserved for a host of reasons, one of which is the lack of foundational research and diagnostic tools. The current study, while modest in scale, is among the first investigations of the grammatical profiles of Mandarin-English bilingual children at risk for DLD. This study made two contributions to the bilingual DLD literature. First, modeled after studies of Spanish-English DLD, we showed how rating scale and test performance can be used together to identify Mandarin-English bilingual children at risk for DLD. This multi-method approach to assessment is emphasized in current conceptualization of bilingual DLD (Castilla-Earls et al., 2020). Moreover, although the current study was unable to investigate "the continuum of bilingualism" by dividing children into subgroups of language dominance, it is worth noting that studies with large samples of US Spanish-English bilingual children have shown that in children who were relatively balanced in the use of the two languages, diagnostic accuracy was higher when both languages were considered than when only one language was considered (Peña et al., 2016); in children who were functionally monolingual, adequate classification accuracy (>80%)

[2] The original discriminant function analysis yielded 100% sensitivity and 100% specificity, consistently with the observation that there is no overlap in group performance. Here we reported cross-validated discriminant function results. Cross validation is the process of testing a model on more than one sample, or in this case, subsamples. This technique is often taken to assess the reliability and generalisability of the findings.

can be achieved by assessing the dominant language only (Peña et al., 2015).

Second, we identified a number of grammatical structures that hold promise as markers of DLD in bilinguals who are learning two typologically distinct languages. While it is well established that verb tense/agreement marking morphemes are reliable clinical markers of monolingual English DLD, previous studies conducted in Canada (Blom & Paradis, 2013, 2015; Paradis et al., 2013, 2016), Singapore (Brebner et al., 2016) and the US (Jia & Fuse, 2007), showed slower acquisition of English tense- and agreement-marking morphemes in TD bilinguals who speak Chinese and/or other non-tense marking L1s. In the current study, The TD bilinguals exhibited accuracy of 66% (SD = 30%) and 56% (SD = 32%) on third person singular and regular past tense, respectively. Both values were below mastery level, replicating patterns of slow acquisition. Several other features also require tense and/or agreement marking and the TD group's scores were also below mastery level: auxiliary+negation (M=.80, SD=.19), auxiliary+ing (M=.67, SD=.36), and copula (M=.71, SD=.33). However, Welch's t-tests indicated that out of the five verb morphological markers, three were significantly different between groups (i.e., third-personal singular, auxiliary+negation, auxiliary+ing), despite large variability in both the TD and AR groups. Furthermore, item analysis indicated that certain regular past tense and copula items also led to substantial group difference. Therefore, the current results are suggestive that tense- and agreement-marking morphemes may still be promising markers of DLD, even in children whose L1s do not mark tense.

Noun-related morphemes such as plural and possessive –s tend to be less problematic in English DLD than verb-related morphemes (Bedore & Leonard, 1998). But items assessing these forms separated the bilingual children with and without risk in the current study. Although the surface forms of noun plural and possessive are identical, the TD children were noticeably more accurate in producing possessives than plurals. These findings warrant further investigations on the underlying mechanisms for the acquisition of various verb- and noun-related morphemes in bilingual children.

Among items assessing English syntactic structures, the production of relative clauses and questions that required subject-verb inversion reliably separated the AR and TD groups. These results converge with the English DLD literature (e.g., Frizelle & Fletcher, 2013; Hewitt et al., 2005; Rice et al., 2009) and the bilingual Spanish-English DLD literature (e.g., Bedore et al., 2018) in showing that problems producing sentences that involve syntactic movement or contain multiple clauses are part of the DLD phenotype across multiple groups.

