Article

JOURNAL OF CHINESE WRITING SYSTEMS

Building a norm-referenced dataset for vocabulary assessment based on Chinese vocD and word classes

Journal of Chinese Writing Systems
2020, Vol 4(1) 5–17

DOI: 10.1177/2513850219894077
journals.sagepub.com/home/cws
SAGE

Yibin Zhang
Institute of Brain and Education Innovation, East China Normal University, China

Jing Zhou
Faculty of Education, East China Normal University, China

Abstract
Exploration of trajectories of expressive language samples is essential for understanding potential indicators for language disorder assessment. This study examined conversational language samples from 341 typically developing Mandarin-speaking children aged 3–7. Through analysis of lexical diversity and word classes, a norm-referenced dataset for vocabulary assessment was built, including indicators such as vocD and the types and tokens of nouns, verbs, measures, adverbs, conjunctions and prepositions. As norm-referenced indicators for the language development of children speaking Mandarin, these developmental data could also inform clinical therapists about the direction of intervention for children with vocabulary deficits.

Keywords
Lexical diversity, Mandarin-speaking children, norm-referenced datasets, vocabulary assessment, word classes

Vocabulary ability is one of the most important language skills for children. The early vocabulary ability of infants and young children could predict language development after preschool (Lin and Liu, 2017) and the grammatical development of Mandarin-speaking children at the age of two to three years (YW Zhang et al., 2002). Studies have shown that children with language disorders had deficiencies in vocabulary learning (Wright et al., 2018), which therefore made it difficult for them to learn new words and use new words (Kan and Windsor, 2010), affected their understanding of the concept of the world, and stopped them from integrating words into their life. Thus, it is quite important to assess their vocabulary skills for further intervention.

There are mainly two types of approach which have been used for assessing language skills within the Chinese context; one was standardized and the other was non-standardized. The former one mostly resorted to laboratory measures, such as Gesell Developmental Schedules, Bayley Scales of Infant Development, 0~6-year-old Pediatric Examination Table of Neuropsychological Development, Sign-Significant Relations – Chinese version, and Wechsler Preschool & Primary Scale of Intelligence (WPPSI) (see meta-analysis; YB Zhang et al., 2016), and there were referenced norms for all of those measures. Although this approach had very high research efficiency, an inevitable low ecological validity problem made the results difficult to be explained. Meanwhile, many researchers had begun to explore non-standard approaches such as language sample analysis to build a norm-referenced dataset used for assessing language development.

Language sample analysis gives us an alternative approach that can trace the characteristics and trajectory of normally developing children's language in action. This study attempted to find relevant indicators from spontaneous language samples based on Child Language Data Exchange System (CHILDES; MacWhinney, 2000), the largest free child corpus. Research in this area had been tested with English-speaking children (Ratner and MacWhinney, 2016). We explored several indicators that reflected the development of overall lexical diversity and different word classes of

Corresponding author:
Jing Zhou, Tianjiabing Building, Room 511, East China Normal University, 3663, North Zhongshan Road, Shanghai, 200062, China.
Email: jzhou@spe.ecnu.edu.cn

normally developing Mandarin-speaking children and established a norm-referenced dataset, which could be used for further language disorder assessment.

Lexical diversity in children speaking Mandarin: vocD and other indexes

Lexical diversity can be used to measure the vocabulary level of normally developing children, and could also be used to distinguish children with language disabilities, such as children with specific language impairment (SLI) (Wong et al., 2010) and children who stutter (Silverman and Ratner, 2002). Although the study of lexical diversity has been around for a long time, there has been a huge controversy about the metrics characteristics of lexical diversity, and this controversy has also witnessed the history of research on lexical diversity.

Internationally, three main indicators have been used in clinical practice to reflect children's lexical diversity: TTR (type/token ratio), NDW (number of different words) and vocD. Those three indicators were also frequently used in the Chinese context, and different researchers applied one or several indicators to evaluate children's lexical diversity. On the one hand, they studied the development of lexical diversity in normally developing children; on the other hand, they tried to explore different lexical diversity indicators that could be used to distinguish children with language disorder from their peers. However, different lexical diversity indicators all had a few problems.

