Original Research

SAGE Open
January-March 2024: 1–12
© The Author(s) 2024
DOI: 10.1177/21582440241239466
journals.sagepub.com/home/sgo
Sage

Indicators for Building a Norm-Referenced Dataset Based on Conversational Language Samples in Typical Mandarin-speaking Children: An Exploratory Study

Yibin Zhang[1], Brian MacWhinney[2], and Jing Zhou[1]

Abstract

For proper assessment, diagnosis, and treatment of language disorders in Mandarin children, it is important to have measures that closely track the course of normal development. The current study uses a large collection of spontaneous conversational language samples to track the developmental course of five language measures: mean length of utterance (MLU), the MLU of the five longest utterances (MLU5), vocD, number of repetitions, and number of retracings. We used cross-validation-based linear regression to estimate the relationship between age, gender and each of the five variables derived from the conversational language samples in 101 typically developing Mandarin-speaking children aged 3 to 7. Each of the five measures showed significant age-related effects during the period from age 3 to age 7. As norm-referenced measures of language development in children speaking Mandarin, these developmental data could inform clinical therapists regarding assessments and interventions for children with language disorder or impairment.

Keywords

Mandarin-speaking children, syntactic complexity, lexical diversity, retracing, repetition, LSA

Introduction

According to the latest survey in 2023, the proportion of children with developmental language disorders in China exceeded 8.5% (S. Wu et al., 2023). Beginning in the 1970s, Chinese scholars have conducted a series of studies of language disorders (see meta-analysis, Y. B. Zhang et al., 2016). However, these studies have not yet produced the type of norm-referenced data that is needed for accurate clinical diagnosis. Without a clear normal comparison database across the full course of language development in childhood, we cannot properly assess, diagnose, and treat language disorders.

Language sample analysis (LSA) is regarded by clinicians as the gold standard procedure for evaluating expressive language ability (Bloom & Lahey, 1978; Brown, 1973; Evans, 1996; Miller et al., 2016; Oh et al., 2020). Spontaneous speech is representative of child language development and can be an effective measure in screening for children at risk of language delay (Guo et al., 2018; MacWhinney & Ratner, 2016; Paul & Norbury, 2012; Yim et al., 2015). Language samples may be elicited from speakers of any age; they are sensitive to change; they minimize cultural bias; and they can be repeated frequently (Heilmann & Westerveld, 2013). In the past years, many indicators based on the analysis of language samples have been produced. For vocabulary analysis indicators, the most common ones are TTR (type/token ratio) (Templin, 1957), NDW (number of different words)(Klee, 1992; Watkins et al., 1995; Wong et al., 2010), and vocD (vocabulary diversity) (Malvern & Richards, 2002). For syntactic analysis, the most common measures are Mean Length of Utterance (MLU)

[1]East China Normal University, Shanghai, China
[2]Carnegie Mellon University, Pittsburgh, PA, USA

Corresponding Author:
Yibin Zhang, Institute of Brain and Education Innovation, East China Normal University, Room 511, Tianjiabing Building, No. 3663, North Zhongshan Road, Shanghai 200062, China
Email: zyb_1989@outlook.com

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

(Blake et al., 1993; Brown, 1973; Rice et al., 2010; Rondal et al., 1987) and longest utterance (Devescovi et al., 2005; Eisenberg et al., 2001; Strömqvist & Verhoeven, 2004). Language samples have also been analyzed by profiling systems, such as LARSP (Crystal et al., 1989), DSS (Lee et al., 1974), and IPSyn (Scarborough, 1990).

As collecting and analyzing language samples takes a lot of time (Pavelko et al., 2016), computer programs have helped standardize the LSA process. Examples include SALT (Miller & Iglesias, 2015), CLAN (MacWhinney, 2000) and SUGAR (Pavelko & Owens, 2017). LSA programs constantly update standards and analysis methods. LSA with SUGAR (Pavelko & Owens, 2017) may be completed in about 20 min, which is even less than the average time for a standardized, normative language assessment, although there can be problems with this level of analysis (Guo et al., 2018). Although both SALT (Miller & Iglesias, 2015) and the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000) are now providing reference corpora for English, there is not yet a systematically constructed normal comparison corpus for child Mandarin.

The current study uses spontaneous language samples from CHILDES to build a database specifically for Mandarin Chinese. CHILDES data from English-speaking children have been configured for rapid LSA (MacWhinney & Ratner, 2016) using the KidEval command within the larger set of CHILDES programs. The numerical results of KidEval analysis for a given child can point to specific weaknesses in vocabulary, fluency, or grammar within a few minutes, saving a lot of time for further clinical practice. For English-speaking children, the LSA measures computed by KidEval are effective for diagnosing language disorders. Looking at data for children aged 2 to 7 years, MacWhinney and Ratner (2016) found that both MLU and IPSyn (MacWhinney et al., 2020; Scarborough, 1990) were successful in diagnosing language disorders. The same methods that are being used for analyzing English-speaking children's data with KidEval can also be used for Mandarin by comparing samples with the large body of Mandarin data in CHILDES. In this study, we analyzed several core indicators representing Mandarin-speaking children's language development through LSA to explore the value of these indicators for the diagnosis and intervention of children with clinical language problems.

Standardized Measures in Mainland China

Standardized assessment of children's language disorders refers to the establishment of reliability, validity, and normative data through formal testing procedures. Measures must be tested according to standard procedures and must demonstrate validity and reliability (S. L. Yang, 2015). For English, there are many standardized measures for clinical diagnosis of childhood language disorder (see review, Wallace et al., 2015). In contrast, valid and reliable methods for Mandarin, based on LSA, are completely lacking. According to Y. B. Zhang et al. (2016), therapists in China are using Gesell Developmental Schedules, Bayley Scales of Infant Development, 0~6-year-old Pediatric Examination Table of Neuropsychological Development, Sign-Significant Relations-Chinese version, and WPPSI for cognitive assessment. For language disorder assessment, clinicians are using tests that emphasize lexical abilities such as PPVT-R (Chinese Peabody Picture Vocabulary Test-Revised, Lu & Liu, 1998) and EVT-R (Chinese Expressive Vocabulary Test, Williams, 1997). One newly developed measure, DREAM (X. L. Liu et al., 2017) examines receptive and expressive language ability. A second recently developed measure, MCELP-CS (H. Wu et al., 2020), covers vocabulary comprehension (VC), sentence comprehension (SC), vocabulary naming (VN), sentence structure imitation (SSI), and story narration (SN). Although these two measures have good reported validity, they are expensive, which seriously affects their accessibility. Most importantly, none of the above tools has established an effective and public norm database in China, and none are based on LSA.