Unlike previous studies of English monolinguals (Leonard et al., 2006) and Spanish-English bilinguals (Bedore et al., 2018), items probing English passive production did not differentiate the two groups. While we were clearly under-powered, the lack of separation may also be partially explained by the supportive cloze context, which provided children with not only exposure to the passive sentence frame (i.e., patient + be + verb past participle + agent) but also the subject/patient of the target response (see Table 2 for an example). Moreover, reduced passive responses without the agent (e.g., "the chair **is/was painted**") were accepted as correct, which may have further diminished group differences. Future studies may consider increasing the task difficulty by reducing similarities between prime and target sentences and changing scoring criteria. Finally, preposition

production was easy for both groups. Only two preposition items (i.e., inside, behind) were included and both had alternative acceptable responses ("in" for "inside", "next to" or "by" for "behind"). Therefore, these results should be interpreted with caution.

Of the seven Mandarin grammatical features, items assessing passive, quantifier, aspect, and relative clause resulted in adequate group separation. These findings are consistent with the monolingual Chinese DLD literature (Fletcher et al., 2005; Hao et al., 2018; He et al., 2013; Leonard et al., 2006; Yu et al., 2017; Yu et al., 2019). Three Mandarin grammatical features did not separate AR and TD groups, including the production of possessive, preposition, and classifier. The Mandarin possessive "*的 de*" is both conceptually transparent and perceptually salient (i.e., fully syllabic). Thus, it is not surprising that both groups of children appeared to have mastered this form. The Mandarin prepositional phrases presented some challenges for both AR and TD groups. The two groups also made similar kinds of errors, mostly missing *在 zai* in the "*在 + noun + location word*" phrase (roughly translated to "at + noun + location"). Prepositional phrases have not received much attention in Chinese acquisition research and warrant closer examination in future studies.

Classifier production did not separate the two groups. Both groups were performing at floor. Two types of errors predominated — substituting the targets with the generic classifier "个 ge" or with the specific classifier appearing in the prompts (e.g., saying "*三朵树*" instead of "*三棵树*"). Despite difficulties producing the specific classifiers, there were no omission errors, indicating that children understood that the classifier was obligatory in this "*numeral + classifier + noun*" syntactic context. These findings are consistent with previous studies of bilingual Chinese-English children growing up in English-speaking countries (Hao et al., 2019; Kan, 2019; Jia & Paradis, 2015; Wei & Lee, 2001), which suggest early acquisition of the syntactic constraints of classifier use but delayed acquisition or failure to master the semantics of classifiers. Chinese classifiers are both numerous and semantically opaque. The production of specific classifiers was challenging even in TD monolingual Mandarin-speaking children (Hao, 2019). Future studies may investigate the utility of classifier comprehension as an alternative in the assessment of bilingual children.

The sentence repetition task achieved successful group separation in both English and Mandarin. This finding adds to the growing evidence base of the diagnostic value of sentence repetition in various languages (Conti-Ramsden et al., 2001; Stokes et al., 2006). The sentences in the current task are either adverbial clauses involving conditional, temporal, and causal relations, or sentences that contained aspect markers, quantifiers, and specific classifiers. Carefully designed sentence repetition stimuli directly tap into children's underlying grammatical knowledge and comprise a sensitive clinical marker of DLD (Redmond et al., 2019).

Finally, discriminant function analyses showed promising diagnostic accuracy by a reduced set of the most sensitive morphosyntax stimuli. Researchers may design an expanded set of items assessing these grammatical constructions for the identification of DLD in this bilingual population. Clinicians who work with Mandarin-English bilingual children may look for errors with these grammatical features in children's spontaneous speech as potential red flags for concern.

Before closing, it would be prudent to point out some limitations of the current study. We did not

collect cognitive measures from the children (e.g., nonverbal IQ) to verify their developmental status. This shortcoming needs to be ameliorated in future studies of bilingual DLD. The present study included a group of US bilingual children who were relatively homogeneous: all children were Mandarin-L1 bilinguals (i.e., received Mandarin exposure from birth and English exposure later in life) and received relatively balanced exposure to the two languages. Thus, the current findings may have limited generalisability to other bilingual children (e.g., Mandarin-dominant, English-dominant, simultaneous bilinguals) of the same community. Also, the results may not be generalisable to Mandarin-English bilingual children who speak different vernaculars of Mandarin and English (e.g., children in Singapore) as the distributional patterns of various Mandarin and English grammatical features may be different in those vernaculars, resulting in different patterns of acquisition.