NDW

Calculation of the NDW was the most straightforward way to measure lexical diversity. A study by Wong et al. (2010) found that children produced more complete and intelligible utterances and had higher NDW compared with children with SLI. Hao et al. (2018) compared the narrative abilities of children with language disorder and children who were not. They found that children with language disabilities had significant weaknesses in lexical diversity, indicating that NDW could be used as an indicator to distinguish children with language disorders. However, there were still two major problems when using NDW for analysis. First, clinical studies had found that the NDW from children's expressive samples was affected by the size of the language samples and the mean length of utterance (MLU). Children who produced longer MLU had a higher NDW level (Klee, 1992). In clinical practice, in order to solve this problem, some researchers have tried to standardize the number of utterances, the duration time for language sample collection and even the total number of words, etc. However, the unification of the number of utterances could not guarantee the unity of the number of words in utterances and the same time duration. There were differences in the number of words produced per minute for children of different ages and different language proficiencies; therefore, the control of the total number of words may not reflect the level of language development of children because the number of words was too small. Second, researchers seemed to confuse the vocabulary skills with conversational abilities that NDW represented. Since the total number of words produced by children was affected by children's conversations per unit time, children with high levels of conversation would naturally produce more utterances, increasing the total number of words they produced. All in all, NDW may be more relevant to children's level of conversation than to lexical diversity.

TTR

Another frequently used indicator for measuring lexical diversity was TTR. The generation of TTR attempted to improve the accuracy of NDW to some extent, which was calculated by the ratio of NDW rather than the frequency of words. TTR was the most widely used indicator for lexical ability assessment. Weng (2016) analyzed the lexical diversity of children in small and medium-sized classes and found that the vocabulary of children increased year by year, but there was a difference between the diversity of each word class. The TTR for nouns was the highest while the pronouns, adverbs and other function words were basically at a relatively low level.

There were also several problems for TTR. Due to the repeated appearance of certain closed words, such as pronouns and prepositions, the children with a larger sample size would have smaller TTR values because when children focused on certain topics or topics, their use of real words would be repeated. It also led to a smaller TTR (Owen and Leonard, 2002), even when more utterances or total number of words were included for analysis (She et al., 2017). In addition, the researchers also found the impact of corpus collection scenarios on lexical diversity. HW Chen (2005) analyzed and compared 12 children's language samples from conversation, storytelling and story retelling. It was found that there was a significant difference between storytelling and story retelling in TTR.

Researchers have also tried to improve the accuracy of TTR, such as using a standard number of utterances (Templin, 1957), measuring closed words and open words separately (Richards, 1987) or using standard number of words (Klee,

1992). However, although it seemed that the use of a standard number of words was an important means of improving the accuracy of TTR, researchers did not agree on the number of words (Owen and Leonard, 2002).

vocD

In addition to the above indexes, Malvern and Richards (2002) created vocD. VocD was mainly achieved by repeatedly calculating the TTR within a certain range of words (35–50), which could not be affected by the number of samples. Lai and Schwanenflugel (2016) found a significant correlation between vocD and expressive verbal test (EVT), and there were only racial differences in EVT, not in vocD. In the study of Chinese, Jin and Jin (2008) firstly applied vocD and found that the vocabulary development of children increased with age before 66 months, but decreased after 66 months. However, it did not reveal the difference between children with SLI and age-matched children in vocD (Wong et al., 2010). Therefore, there was still disagreement about the use of the vocD in Mandarin-speaking children's language assessment.

Few studies had put the three indicators together in the Chinese context, and no studies had applied these indicators to construct a vocabulary-based norm-referenced dataset that could be directly used for clinical diagnosis. Therefore, in this study, we first compare three indicators, and try to find an index representing overall development of vocabulary skills and apply it to construct referenced data for further vocabulary assessing.

The development of word class in children speaking Mandarin

There are real words and function words in Mandarin. Real words are words that can serve as sentence components and have certain meanings, including nouns, verbs, adjectives, numerals, measures, pronouns, adverbs, etc., while function words have no clear semantics, and mainly include prepositions, conjunctions, auxiliary words, sound words and interjections (Liu et al., 2001). She et al. (2017) showed that the ratio between real words and function words was 8:2 and the verbs and nouns were the top 2 in number among real words. Although the researchers analyzed the real words and function words, more research focused on a certain type of real words or function words. The specific studies were as follows.

Nouns

GM Chen (2009) found that real words were firstly acquired by children, of which the nouns were the earliest and most they came into contact with, followed by verbs and adjectives. TX Zhang (2010) found that the total number of nouns used by children aged three to six accounted for more than half of their total vocabulary, and children produced the most nouns representing people and things. Nouns for time and direction were produced relatively rarely. In addition to the study of the development of nouns as a whole word class, many researchers have conducted more detailed research on the subclasses of nouns, to explore the microscopic features of children's noun development. Ou (2009) found that the acquisition of nouns generally followed the order from individual nouns to proper nouns to collective nouns and abstract nouns, which was the same as English-speaking children. All of the above studies illustrate the important value of nouns in early vocabulary development.