A meta-analysis of language disorder studies for Chinese has shown that different measures motivate very different intervention effects/outcomes (Y. B. Zhang et al., 2016). Furthermore, tests conducted based on PPVT-R and EVT-R have produced the controversial result that receptive language skills are weaker than expressive language skills for Mandarin-speaking children (Zhou et al., 2014, 2017). In addition, children may be unfamiliar with test formats and anxious about the testing situation in ways that will affect results. Moreover, because the forms of standardized test form are often limited to tasks such as picture naming, sentence repetition, or sentence completion, they can fail to provide a full picture of children's expressive language ability (Gai et al., 2009). As noted earlier, the high price of language assessment instruments can make them inaccessible to many sites in China.

Utility of LSA in Mainland China: A Short Review

When used in conjunction with these standardized tests, LSA can provide better ecological validity and a fuller measure of a child's functional language usage. The growth of LSA in China has relied in large part on the introduction of CHILDES methods in the 1990s. The first study using CHILDES corpus methods (Zhou, 2001) focused on pragmatic development of children

Table 1. MLU and MLU5 Values of Different Studies on Mandarin-Speaking Children.

Author(s)	Context	Age	36 m	42 m	48 m	54 m	60m	66 m	72 m
Jin and Jin (2008)	Free play	MLU	3.8	3.9	3.99	4.07	4.38	4.5	4.55
Cai (2008)	Free play	MLU	3.99	/	4.29	/	/	/	/
L. Li (2014)	Narrative	MLU	/	/	5.06	5.88	6.14	6.57	6.72
		MLU5	/	/	5.8	6.52	8.64	9.4	9.96

aged 0 to 3. This study reversed the earlier tradition of child language research which relied on speculative reasoning using single sample sentences. Since that time, investigators have contributed over a dozen new Mandarin child language corpora to CHILDES. Researchers are using CLAN, the free software that comes with the CHILDES system, to automatically calculate various measures of vocabulary and grammar for children with specific language impairment (F. Zhang, 2013), children with autism (X. Y. Li, 2008), children with hearing impairment (He & He, 2009), children with mental retardation (Chen, 2007), as well as typically developing children (H. M. Li, 2003; L. H. Li, 2008; S. Y. Liu, 2018; Ng, 2017; Niu, 2018; Ouyang, 2003; X. L. Yang, 2018; S. Y. Yu, 2017).

Although the CHILDES Mandarin database includes 19 corpora collected from hundreds of children, the studies did not implement a standard elicitation method or recording of demographic information. In order to establish a database for normative validation, it will be necessary to implement elicitation standards and systematic demographic recording.

Context and Measures Used in LSA for Mandarin-Speaking Children

When collecting language samples, it is important to realize that children's productions may be markedly affected by the sampling context (Eisenberg et al., 2001, 2018; Klein et al., 2010; Nippold et al., 2014; Rice et al., 2010; Sealey & Gilmore, 2008; Southwood & Russell, 2004). This means that a language sample taken from a child in a given context should be compared to a collection of samples from other children taken in the same context. The most common contexts for sampling include toy play (or free play), narratives, story-telling, and dinner talk. At present, the majority of samples are taken from parent-child free play (Chen, 2007; He & He, 2009; Ng, 2017; Ouyang, 2003; X. Y. Li, 2008; X. L. Yang, 2018). Less common are samples from narrative (Niu, 2018; S. Y. Yu, 2017; F. Zhang, 2013), storytelling (L. H. Li, 2011; S. Y. Liu, 2018), and dinner talk (H. M. Li, 2003). Despite the preponderance of data from free play, publications of the utility of LSA have so far focused on samples generated from narration. For example, Hao et al. (2018) compared the narrative abilities of children with language disorder and their peers and found that children with language impairment had lower lexical diversity (NDW) and fewer complex sentences. The only study published with samples elicited in free play with Chinese children focused on Cantonese-speaking children. Wong et al. (2010) found that normally developing children produced more complete and intelligible utterances and higher lexical diversity (NDW) than SLI children.

Prior work has demonstrated that lexical diversity, MLU, and MLU5 (the MLU of the five longest utterances) are good metrics for distinguishing SLI from typically developing children (Cai, 2008; Cheung, 1998; Jin & Jin, 2008; Weng, 2016; X. Y. Li, 2008; L. Yu et al., 2017; L. Zhang & Zhou, 2009; Zhou & Zhang, 2020; Zhu, 1986). For lexical diversity, people are using NDW, vocD and TTR. NDW can be used to differentiate developmental language disorders from typically-developing controls (Sheng et al., 2020). Jin and Jin (2008) used vocD to explore the relationship between lexical diversity and age, and they found that children's vocabulary diversity increased with age before 66 months, while the speed of increasing decreased after 66 months, indicating the language development may be dominated by grammatical development. VocD can also be used to distinguish typically-developing children from children with developmental delays (Cai, 2008). However, Wong et al. (2010) found that the combination of vocD and MLU cannot distinguish Cantonese-speaking SLI from typically-developing children. Therefore, the use of vocD in the Mandarin context still needs further discussion. Weng (2016) also analyzed the TTR of children in small, middle and large classes and found the same negative results.

Several studies have used CLAN to calculate MLU for Chinese children (Table 1). These studies show that the MLU of Chinese children increases gradually with age. In comparison with overall MLU, L. Zhang and Zhou (2009) and L. Li (2014) found that MLU5 was more able to reflect the grammatical level of Chinese children after four or four and a half years old. Niu (2018) also found that compound sentences appeared more and more frequently in children's five longest sentences, reflecting the development trend of children's grammatical structure from simple sentences to compound sentences, especially at the age of 4 to 5. Although many studies have

examined the development of MLU or MLU5 in Mandarin-speaking children, few have concentrated on children with language disorders, especially in a clinical context. Sheng et al. (2020) found that there was no significance between children with developmental language disorders and their matched typically developing controls on MLU in narrative. Therefore, the exploration of the development of MLU will provide important evidence to further clarify the clinical significance of this indicator.

Disfluencies such as repetitions and retracings have also been considered as important indicators of developmental status for typical Mandarin-speaking children (Liang, 2018; Tseng, 2006). Often, when they detect a mismatch with a stored form, children will repair or retrace words or even whole phrases (MacWhinney & Osser, 1977). These can be calculated automatically by CLAN's KidEval and FluCalc programs. Liang (2018) analyzed the narrative corpus of 40 6-7-year-old children using *The Frog Story*. She found that pragmatics played an important role in promoting repairing. Through the interaction with other people, children continuously modified and adjusted themselves to adapt to their own language needs and external communication needs. However, both studies did not offer any quantitative data for further analysis.

This review has identified four gaps in previous research. First, different language sample analysis indicators have not been explored and analyzed in a larger and more integrated framework. Second, the LSA method has not been used for developing norms for Mandarin development (see Sheng et al., 2020). Third, despite this, the research has not been extended to explore the applicability of language sample indicators, or whether the language sample indicators can reflect the level or gradient of language development of Mandarin-speaking children, and as most of the corpora collected are in state of sharing, no data is used to establish a normative reference database. Although lot of corpora has been pushed on CHILDES with more children included, it is hard to find the information on socioeconomic status. As a seemingly old but innovative research, this research uses some language sample analysis indicators commonly used in English, combined with some indicators in exploring the language development of Mandarin speaking children, and then try to explore the value of these indicators for measuring the language development of Mandarin-speaking children and for constructing a diagnostic system with normative meaning. The exploration of the indicator system will provide important support for the subsequent construction of a norm reference dataset, which can be used for preliminary diagnosis of children with clinically delayed or impaired language development, thus providing a diagnostic path based on "language in action."