To conclude, the current study represents a modest yet important first step in delineating the grammatical profiles in both languages of Mandarin-English bilinguals at risk for DLD. With caution, we draw the following conclusions: there are many similarities in the manifestations of DLD between monolingual and bilingual learners, especially in the domain of English verb morphology and complex syntax in both English and Mandarin. Deviations from the English monolingual DLD profile may manifest as the added value of noun-related morphemes. A major deviation from the Mandarin monolingual DLD profile is the insensitivity of classifier production in detecting the disorder. More than half of the items on the experimental screener achieved considerable separation between AR and TD groups. Future larger-scale studies may build on the current findings to generate more definitive evidence regarding the grammatical profile of this bilingual group to guide the design of effective assessment and intervention.

Reference list

Allen, M. J., & Yen, W. (2002). Introduction to measurement theory. Long Grove, IL: Waveland.

Australia Bureau of Statistics. (2016). Australia language spoken at home. http://profile.id.com.au/australia/language. Accessed 30th June, 2017.

Bedore, L. M., & Leonard, L. B. (1998). Specific Language Impairment and Grammatical Morphology: A Discriminant Function Analysis. *Journal of Speech Language and Hearing Research*, *41*(4), 1185–1192. https://doi.org/1092-4388/98/4105-1185.

Bedore, L. M., Peña, E. D., Anaya, J. B., Nieto, R., Lugo-neris, M. J., & Baron, A. (2018). Understanding Disorder Within Variation: Production of English Grammatical Forms by English Language Learners. *Language, Speech, and Hearing Services in Schools*, *49*, 277–291. https://doi.org/10.1044/2017_LSHSS-17-0027

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the Catalise-2 Consortium. (2017). Phase 2 of CATALISE : a multinational and multidisciplinary Delphi consensus study of problems with language development : Terminology. *The Journal of Child Psychology and Psychiatry*, *10*, 1068–1080. https://doi.org/10.1111/jcpp.12721

Blom, E., & Paradis, J. (2013). Past tense production by English second language learners with and

without language impairment. *Journal of Speech, Language and Hearing Research*, *56*(1), 281–294. https://doi.org/10.1044/1092-4388(2012/11-0112)

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*, 953–976. https://doi.org/10.1017/S014271641300057X

Brebner, C., Mccormack, P., & Liow, S. R. (2016). Marking of verb tense in the English of preschool English – Mandarin bilingual children : evidence from language development profiles within subgroups on the Singapore English Action Picture Test. *International Journal of Language & Communication Disorders*, *51*(1), 31–43. https://doi.org/10.1111/1460-6984.12181

Castilla-Earls, A., Bedore, L., Fabiano-smith, L., Pruitt-lord, S., Restrepo, M. A., & Peña, E. D. (2020). Beyond scores: Using converging evidence to determine speech and language services eligibility for dual language learners. *American Journal of Speech-Language Pathology*, *29*, 1116–1132. https://doi.org/10.1044/2020_AJSLP-19-00179

Cooc, N. (2019). Disparities in the Enrollment and Timing of Special Education for Asian American and Pacific Islander Students. The Journal of Special Education. https://doi.org/10.1177/0022466919839029.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of child psychology and psychiatry*, *42*(6), 741-748. https://doi.org/10.1111/1469-7610.00770.

Fletcher, P., Leonard, L. B., Stokes, S. F., & Wong, A. M.-Y. (2005). The expression of aspect in Cantonese-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *48*(3), 621–634. https://doi.org/1092-4388/05/4803-0621

Frizelle, P., & Fletcher, P. (2014). Relative clause constructions in children with specific language impairment. *International Journal of Language & Communication Disorders*, *49*(2), 255–264. https://doi.org/10.1111/1460-6984.12070

Grela, B. G., & Leonard, L. B. (1997). The use of subject arguments by children with specific language impairment. *Clinical Linguistics & Phonetics*, *11*(6), 443–453. https://doi.org/10.3109/02699209708985206