Verbs

The development of verbs was second only to nouns. TX Zhang (2010) found that verbs account for about one-third of the high-frequency words used by children, second only to nouns. A subclass analysis of verbs found that the earliest verbs in children were the existential verb 有 *you*, psychological verb 喜欢 *xihuan*, auxiliary verb 能 *neng*, directive verb 下 *xia* and verb of judging 是 *shi* (Y Liu, 2016). Sometimes children used auxiliary verbs such as 会 *hui* as active verbs; for example, 我会这个 *wo hui zhe ge* (I will do this) (Erbaugh, 1992). Tardif's (1995) analysis of children's word classes showed that there were differences in age between the total number of different word classes. Children with high socioeconomic status also produced more words than children with low socioeconomic status. After analyzing the top 50 words of the 4 children, Peng (2009) found that the first words acquired by 2 children were not nouns, but verbs, such as 打*hit* and 搞 *gao*. The three children had more nouns than verbs.

Adjectives

Zhu's (1987) survey on adjectives found that the use of adjectives increased from two to six years of age, with the fastest increase at the age of four years and six months. The use of adjectives was more related to the description of the object and was also used as early as two years old. At the age of two years and six months, children began to use words related to their body. At the age of three, they produced adjectives describing movements. At the age of three years and six months, they used adjectives that describe the shape of the human body. At the age of four years and six months, adjectives describing individual qualities, expression of emotions, and event descriptions were first produced. TX Zhang (2010) found that there was a significant age effect for the number of words, the amount of words and the frequency of adjectives used by children aged three to six. The most acquired adjectives were monosyllabic adjectives, while polysyllabic adjectives appeared less (Y Liu, 2016). Children first acquired adjectives that express external characteristics and nature evaluation.

Adverbs

Che (2009) analyzed three children's longitudinal corpus and found that the main adverbs used by children were 不 *bu,*, 还 *huan* and 好 *hao*. The frequency of adverbs increased significantly with age, and the negative adverb 不 *bu* had the highest frequency, followed by 就 *jiu,* 还 *hai*, 都 *dou*, 也 *ye*, 没 *mei*, 就是 *jiushi*, 很 *hen*, 再 *zai*, 又 *you* and 没有 *meiyou* (TX Zhang, 2010), which was basically similar to the characteristics of the development of the two-year-old children's adverbs studied by Che (2009). J Zhang (2013) analyzed the longitudinal corpus of two children and found that the order of acquisition of adverbs was 好 *hao*, 太 *tai*, 很 *hen*, 这么 *zheme*, 最 *zui*, 还 *hai*, 更 *geng*, 那么 *name*, 真 *zhen*, 可 *ke*, 老 *lao*, 蛮 *man*, 非常 *feichan*g and 特别 *tebie*, of which the most frequent ones were 好 *hao*, 太 *tai*, 很 *hen*, 这么 *zheme* and 最 *zui.*

Measures

Measures was a specific part of speech for Mandarin. Zhu's (1987) analysis revealed that children at the age of four had acquired the individual measures 个 *ge* and 只 *zhi*, but their generalization was common. Children acquired 条 *tiao* and 本 *ben* by the age of 5, and 10 individual classifiers by the age of 6. The correct rate of using temporary measures also increased year by year, while the use of aggregate measures had developed relatively slowly. Even by the age of seven, only 双 *Shuang* had been acquired. TX Zhang (2010) found that the most used measure was个 *ge*, which exceeded the total number of other measures by frequency. Although the use of specific measures gradually increased after two years and six months, it still only reached half the level of adult measures before the age of four, and there were still some errors in using measures by the age of seven (Erbaugh, 1992). Yang and Zhou (2010) analyzed eight pairs of mother–child conversational samples and found that the number of measures children used gradually increased, but there was no age effect in frequency. A separate survey of the individual measure 个 *ge* found that it gradually decreased from three years and six months to four years old, but the types of measures were gradually produced increasingly, leading to the overall increasing use of measures. Yan (2010) found that four to five years old was an important stage in the acquisition of measures, and measures produced by children with retardation could reach the level of normally developing children later. In general, it followed the order from life measures to shape measures to functional measures.

Conjunctions

Conjunctions, as a type of function word, had the main function of connecting two or more words, phrases or clauses. Zhu (1987) found that the number of conjunctions used by three-year-old children was few, mainly including 还有 *hai-you*, 也 *ye*, 又 *you* and 就 *jiu*, but the frequency of use increased with age. In addition, L Li's (2014) analysis of children's narrative samples found that children's use of conjunctions increased slowly from 48 to 60 months and grew rapidly after 60 months. By analyzing the peer talk in dramatic play, WJ Zhang and Zhou (2016) found that the ratio of children using conjunctions was not high, but the types and tokens of conjunctions increased with age.