Present Study

This study aims to explore the developmental trajectory of the LSA measures of MLU, MLU5, vocD, number of retracings, and number of repetitions. These data can provide important guidance for further measurement of children with possible language impairment by delineating the upper and lower limits of development for children across different ages. For example, MLU and lexical diversity obtained from conversational language samples can reveal the extent to which they can represent the grammar and vocabulary level of Chinese children. From the perspective of developmental research, the above data will further enhance our understanding of LSA on Mandarin children.

The study's second objective is to explore gender differences between different types of indicators. Few studies have examined gender differences in the language ability of Chinese children (Cao et al., 2018). We do not know whether there are gender differences in spoken language among children with/without language disorders. This exploration will further strengthen the feasibility of applying LSA in clinical practice for children of both genders.

Method

Participants

Participants were drawn from the control group used in a larger study of children intellectual disabilities, hearing loss, autism and specific language impairment. These data are available in the ZhouAssessment corpus at https://childes.talkbank.org. Public access to the anaonymized data was approved by the Ethics Committee of University.

We recruited 112 normally-developing children (60 boys and 52 girls) from preschools in Zibo, Shandong Province, China. The final sample consisted of 101 typically developing children (54 boys and 47 girls), because 11 families did not finish the language background survey. The children were 34 to 72 months of age (M = 51.34, SD = 12.51), and the sample was 47% female and 53% male. All children spoke Putonghua (Mandarin Chinese) at preschool, and 99 spoke Putonghua at home, while only 2 spoke Shandong dialect at home. In regard to maternal education, 7% had completed less than high school, 10% had completed high school, 3% had completed some college, 66% had graduated from college, and 14% had obtained a graduate-level or professional degree, indicating a medium to high SES compared to other studies (e.g., Y. Zhang et al., 2021). We required teachers to confirm that the children's performance was similar to those of their peers and that they had never received special services.

Material and Procedure

Different approaches for eliciting language samples have been used by researchers in China, including spontaneous free play, peer talk and storytelling. When attempting to assess the language of very young children, particularly those with language disorders or delay under the age of 5 (Pezold et al., 2020), spontaneous free play has the advantage of making the child comfortable enough to maximize language production. In this study, we also used spontaneous child free play to obtain language samples, because this allows teachers to elicit language samples most easily. The elicitation procedure followed a tightly scripted protocol. We limited the entire game process to 20 min. Even if children were still immersed in the game in most cases, we still chose to stop the game in time. In order to ensure the continuity and fun of the game, we chose the dollhouse and dolls that children prefer, and placed them on the table in the play center of the kindergarten. The teacher and the child would play kitchen games and games in which guests would come to visit. Materials for promoting and extending the conversation included a big toy house with toy furniture, such as desks, chairs, beds, cooking devices, and toy pigs and cats frequently shown in cartoons.

The teacher would bring the child to a play center in the preschool familiar to children and would start the procedure by asking *shall we start playing the game*? They then started recording the conversation, but only if the child gave an answer of *yes*. After 20 min, the play session was ended by teacher's question, *shall we take a break?* The play sessions were recorded via audio-recording devices, such as cellphones.

It is difficult to fully control what teachers say in dialog with different children. We decided that it was best to allow the game process to be maximally natural to increase ecological validity. Therefore, we did not design problems when teachers need to talk to children. In order to make the data of this research more representative and reliable, we still investigated the basic information of teachers. The 6 teachers participating in this study all have bachelor's degree, and they all have junior teacher qualification certificates.

Reliability

The language samples were transcribed by two graduate students and rechecked by the author. The two graduate students participated in a one-week-training workshop on the word and utterance segmentation guidelines held by the second author, the coordinator of CHILDES. A graduate student listened to the audio recordings while reading the transcript and entering queries and corrections directly onto the transcripts. All discrepancies were discussed and resolved by the graduate students, and 10% of each set of transcripts were double-transcribed, yielding high reliability, Cohen's Kappa = 0.85 for all transcripts. Where they could not agree, the words or utterances were indicated as unintelligible.

Transcription and LSA

During transcription, everything the child said was written down, including pauses, repetitions and retracings. Conversational turns were divided into utterances based on the rules of intonation patterns and the Chinese grammar. All transcripts were coded in CHAT (Codes for the Human Analysis of Transcripts) format, a standardized format required for further LSA with CLAN (MacWhinney, 2000). Unintelligible utterances or single-word responses to teacher's questions were excluded in this study. All children produced 50 utterances or more (see Guo & Eisenberg, 2015), and in total, 9,337 utterances were transcribed and store and coded digitally.

Segmentation into words was conducted based on a Java program developed by CHILDES systems (https://talkbank.org/morgrams/zhoseg.zip) that relied on commonly accepted segmentations from the CEDICT dictionary and morphemes Lexicon in CLAN. The Java program would first check morphemes Lexicon in CLAN, where children's speech morphemes were stored and segmented based on Mandarin rules proposed by Cheung (1998) and Zhou (2001). For morphemes not found in Lexicon, the java program would automatically segment them refer to the CEDICT dictionary. If morphemes could be found in either Lexicon or CEDICT dictionary, a file would be outputted, and the authors would segment add them into Lexicon by hand based on Cheung (1998) and Zhou (2001).

After segmentation, all CHAT files were supplemented with a %mor tier by running the MOR command in CLAN. The %mor tier is the basis for morphemes analysis as it labels the part of speech for each morpheme. Then, all of the language indices were calculated automatically using CLAN. In this study, we focused analysis on the three dimensions of syntactic complexity, lexical diversity, and fluency.

Syntactic complexity. We used the mean length of utterance (MLU) as the index of syntactic complexity. MLU is sensitive to age level in younger children, but this sensitivity decreases with age. In addition, we utilized MLU5 an indicator verified in previous Mandarin child language research (e.g., L. Zhang & Zhou, 2009).

Lexical diversity. We used vocD as the measure of lexical diversity. VocD was widely used in the study of Chinese typically-developing children and children with language disorders (Cai, 2008; Jin & Jin, 2008; Wong et al., 2010). Higher values of vocD reflect greater lexical diversity in children's expressive language.

Fluency. We used repetition and retracing as our two measures of disfluency. Repetition is a simple repetition of certain words, such as

*CHI: <炒 鸡蛋> [/] 炒 鸡蛋 已经 做 好 了.
[*CHI: <fried eggs> [/] fried eggs have already been done.]

Retracing is a vocabulary modification or syntactic modification with a correct meaning, for example,

*CHI: <放 这边>[//] 把 椅子 放 这边.
[*CHI: < put this side> [//] put the chair here.]