Grela, B., Rashiti, L., & Soares, M. (2004). Dative prepositions in children with specific language impairment. *Applied Psycholinguistics*, *25*, 467–480. https://doi.org/10.1017.S0142716404001225

Hao, Y., Bedore, L. M., Sheng, L., & Peña, E. D. (2019). Narrative skills in two languages of Mandarin–English bilingual children. International Journal of Speech-Language Pathology, 21(4), 325–335. https://doi.org/10.1080/17549507.2018.1444092

Hao, Y., Bedore, L., Sheng, L., Zhou, P., & Zheng, L. (2021). Exploring influential factors of shape classifier comprehension and production in Mandarin-speaking children. *First Language, 41*(5), 573-604. https://doi.org/10.1177/01427237211026435

Hao, Y., Sheng, L., Zhang, Y., Jiang, F., de Villiers, J., Lee, W., & Liu, X. L. (2018). A narrative evaluation of Mandarin speaking children with language impairment. Journal of Speech, Language, and Hearing Research, 61(2), 345–359. https://doi.org/10.1044/2017_JSLHR-L-16-0367

He, X., & Yu, H. (2013). 汉语特殊型语言障碍儿童关系从句理解研究 [The grammar impairment of Mandarin Chinese SLI children: Evidence from relative clause comprehension]. 现代外语, 36(4), 340-436. https://doi.org/1003-6105(2013)04-0340-07

Hendricks, A. E., Adlof, S. M., Alonzo, C. N., Fox, A. B., & Hogan, T. P. (2019). Identifying Children at Risk for Developmental Language Disorder Using a Brief, Whole-Classroom Screen. *Journal of Speech Language and Hearing Research*, *62*, 896–908. https://doi.org/10.1044/2018_JSLHR-L-18-0093

Hesketh, A. (2006). The use of relative clauses by children with language impairment. *Clinical Linguistics & Phonetics*, *20*(7–8), 539=546. https://doi.org/10.1080/02699200500266398

Hewitt, L. E., Scheffner, C., Yont, K. M., & Tomblin, J. B. (2005). Language sampling for kindergarten children with and without SLI : Mean length of utterance, IPSYN , and NDW. *Journal of Communication Disorders*, *38*, 197–213. https://doi.org/10.1016/j.jcomdis.2004.10.002

Jia, G. (2003). The acquisition of the English plural morpheme by native Mandarin Chinese-speaking children. *Journal of Speech, Language and Hearing Research*, *46*, 1297–1311. https://doi.org/1092-4388/03/4606-1297

Jia, G., & Fuse, A. (2007). Acquisition of English gammatical morphology by native Mandarin-speaking children and adolescents. *Journal of Speech Language and Hearing Research*, *50*, 1280–1299. https://doi.org/1092-4388/07/5005-1280

Jia, R., & Paradis, J. (2015). The use of referring expressions in narratives by Mandarin heritage language children and the role of language environment factors. *Bilingualism: Language and Cognition*, *18*(4), 737–752. https://doi.org/10.1017/S1366728914000728

Jia, R., & Paradis, J. (2020). The acquisition of relative clauses by Mandarin heritage language children. *Linguistic Approaches to Bilingualism*, *10*(2), 153–183. https://doi.org/10.1075/lab.16015.jia

Kan, R. T. Y. (2019). Production of Cantonese classifiers in young heritage speakers and majority language speakers. *International Journal of Bilingualism*, *23*(6), 1531–1548. https://doi.org/10.1177/1367006918808002

Leonard, L. B. (1995). Functional Categories in the Grammars of Children With Specific Language Impairment. *Journal of Speech and Hearing Research*, *38*, 1270–1283. https://doi.org/10.1044/jshr.3806.1270.

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). MIT Press.