Pronouns

There are three types of pronouns in Mandarin: personal, interrogative and demonstrative. Names and nouns are often used to substitute words, so the use of pronouns is much less than in English. However, personal pronouns are acquired

earlier. Kong et al. (2004) studied the acquisition of personal pronouns and found that children basically learned the main personal pronouns, 我 *wo*, 我们 *women*, 你 *ni*, 你们 *nimen*, 他 *ta*, 他们 *tamen*, 自己 *ziji*, 人家 *renjia*, 大家 *dajia* and 咱们 *zanmen* before the age of three. TX Zhang (2010) found that the number of interrogative pronouns changed little before the age of six and increased after six years old. The most used pronouns were 什么 *shenme*, 啥 *sha*, 怎么 *zenme*, 谁 *shui*, 哪 *na*, 哪个 *nage*, 为什么 *weishenme*, 干嘛 *ganma*, 咋 *za* and 哪儿 *na'er*, and the earliest appearance of demonstrative pronouns in children was the proximal demonstrative pronoun 这 *zhe*, which was also the most productive pronouns.

Prepositions

Prepositions cannot be used alone in Mandarin. They need to be combined with nouns, verbs and pronouns to form a prepositional phrase or a prepositional structure. The prepositional phrase and prepositional structure cannot be used alone and need to be attached to the predicate as an adverbial or complement. There were not too much prepositions, but the 把 *ba* and 被 *bei* were the most important. XN Li et al. (1990) found nine categories for *ba* constructions, and children acquired *ba* constructions from the age of two and became quite familiar with *ba* constructions at the age of four years and six months. Tian (2008) found the onset time for producing *bei* was at the age of one year and four months. Also, K Li (2015) found that children with SLI produced fewer *bei* than those in an age-matched group and had more errors in *bei* productions.

In conclusion, for measuring the overall level of vocabulary development, different researchers preferred different indicators, including NDW, TTR and vocD, but the results based on these three indicators were not consistent. Researchers also did not have referenced data for these indicators. In the number of outputs of different word classes, the existing research mainly uses the types or tokens to measure (e.g. Xia, 2008); however, the research found that the development of different word classes had very strong asynchrony, which required us to further analyze to construct referenced data based on developmental analysis. Therefore, based on the above two dimensions for the construction of a vocabulary diagnosis system, it is necessary to determine the indicators that need to be included. A corpus-based vocabulary diagnosis system based on CHILDES consists of two main dimensions; one is the overall level of lexical diversity, and the other is the different levels of different word classes. We try to solve two questions in this study, based on those considerations.

First, could NDW, TTR and vocD measure the development of lexical diversity for Chinese children at different ages? What is the referenced data constructed by this indicator based on development characteristics?

Second, in terms of different word classes, which one will increase with age? Which one will not? What is the referenced data based on the development of different word types and tokens?

Method

Participants

In this study, two parts of a corpus from the same context were recruited; one was from CHILDES-zhou2 and the other was zhou4, a newly collected corpus. All corpus were collected from conversational free play and could be traced in CHILDES (https://childes.talkbank.org/access/Chinese/Mandarin/ZhouAssessment.html). Three hundred and forty-one normally developing children were recruited randomly from different preschools of moderate social-economic status in China. The children were 36–72 months of age (M = 53.66, SD = 11.51) (Table 1) and spoke Mandarin as their native language. We required teachers to confirm that all children's performance was similar to those of their peers and that they had never received special services. The language sample collection procedure had the approval of the Ethics Committee of East China Normal University in China, and every child's parent signed consent forms.

Table 1. Participants in this study.

Group	*N*	Mean age (months)	SD	Sources
3–4 yr	119	40.73	3.349	zhou2(32); zhou4(87)
4–5 yr	103	53.62	4.076	zhou2(33); zhou4(70)
5–6 yr	119	66.61	4.294	zhou2(49); zhou4(70)
Total	341	53.66	11.511	/

Material and procedure

In this study, we used adult (teacher/mother)–child free play to obtain spontaneous language samples, because teachers or mothers could elicit more language samples as they were familiar people. The adult brought the child to a play center at preschool or at home and started the procedure by asking *Shall we start playing the game?* and starting to record the conversation if the child gave an answer of *Yes*. After 20 minutes, the play session was ended by a question: *Shall we take a break?* The whole play sessions were recorded via audio-recording devices, such as cellphones. Materials for extending conversation were used, including a big toy house with toy furniture like desks, chairs, beds, cooking devices, and some animal toys like toy pigs and toy cats that are frequently shown in cartoons, which would increase the children's gaming time.