Although there are studies that have systematically analyzed different types of disfluencies (e.g., Liang, 2018), none of these studies have explored these features developmentally.

Data Analysis

In order to increase the reliability of data analysis, 101 children were randomly divided into three groups using the Excel RANK function, with 34, 34, and 33 children in each group respectively, and a cross-validation-based linear regression was conducted for each group (Group1, Group 2, and Group 3) independently. The independent variables were age and gender; the dependent variables were MLU, MLU5, vocD, repetition, and retracing. In total, 15 separate analyses (3 groups × 5 dependent variables) were conducted.

Results. We used cross-validation-based linear regression to estimate the relationship between age, gender and each of the five variables derived from the conversational free play samples. See Table 2 for the descriptive statistics for these variables and Table 3 for linear analysis on five variables.

Table 2. Language Variables.

Variables	Group	*N*	Sex	Mean	*SD*
MLU	Group 1	12	Boy	3.704	1.026
		22	Girl	4.054	0.960
	Group 2	21	Boy	3.565	0.890
		13	Girl	3.822	1.181
	Group 3	21	Boy	4.032	1.327
		12	Girl	4.500	1.545
MLU5	Group 1	12	Boy	9.417	2.401
		22	Girl	10.855	3.300
	Group 2	21	Boy	9.514	2.366
		13	Girl	12.231	4.401
	Group 3	21	Boy	11.352	3.853
		12	Girl	10.917	2.067
NDW	Group 1	12	Boy	54.250	7.047
		22	Girl	56.318	5.541
	Group 2	21	Boy	54.762	6.833
		13	Girl	55.385	6.640
	Group 3	21	Boy	58.286	4.161
		12	Girl	55.333	6.513
VocD	Group 1	12	Boy	58.758	15.493
		22	Girl	62.232	16.934
	Group 2	21	Boy	56.042	17.836
		13	Girl	54.255	14.398
	Group 3	21	Boy	63.894	16.112
		12	Girl	56.665	13.397
Retracing	Group 1	12	Boy	0.167	0.389
		22	Girl	0.727	1.695
	Group 2	21	Boy	0.429	0.676
		13	Girl	0.385	1.121
	Group 3	21	Boy	0.810	1.401
		12	Girl	0.417	0.900
Repetition	Group 1	12	Boy	3.917	3.965
		22	Girl	3.182	4.316
	Group 2	21	Boy	2.048	3.122
		13	Girl	2.538	4.943
	Group 3	21	Boy	2.524	3.108
		12	Girl	4.417	7.090

Syntactic Complexity. The linear regression showed that age significantly predicted the size of MLU for each group [Group1: $\beta = 0.668$, $t(33) = 5.102$, $p = .000$; Group2: $\beta = 0.583$, $t(33) = 4.034$, $p = .000$; Group3: $\beta = 0.654$, $t(32) = 4.791$, $p = .000$](Figure 1), indicating a linear trajectory with age. Age and gender explained 47.3%, 35.5% and 44.8% of the variance in the size of MLU [Group 1: $F(2,31) = 13.890$, $p = .000$; Group 2: $F(2,31) = 8.522$, $p = .001$; Group 3: $F(2,30) = 12.196$, $p = .000$] in each group respectively, though gender was not a significant predictor.

The linear regression showed that age significantly predicted the size of MLU5 for group one and group three [Group1: $\beta = 0.588$, $t(33) = 4.189$, $p = .000$ and Group3: $\beta = 0.718$, $t(32) = 5.587$, $p = .000$] (Figure 2), indicating a linear trajectory with age. Age and gender explained 39.5% and 51.2% of the variance in the size of MLU5 [$F(2,31) = 10.108$, $p = .000$ and $F(2,30) = 15.735$, $p = .000$, respectively]. Meanwhile, age and gender significantly predicted the size of MLU5 [Group1: $\beta = 0.451$, $t(33) = 3.106$, $p = .004$; Group3: $\beta = 0.355$, $t(33) = 2.449$, $p = .02$], and the model explained 34.9% of the variance in the size of MLU5 [$F(2,31) = 8.319$, $p = .001$]. However, age and gender do not contribute to the variance in the size of MLU 5 in group two.

Lexical Diversity. The linear regression showed that age significantly predicted the size of vocD for group two and group three [$\beta = 0.397$, $t(33) = 2.404$, $p = .022$ and $\beta = 0.366$, $t(32) = 2.207$, $p = .035$, respectively] (Figure 3), indicating a linear trajectory with age. Age and gender explained 16% and 18.5% of the variance in the size of vocD [$F(2,31) = 2.944$, $p = .068$ and $F(2,30) = 3.405$, $p = .046$, respectively]. Meanwhile, age does not contribute to group one, and gender does not contribute to any group.

Table 3. Linear analysis on five variables.

Indicators		Variable	Group 1 [*N* = 34]			Group 2 [*N* = 34]			Group 3 [*N* = 33]		
			B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Syntactic complexity	MLU	Age (Month)	0.052	0.010	0.668	0.046	0.011	0.583	0.075	0.016	0.654
		Sex	0.227	0.266	0.112	0.185	0.294	0.091	0.247	0.393	0.086
		R^2		0.473			0.355			0.448	
		F		13.890***			8.522**			12.196***	
	MLU5	Age (Month)	0.142	0.034	0.588	0.125	0.040	0.451	0.192	0.034	0.718
		Sex	1.102	0.884	0.175	2.520	1.029	0.355	−1.003	0.864	−0.149
		R^2		0.395			0.349			0.512	
		F		10.108***			8.319**			15.735***	
Lexical diversity	VocD	Age (Month)	0.334	0.223	0.260	0.514	0.214	0.397	0.458	0.207	0.366
		Sex	2.681	5.816	0.080	−2.596	5.484	−0.078	−8.585	5.224	−0.273
		R^2		0.078			0.160			0.185	
		F		1.307			2.944			3.405*	
Disfluencies	Repetition	Age (Month)	−0.153	0.052	−0.468	−0.135	0.049	−0.442	−0.125	0.068	−0.313
		Sex	−0.371	1.359	−0.043	0.703	1.259	0.090	2.264	1.724	0.225
		R^2		0.225			0.198			0.132	
		F		4.495*			3.837*			2.285	
	Retracing	Age (Month)	0.040	0.018	0.367	0.004	0.012	0.064	0.030	0.017	0.296
		Sex	0.465	0.473	0.161	−.051	0.312	−0.029	−0.481	0.441	−0.189
		R^2		0.171			0.005			0.110	
		F		3.199			0.073			1.863	

*p < 0.05, **p < 0.01, ***p < 0.001.