Leonard, L. B., Owen, A., & Elam, D. (2007). Lexical Aspect and the Use of Verb Morphology by Children With Specific Language Impairment. *Journal of Speech, Language and Hearing Research*, *50*, 759–777. https://doi.org/1092-4388/07/5003-0759

Leonard, L. B., Wong, A. M.-Y., Deevy, P., Stokes, S. F., & Fletcher, P. (2006). The production of passives by children with specific language impairment : Acquiring English or Cantonese. *Applied Psycholinguistics*, *27*, 267–299. https://doi.org/10.1017.S0142716406060280

Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., & Gillam, R. B., (2015). Utility of a language screening measure for predicting risk for language impairment in bilinguals. *American Journal of Speech-Language Pathology, 24*, 426-437. https://doi.org/10.1044/2015_AJSLP-14-0061.

Marshall, C. R., van der Lely, H. K. J. (2012). Irregular past tense forms in English : how data from children with specific language impairment contribute to models of morphology. *Morphology*, *22*, 121–141. https://doi.org/10.1007/s11525-011-9195-4

McGregor, K. K. (2020). How We Fail Children With Developmental Language Disorder. *Language, Speech, and Hearing Services in Schools*, *51*(October), 981–992. https://doi.org/10.1044/2020_LSHSS-20-00003

Moser, B. K., & Stevens, G. R. (1992). Homogeneity of variance in the two-sample means test. *The American Statistician*, *46*(1), 19–21. https://doi.org/10.2307/2684403.

Oetting, J. B., & Rice, M. L. (1993). Plural Acquisition in Children With Specific Language Impairment. *Journal of Speech and Hearing Disorders*, *36*, 1236–1248. https://doi.org/10.1044/jshr.3606.1236.

Ooi, C. C., & Wong, A. M. (2012). Assessing bilingual Chinese – English young children in Malaysia using language sample measures. *International Journal of Speech-Language Pathology*, *14*(6), 499–508. https://doi.org/10.3109/17549507.2012.712159

Paradis, J., Emmerzael, K., & Sorenson Duncan, T. (2010). Assessment of English Language Learners: Using Parent Report on First Language Development. *Journal of Communication Disorders, 4*3, 474-497. https://doi.org/10.1016/j.jcomdis.2010.01.002.

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

4388(2012/12-0050)

Paradis, J., Tulpar, Y., & Arppe, A. (2016). Chinese L1 children' s English L2 verb morphology over time : Individual variation in long-term outcomes. *Journal of Child Language*, *43*, 553–580. https://doi.org/10.1017/S0305000915000562

Peña, E. D., Bedore, L. M., Gutiérrez-Clellen, V. F., Iglesias, A., & Goldstein, B. (2008). Bilingual English Spanish Oral Screener: Experimental Version.

Peña, E. D., Bedore, L. M., Iglesias, A., Gutiérrez-Clellen, V. F., & Goldstein, B. (2010). Bilingual English Spanish Assessment-Middle Elementary (BESA-ME). Agile Testing Solutions: Austin.

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

Peña, E. D., Bedore, L. M., & Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: two languages classify better than one. *International Journal of Language and Communication Disorders, 51*(2), 192-202. https://doi.org/ 10.1111/1460-6984.12199.

Peña, E. D., Bedore, L. M., Lugo-Neris, M. J., & Albudoor, N. (2020). Identifying developmental language disorder in school age bilinguals: Semantics, grammar, and narratives. *Language Assessment Quarterly, 17*, 541-558. https://doi.org/10.1080/15434303.2020.1827258.

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

Redmond, S. M., Ash, A. C., Christopulos, T. T., & Pfaff, T. (2019). Diagnostic accuracy of sentence recall and past tense measures for identifying children's language impairments. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2438-2454. https://doi.org/10.1044/2019_JSLHR-L-18-0388.

Rice, M. L., Hoffman, L., & Wexler, K. (2009). Judgments of Omitted BE and DO in Questions as Extended Finiteness Impairment (SLI) to 15 Years : *Journal of Speech, Language, and Hearing Research*, *52*, 1417–1433. https://doi.org/1092-4388/09/5206-1417

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., & Blossom, M. (2010). Mean Length of Utterance Levels in 6-month Intervals for Children 3 to 9 Years with and without Language Impairments. *Journal of Speech Language and Hearing Research*, *53*(2), 333–349. https://doi.org/10.1044/1092-4388(2009/08-0183)

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*, 1239–

1257. https://doi.org/10.1044/jshr.3906.1239.