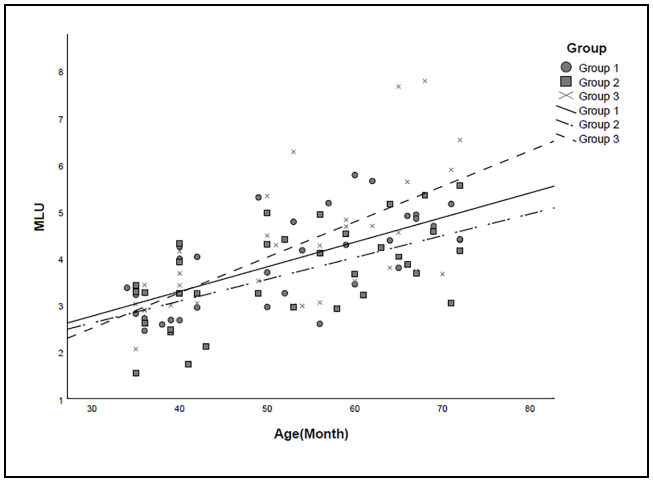

Figure 1. Cross-sectional trajectory of syntactic complexity (mean length of utterance [MLU]) across age.

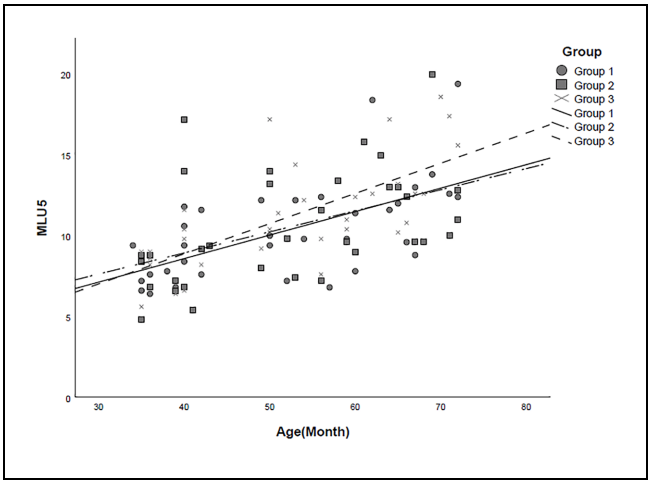

Figure 2. Cross-sectional trajectory of syntactic complexity (mean length of utterance [MLU5]) across age.

Disfluencies. There was an inverse relationship between age and repetition whereas there was a positive relationship between age and retracing. In other words, children retrace more frequently as they get older, whereas they do less repetition.

The linear regression showed that age significantly predicted the size of retracing only for group one [β = 0.367, $t(33) = 2.232$, $p = .033$] (Figure 4), indicating a linear trajectory with age. Age and gender explained 17.1% of the variance in the size of retracing [F (2,31) = 3.199, $p = .055$]. Meanwhile, age and gender do not contribute to the other two groups.

The linear regression showed that for age significantly predicted the size of repetition for group one and group two [β = −0.468, $t(33) = -2.949$, $p = .006$ and β = −0.422, $t(33) = -2.743$, $p = .010$, respectively] (Figure 5). Age explained 22.5% and 19.8% of the variance in the size of repetition [F (2,31) = 4.495, $p = .019$ and F (2,31) = 3.847, $p = .032$, respectively]. Meanwhile, age and gender do not predict the size of repetition in

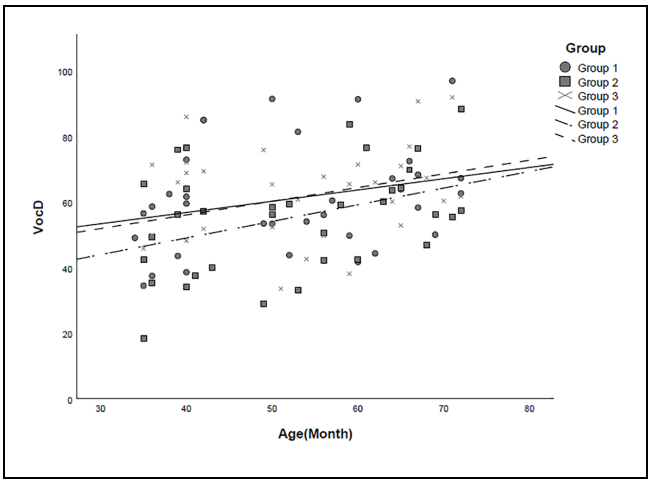

Figure 3. Cross-sectional trajectory of lexical diversity (vocD) across age.

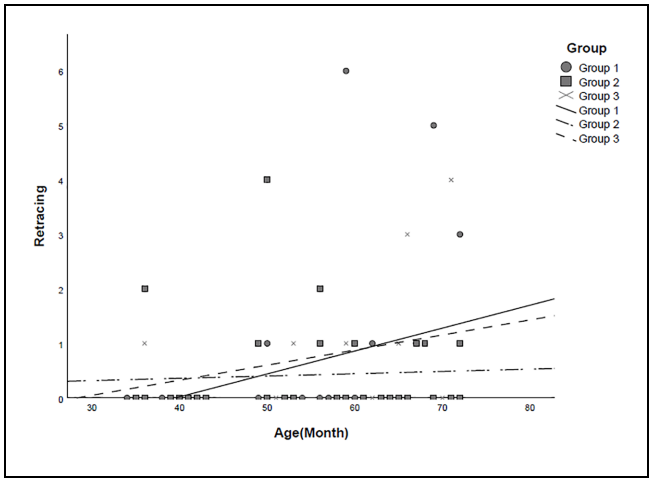

Figure 4. Cross-sectional trajectory of disfluencies (Retracings) across age.

group three, and gender do not contribute other two groups either.

Discussion

This study examined the developmental patterns for measures based on spontaneous conversational language samples of Mandarin-speaking children. Based on existing language research on Mandarin-speaking children, we found and explored the developmental trend of five indicators: syntactic complexity: MLU, MLU5, vocD, retracing, and repetition.

Syntactic complexity, as measured by MLU and MLU5, increased significantly with age. Although there are inconsistencies in the regression analysis on MLU5

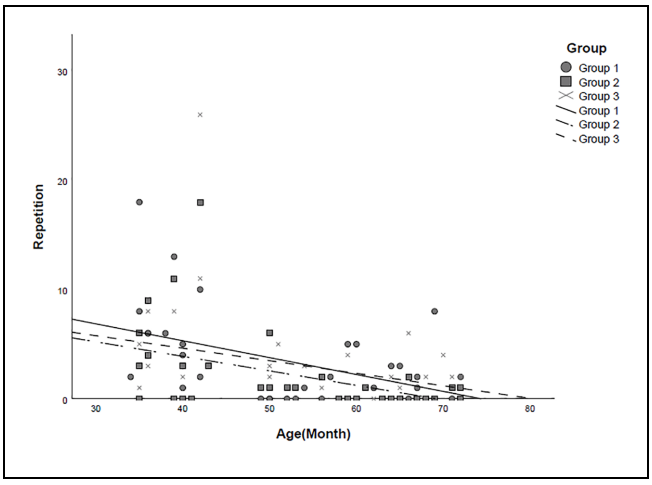

Figure 5. Cross-sectional trajectory of disfluencies (Repetitions) across age.

and two of three groups were shown significant age effects, but the linear fit showed that there is still an increasing trend. This result is different from Cheung (1998)'s finding that for children before 3.5 years of age, which MLU is valuable, but for children older than 3.5, MLU may not reflect their real language level. However, our study found that increases in MLU continue for each age group up to age 6. The result of regression analysis on MLU5 was consistent with L. Zhang and Zhou (2009), who found both MLU5 and MLU could demonstrate a developmental trend for children above age 4. It is worth mentioning that the growth trend of MLU found in this study did not include a nonlinear component such as that found for English by Channell et al. (2018) or for Mandarin by Cheung (1998). Those studies found a nonlinear relationship between syntactic complexity and age based on narrative LSA.

Prior studies on vocD have failed to find age-related trends as strong as those found here. Research on children speaking different Chinese languages (Cantonese and Mandarin) has found that vocD has different sensitivity in distinguishing between SLI and normally developing children (Cai, 2008; Wong et al., 2010). However, the above research did not explore the development trends of children of different ages. Our study found that vocD was significantly related to age and that this relation was linear, at least for two subgroups, besides, as shown in Figure 2, the values of vocD of all three groups increased continuously with age. This result supports the claim from Cai (2008) that lexical diversity can be used as an index for language disorder assessment. However, this linear correlation in this study is not so stable, and this unstable age growth effect may exhibit a mutual constraint effect between children's vocabulary and syntactic development. Jin and Jin (2008) found that children's lexical diversity increased

before 66 months, and the speed of increasing decreased after that time. These differences in findings could be caused by the ways that different speakers interacted with the children. In their study, clinical doctors talked with the children, whereas teachers conversed with children in this study. Preschool teachers are quite familiar with every child and they know how to talk effectively with children using different strategies, but clinical doctors may not be as effective in this regard.

Interestingly, we found that the number of retracings used by children increased with age, indicating that language is more precisely used for serving specific interactional needs by children (Liang, 2018), as they need more and more time to make adjustments and corrections in order to interact most effectively with their teacher. In addition, repetition may arise from the interactive nature of free play. In free play, children needed to constantly adjust syntactic structures to make it easier and clearer for listeners to understand (Tao, 2019). However, only one of the three groups of children showed a significant age effect, which is obviously not too convincing. In our opinion, this may be closely related to the relatively small number of retracing children producing. Nevertheless, this indicator still deserves further attention.

The type of disfluency that leads to repetitions is very common in young typically developing children (MacWhinney & Osser, 1977), in both narrative and conversational samples. Tseng (2006) found that of all repair types used by Mandarin-speaking children, repetition was the most frequent. In this study, we further found that repetition in early childhood showed a significant decrease with age through quantitative analysis of language samples, which was consistent with prior work from narrative-based LSA (Channell et al., 2018; Liang, 2018). For example, Channell et al. (2018) found a linear decrease in dysfluency across the full age range. The decrease in repetition may reflect the improvement of children's pragmatic abilities, as children continuously changed their speech to serve the partners they were talking with (Liang, 2018). Again, this is an important variable to consider when examining language samples in individual with language disorders of all ages, given that repetition is common during conversation.

We did not find any significant gender effects. Cao et al. (2018) found that, before age 3, girls had a higher MLU and MLT (mean length of turns) than boys. However, for older children, this difference disappeared. The differentiation of gender differences may not be significant in the early stages. Research on English speaking children had found that this difference does not appear until the age of 11, while research on Chinese speaking children suggested that it is around the age of 7 (Huang, 2014). Nonetheless, for language analysis, gender is still an important factor to be considered, especially for children under 3.

Limitations and Future Directions

There are certain limitations in this study. First, the research design is cross-sectional, making it impossible to track individual developmental characteristics. Although we see the trend of the five indicators changing systematically with age for the sub-group as a whole, we also saw differences between individuals. The linear correlation in this study is not so stable, because it cannot be well represented in all three groups for MLU5, vocD, repetition and retracing, though most indictors had great age effects for at least two groups. A larger corpus data set should be used in the future to increase the reliability of research indicators, in order to be strong enough to serve as a good clinical measure. Besides, a longitudinal design could also reveal how the above five indicators change with age in individual children. Second, demographic indicators such as family cultural and socioeconomic status should be included in future analyses. The socioeconomic status was relatively high because children were recruited in the experimental preschool, and children's mother's education level were higher than other studies (e.g., Y. Zhang et al., 2021), so, in order to increase the scope of application of the norm-referenced data set, children from more families with different socioeconomic status should be included in future. Third, further research should examine different dialect areas, apart from Shandong. Fourth, it would be good to include these five measures along with current standardized measures such as DREAM or MCELP-CS on the same group of children to examine the mutual validity of the two modes of analysis.

Conclusion

In conclusion, by exploring cross-sectional data, this study pinpointed five measures that reflect the developmental trend of typically developing children speaking Mandarin. These five LSA based variables provide options beyond standardized tools for evaluating Chinese children with developmental language disorders. The norm reference data established based on the five variables, although the current sample size is still small, can help SLP and kindergarten teachers screen for children who may be at risk of language disorders, and develop relevant intervention plans as soon as possible to help children improve their language abilities as soon as possible. Given the free nature of LSA based on CHILDES, this will further promote the rapid development of related evaluation work.

Acknowledgments

The authors would like to thank the teachers and children at Zibo, China, and graduate students at East China Normal

University for their collaboration in the collection and transcription of language samples. At the same time, the first author would like to express his deepest gratitude to his wife, Qian, and his son, Muyang, for their unconditional love toward him.

Credit Author Statement

Yibin Zhang: Conceptualization, Methodology, Resources, Writing – Original Draft. Brian MacWhinney: Methodology, and editing the grammatical and lexical errors. Jing Zhou: Writing – Review & Editing, Funding Acquisition, Project Administration.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by National Social Science Foundation (17BYY093), Ministry of Finance Fundamental Research Funds for the Central Universities (2023ECNU-YY1046) and Shanghai Philosophy and Social Sciences Fund (2023EYY004).

Ethics Statement

The study was approved by the University Committee on Human Research Protection (UCHRP) of East China Normal University, Shanghai, China. All children's parents provided written informed consent.

ORCID iD

Yibin Zhang https://orcid.org/0000-0002-3111-7143

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language*, *20*(1), 139–152.

Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. Wiley.

Brown, R. (1973). *A first language: The early stages*. Harvard University Press.

Cai, Y. F. (2008). *A study of Chinese language sample analysis for 3-5 years old children* [Master's thesis]. https://hdl.handle.net/11296/3pa2mt

Cao, L., Ran, Y. C., & Meng, X. (2018). Characteristics of language development in early childhood. *Chinese Journal of Child Health Care*, *26*(4), 437–439.

Channell, M. M., Loveall, S. J., Conners, F. A., Harvey, D. J., & Abbeduto, L. (2018). Narrative language sampling in typical development: Implications for clinical trials. *American Journal of Speech-Language Pathology*, *27*(1), 123–135.