Sheng, L., Yu, J., Su, P.L., Wang, D., Lu, T.-H., Shen, L., Hao, Y., & Lam, B.P.W. (under review). Developmental language disorder in Chinese children: A systematic review of research.

Statistics Canada (2016). Language spoken at home. https://www12.statcan.gc.ca/census recensement/2016/dp-pd/index-eng.cfm Accessed 30th June, 2017.

Stokes, S. F., & So, L. K. H. (1997). Classifier use by language-disordered and age-matched Cantonese-speaking children. *Asia Pacific Journal of Speech, Language and Hearing*, *2*(2), 83–101. https://doi.org/10.1179/136132897805577413

Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword Repetition and Sentence Repetition as Clinical Markers of Specific Language Impairment: The Case of Cantonese. *Journal of Speech Language and Hearing Research*, *49*(2), 219–236. https://doi.org/10.1044/1092-4388(2006/019)

Wang, D., Zheng, L., Lin, Y., Zhang, Y., & Sheng, L. (2022). Sentence repetition as a clinical marker for Mandarin-speaking preschoolers with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 65*, 1543-1560. https://pubs.asha.org/doi/10.1044/2021_JSLHR-21-00401.

Wei, L., & Lee, S. (2001). L1 Development in an L2 Environment : The Use of Cantonese Classifiers and Quantifiers by Young British-born Chinese in Tyneside. *International Journal of Bilingual Education and Bilingualism*, *4*(6), 359–382. https://doi.org/10.1080/13670050108667738

Welch, B. L. (1947). The Generalization of `Student's' Problem when Several Different Population Variances are Involved. *Biometrika*, *34*(1/2), 28–35. https://doi.org/10.2307/2332510.

Yu, H., He, X. & Wang, H. (2017). 普通话特殊型语言障碍儿童关系从句产出研究 [The production of relative clauses by Mandarin preschool children with specific language impairment]. 现代外语, 40(4), 495-584. https://doi.org/1003-6105(2017)04-0495-012

Yu, J., Wang, L., & Liang, D. (2019). 学龄期普通话特异性语言损伤儿童体习得和语法体组合加工的证据 [The acquisition of aspect in Mandarin-speaking school age children with SLI—Evidence from processing the combination of of lexical and grammatical aspects]. 心理发展与教育, 35(2), 203-209. https://doi.org/10.16187/j.cnki.issn1001-4918.2019.02.09