Chen, Z. H. (2007). *The communicative acts in 4-5 years old children with mental retardation: A pragmatic research in mother-child's communicative interactive* [Doctoral dissertation, East China Normal University, Shanghai, China]. http://www.cnki.net.

Cheung, H. (1998). *Utterance length and the development of Mandarin Chinese*. Paper presented at the First Asia-Pacific Conference on Speech, Language and Hearing, University of Hong Kong, Hong Kong.

Crystal, D., Fletcher, P., & Garman, M. (Eds.). (1989). *Grammatical analysis of language disability* (2nd ed.). Whurr.

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language*, *32*, 759–786.

Eisenberg, S. L., Fersko, T. M., & Lundgren, C. (2001). The use of MLU for identifying language impairment in preschool children. *American Journal of Speech-Language Pathology*, *10*, 323–342.

Eisenberg, S. L., Guo, L.-Y., & Mucchetti, E. (2018). Eliciting the language sample for developmental sentence scoring: A comparison of play with toys and elicited picture description. *American Journal of Speech-Language Pathology*, *27*(2), 633–646.

Evans, J. (1996). Plotting the complexities of language sample analysis: Linear and non-linear dynamical models of assessment. In K. Cole, P. Dale, & D. Thal (Eds.), *Assessment of communication and language* (Vol. 6, pp. 207–256). Baltimore, MD: Brookes.

Gai, X. S., Yang, W., & Tai, Y. (2009). Review of foreign studies on children language sample analysis. *Advances in Psychological Science*, *17*(6), 1242–1249.

Guo, L., & Eisenberg, S.L. (2015). Sample length affects the reliability of language sample measures in 3-year-olds: evidence from parent-elicited conversational samples. *Language, speech, and hearing services in schools, 46*(2), 141–53

Guo, L. Y., Eisenberg, S., Ratner, N. B., & MacWhinney, B. (2018). Is putting SUGAR (Sampling utterances of grammatical analysis revised) into language sample analysis a good thing? A response to Pavelko and Owens (2017). *Language Speech and Hearing Services in Schools*, *49*(3), 622–627.

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*, 345–359.

He, H. Z., & He, L. Z. (2009). A comparative study of pragmatic development of children with hearing impairment aged from 4 to 6 and their hearing counterparts. *Journal of East China Normal University: Educational Sciences*, *27*(1), 63–71.

Heilmann, J. J., & Westerveld, M. F. (2013). Bilingual language sample analysis: Considerations and technological advances. *Journal of Clinical Practice in Speech-Language Pathology*, *15*, 87–93.

Huang, L. L. (2014). Gender differences in the comprehension of Chinese word meanings among school children. *Journal of Shenzhen University (Humanities & Social Science)*, *20*, 114–117.

Jin, X. M., & Jin, Z. J. (2008). The research on mean length of utterance and lexical diversity in preschool Mandarin-speaking children. *Chinese Journal of Evidence-Based Pediatrics*, *3*(4), 261–266.

Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders*, *12*(2), 28–41.

Klein, H. B., Moses, N., & Jean-Baptiste, R. (2010). Influence of context on the production of complex sentences by typically developing children. *Language Speech and Hearing Services in Schools*, *41*(3), 289–302.

Lee, L. L., Koenigsknecht, R. A., & Mulhern, S. T. (1974). *Developmental sentence scoring*. North Western University.

Liang, D. D. (2018). *Language disorders and acquisition in children*. Nan Jing Normal University (in Chinese.

Li, H. M. (2003). *Study on dinner table talk of preschool children family in Shanghai* [Master's thesis, East China Normal University, Shanghai, China]. http://www.cnki.net.

Li, L. (2014). *Study on development of preschool children's narrative abilities in a Mandarin-speaking context* [Doctoral dissertation, Shanghai International Studies University, Shanghai, China]. http://www.cnki.net.

Li, L. H. (2008). *Metacommunication development of children age three to six during their collaborative pretend play* [Master's thesis, East China Normal University, Shanghai, China]. http://www.cnki.net.

Li, L. H. (2011). *Preschool children's development of reading comprehension of picture storybook: From a perspective of multimodel meaning making* [Doctoral dissertation, East China Normal University, Shanghai, China]. http://www.cnki.net.

Liu, S. Y. (2018). *Study on the development of learning story completion skills among children aged 3-5* [Master's thesis, East China Normal University, Shanghai, China]. http://www.cnki.net.

Liu, X. L., de Villiers, J., Ning, C., Rolfhus, E., Hutchings, T., Lee, W., Jiang, F., & Zhang, Y. W. (2017). Research to establish the validity, reliability, and clinical utility of a comprehensive language assessment of mandarin. *Journal of Speech Language and Hearing Research*, *60*(3), 592–606.

Li. X. Y. (2008). A case study of the language development and communicative skills for Chinese young children with autism [Doctoral dissertation, East China Normal University, Shanghai, China]. http://www.cnki.net.

Lu, L., & Liu, H. (1998). *The peabody picture vocabulary test-revised in Chinese*. Psychological Publishing.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs*. Psychology Press.

MacWhinney, B., & Osser, H. (1977). Verbal planning functions in children's speech. *Child Development*, *48*, 978–985.

MacWhinney, B., & Ratner, N. (2016). Your laptop to the rescue: Using the Child Language Data Exchange System Archive and CLAN utilities to improve child language sample analysis. Seminars in Speech and Language, *37*(2), 74–84.

MacWhinney, B., Roberts, J. A., Altenberg, E. P., & Hunter, M. (2020). Improving automatic IPSyn coding. *Language Speech and Hearing Services in Schools*, *51*, 1187–1189.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*(1), 85–104.

Miller, J. F., Andriacchi, K., & Nockerts, A. (2016). Using language sample analysis to assess spoken language production in adolescents. *Language Speech and Hearing Services in Schools*, *47*(2), 99–112.

Miller, J. F., & Iglesias, A. (2015). *Systematic analysis of language transcripts (SALT)* [Version 16.1. 5. Computer software]. SALT Software, LLC.

Ng, S. H. (2017). *A study of Chinese lexical and syntactic development of Chinese preschoolers from English speaking families in Singapore* (Doctoral dissertation, East China Normal University, Shanghai, China]. http://www.cnki.net.

Nippold, M. A., Frantz-Kaspar, M. W., Cramond, P. M., Kirk, C., Hayward-Mayhew, C., & MacKinnon, M. (2014). Conversational and narrative speaking in adolescents: Examining the use of complex syntax. *Journal of Speech Language and Hearing Research*, *57*, 876–886.

Niu, M. M. (2018). *Research on the development of expository language in children aged 4-6 years* [Master's thesis, East China Normal University, Shanghai, China]. http://www.cnki.net.

Oh, S. J., Yoon, J. H., & Lee, Y. (2020). A qualitative study on experiences and needs of language sample analysis by speech–language pathologists: Focused on children with language disorders. *Communication Sciences & Disorders*, *25*(2), 169–189.

Ouyang, X. M. (2003). *The influence of children's pragmatic development on their mothers' speech use* [Master's thesis, Nan Jing Normal University, Nanjing, China]. http://www.cnki.net.

Paul, R., & Norbury, C. F. (2012). *Language disorders from infancy through adolescence*. St. Elsevier Health Sciences.

Pavelko, S. L., & Owens, R. E. (2017). Sampling utterances and grammatical analysis revised (SUGAR): New normative values for language sample analysis measures. *Language Speech and Hearing Services in Schools*, *48*(3), 197–215.

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

Pezold, M. J., Imgrund, C. M., & Storkel, H. L. (2020). Using computer programs for language sample analysis. *Language Speech and Hearing Services in Schools*, *51*(1), 103–114.

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & 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.

Rondal, J. A., Ghiotto, M., Bredart, S., & Bachelet, J. F. (1987). Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of Child Language*, *14*(3), 433–446.

Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics*, *11*(1), 1–22.

Sealey, L. R., & Gilmore, S. E. (2008). Effects of sampling context on the finite verb production of children with and

without delayed language development. *Journal of Communication Disorders*, *41*(3), 223–258.

Sheng, L., Shi, H., Wang, D., Hao, Y., & Zheng, L. (2020). Narrative production in Mandarin-speaking children: Effects of language ability and elicitation method. *Journal of Speech Language and Hearing Research*, *63*(3), 774–792.

Southwood, F., & Russell, A. F. (2004). Comparison of conversation, freeplay, and story generation as methods of language sample elicitation. *Journal of Speech Language and Hearing Research*, *47*(2), 366–376.

Strömqvist, S., & Verhoeven, L. (Eds.). (2004). *Relating events in narrative: Typological and contextual perspectives: Volume 2 of the frog story studies*. Lawrence Erlbaum Associates, Inc.

Tao, L. (2019). Self-repair in Mandarin Chinese: The multimodality of conversation. In X. Li & T. Ono (Ed.), Multimodality in Chinese Interaction (pp. 255-299). Berlin, Boston: De Gruyter Mouton.

Templin, M. C. (1957). *Certain language skills in children*. University of Minnesota Press.

Tseng, S. C. (2006). Repairs in Mandarin conversation. *Journal of Chinese Linguistics*, *34*(1), 80–120.

Wallace, I. F., Berkman, N. D., Watson, L. R., Coyne-Beasley, T., Wood, C. T., Cullen, K., & Lohr, K. N. (2015). Screening for speech and language delay in children 5 years old and younger: A systematic review. *Pediatrics*, *136*(2), e448–e462.

Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech Language and Hearing Research*, *38*(6), 1349–1355.

Weng, C. Q. (2016). *The syntactic level of children in the context of autonomous expression* [Master's thesis, Shan Xi Normal University, Shanxi, China]. http://www.cnki.net.

Wong, A. M. Y., Klee, T., Stokes, S. F., Fletcher, P., & Leonard, L. B. (2010). Differentiating Cantonese-speaking preschool children with and without SLI using MLU and lexical diversity (D). *Journal of Speech Language and Hearing Research*, *53*(3), 794–799.

Wu, H., Liu, Q., Yu, B., Zhang, Y., Ren, D., Serdarevic, M., Liang, Z., Wang, Y., Chen, S., Zhang, K., & Chen, S. (2020). Psychometric properties of the mandarin clinical evaluation of language for preschooler's core scale. *Journal of Communication Disorders*, *87*, 106000.

Wu, S., Zhao, J., de Villiers, J., Liu, X. L., Rolfhus, E., Sun, X., Li, X., Pan, H., Wang, H., Zhu, Q., Dong, Y., Zhang, Y., & Jiang, F. (2023). Prevalence, co-occurring difficulties, and risk factors of developmental language disorder: first evidence for Mandarin-speaking children in a population-based study. *The Lancet Regional Health – Western Pacific*, *34*, 100713.

Yang, S. L. (2015). *Communication and communication disorders: Theory and practice*. Psychological Publishing.

Yang, X. L. (2018). *Peer interactive language in role play situation with children from 4 to 6 years old* [Doctoral dissertation, East China Normal University, Shanghai, China]. http://www.cnki.net.

Yim, D., Park, W., Cheon, S., Lee, Y. J., & Lee, J. (2015). Interaction Skills via spontaneous. Speech Sample Analysis: 2- to 5-year-old children with and without language impairment. *Communication Sciences & Disorders*, *20*(3), 364–373.

Yu, L., Yang, C. K., Liao, Y. K., & Zhu, Y. S. (2017). The study of validity in language sampling indicators for Chinese children in Guangzhou. *Bulletin of Eastern-Taiwan Special Education*, *106*(19), 129–155.

Yu, S. Y. (2017). *Study on narrative skills of preschool children based on IPAD* [Master's thesis, East China Normal University, Shanghai, China]. Retrieved from http://www.cnki.net.

Zhang, F. (2013). A study of narrative development of young Chinese children with specific language impairment aged four to six years. In A. McCabe & C.-J. Chang (Eds.), *Chinese language narration: Culture, cognition, and emotion* (pp. 143–179). John Benjamins.

Zhang, L., & Zhou, J. (2009). The development of mean length of utterance in Mandarin-speaking children. In J. Zhou (Ed.), *The application and development of international corpus-based research methods (in Chinese)* (pp. 40–58). Education Science Publishing House.

Zhang, Y., Xu, X., Jiang, Y., Sun, W., Wang, Y., Song, Y., Dong, S., Zhu, Q., Jiang, F., & Sheng, L. (2021). Early language and communication development in Chinese children: Adaption and validation of a parent report instrument. *International Journal of Speech-Language Pathology*, *23*, 372–381.

Zhang, Y. B., Jiang, Z. X., Zhang, H. Q., & Guo, L. P. (2016). The efficacy of the interventions for children with primary speech and language delay or disorder on the Chinese mainland: A meta-analysis and its implications. *Chinese Journal of Special Education*, *196*(10), 19–25.

Zhou, J. (2001). *Pragmatic development of Mandarin speaking young children: From 14months to 32 months* (Unpublished doctoral dissertation). The University of Hong Kong, Hong Kong.

Zhou, J., Li, C. J., Du, L. J., Wang, F. X., & Chen, S. (2014). The development of Mandarin acquisition of Uygur children in bilingual preschools in Xinjiang. *Journal of East China Normal University: Educational Sciences*, *32*(1), 11–19.

Zhou, J., Zhang, L., & Li, C. J. (2017). Vocabulary development of pre-school Mandarin-speaking children. *Chinese Journal of Language Policy and Planning*, *2*(6), 89–96.

Zhou, J., & Zhang, Y. B. (2020). The computerized child language analysis: Building a non-standardized child language measuring system based on Chinese language development corpus. *Studies in Early Childhood Education*, *306*(6), 72–84.

Zhu, M. (1986). *Research in child language development*. East China Normal University (in Chinese).