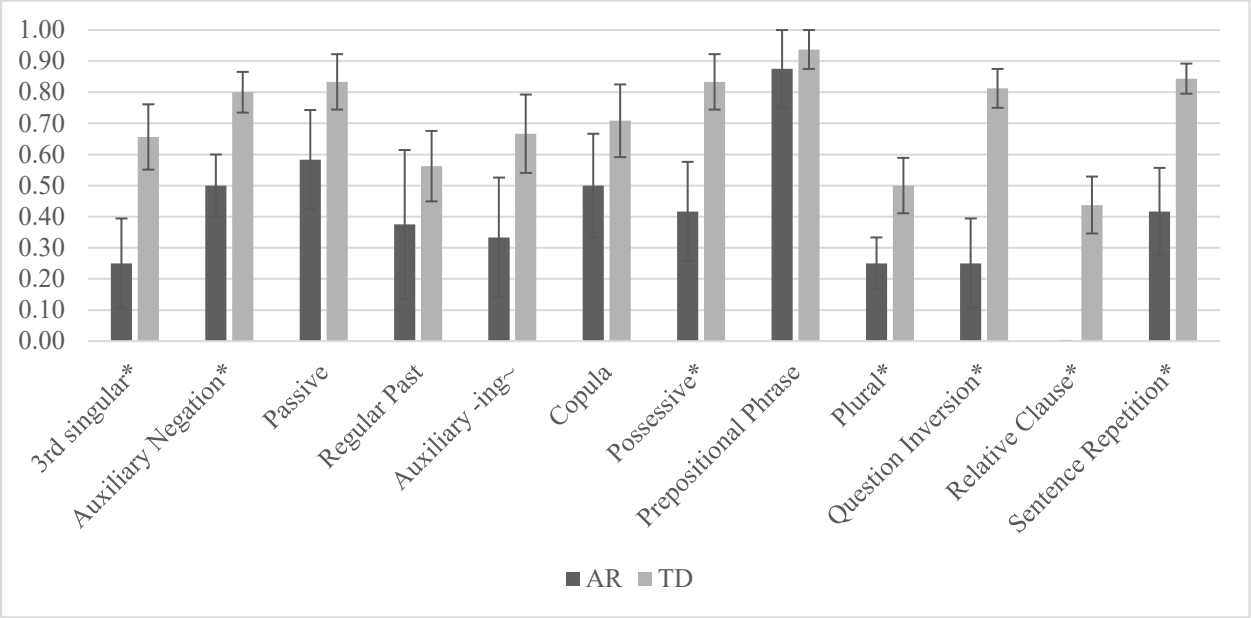

Figure 1. Proportion correct on the English morphosyntax subtest as a function of group and grammatical structure. *p<.05; ~p<.10.

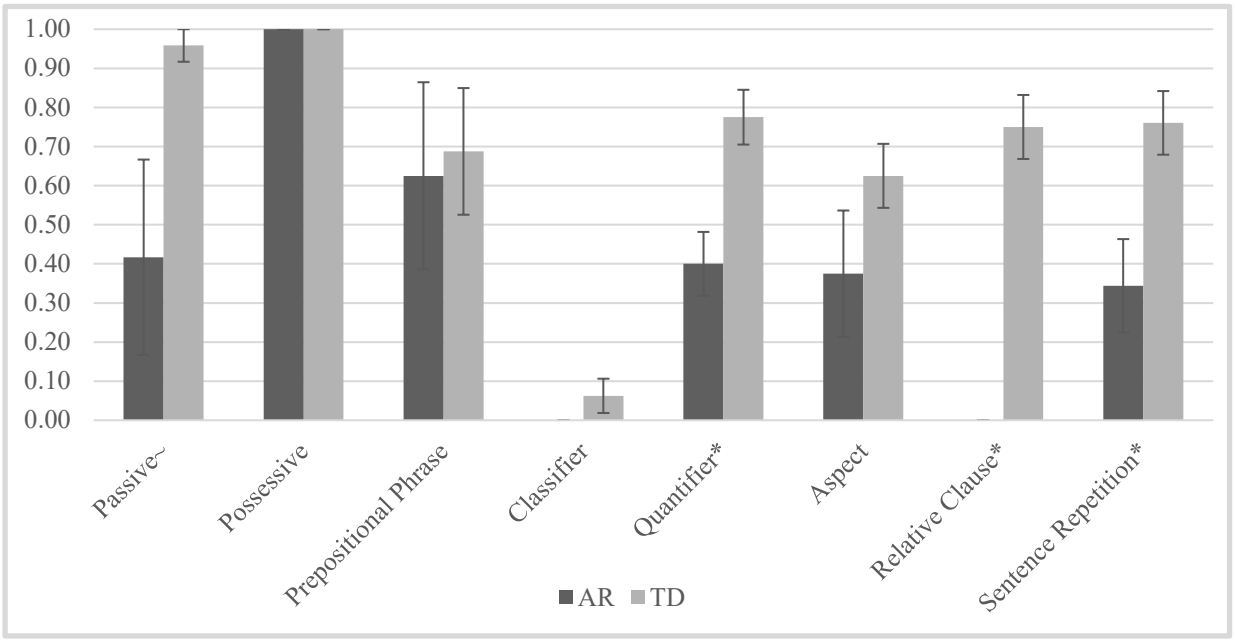

Figure 2. Proportion correct on the Mandarin morphosyntax subtest as a function of group and grammatical structure. *p<.05; ~p<.10.