Reliability

Language samples were transcribed by graduate students and rechecked by the first author. All students were accepted onto a one-week-training workshop held by the first author. Disagreements were identified and solved by discussion between the first author and all graduate students.

Transcription and language sample analysis

The whole play recordings were firstly transcribed by graduate students. Everything the child said was written down, including pause, repetition and retracing. All transcripts were stored on computers in CHAT (Codes for the Human Analysis of Transcripts) format, a standardized format required for further language sample analysis based on CLAN (Computerized Language ANalysis). After transcription, all CHAT files were added %mor tier and %gra tier by running command MOR in CLAN, and all indicators would be calculated automatically via computer.

Results

Lexical diversity

Through ANOVA, we explored the development trends of NDW, TTR and vocD, which measured the overall development of vocabulary. The results showed that the age effect of NDW and TTR did not exist. However, there was a significant age effect of vocD, with five to six years significantly higher than three to four years (Table 2). Therefore,

Table 2. Descriptive analysis, F-test and multiple comparison of NDW, TTR and vocD.

Indicators	Group	N	M	SD	Min	Max	F	Multiple comparation
NDW	3–4 yr	119	52.13	7.786	33	68	F = 1.429	/
	4–5 yr	103	52.78	7.653	36	75		
	5–6 yr	119	53.80	7.513	35	69		
TTR	3–4 yr	119	0.36	0.074	.220	.656	F = .577	/
	4–5 yr	103	0.37	0.077	.175	.627		
	5–6 yr	119	0.36	0.082	.184	.597		
vocD	3–4 yr	119	49.85	15.226	22.52	94.94	F = 4.909**	5–6 yr > 3–4 yr
	4–5 yr	103	53.21	14.154	22.80	95.08		
	5–6 yr	119	55.74	14.211	23.25	97.48		

*p < .05, **p < .01, ***p < .001.

Table 3. Norm-referenced data for Chinese vocD.

	Group	M	SD	Min	Max
vocD	3–4 yr	49.85	15.226	22.52	94.94
	4–5 yr	53.21	14.154	22.80	95.08
	5–6 yr	55.74	14.211	23.25	97.48

vocD could be used as an index to construct reference data for the overall development of Mandarin-speaking children's vocabulary skills. The reference data based on vocD is shown in Table 3.

Word classes

The above analysis mainly revealed the overall development characteristics of Chinese children's vocabulary, and established reference data based on vocD. Then, specific to different word classes, what indicators can be used for constructing norm-referenced datasets? In this part, we mainly analyze the type and tokens of eight different word classes. The specific results are shown in Table 4 below.

Nouns

ANOVA showed that there was a significant age effect on the types of nouns produced by Mandarin-speaking children. Multiple comparisons showed that the five to six age group was significantly higher than the four- to five-year-old group and the three- to four-year-old group. In addition, there was a significant age effect of tokens. Multiple comparisons showed that the five to six age group was higher than the four to five-year-old group and the three- to four-year-old group, while the difference between the other groups was not significant.

Verbs

ANOVA showed that the types of verbs produced by children had significant age effects. Multiple comparisons showed that the five to six age group was significantly higher than the four- to five-year-old group and the three- to four-year-old group. There was also significant age effect on the tokens of verbs. Multiple comparisons showed that the five- to six-year-old group was significantly higher than the four- to five-year-old group and the three- to four-year-old group, with the same developmental pattern as the types of verbs.

Pronouns and adjectives

There was no significant age effect in types and tokens of pronouns and adjectives that children produced.

Adverbs

There was a significant age effect on the types of children's adverbs. Multiple comparisons showed that the five to six age group was significantly higher than four to five years old and three to four years old, and the four- to five-year-old group was significantly higher than the three- to four-year-old group. Furthermore, the tokens of adverbs also had significant age effect. The five to six years old group was significantly higher than the four to five years old group and the three to four years old group, indicating that five to six years old was a critical period for producing adverbs.

Table 4. Descriptive analysis, F-test and multiple comparation of different word classes.

Indicators	Group	N	M	SD	Min	Max	F	Multiple comparation
Noun/types	3–4 yr	119	30.83	9.463	10	55	F = 16.447**	5–6 yr > 4–5 yr; 5–6 yr > 3–4 yr
	4–5 yr	103	33.90	10.178	15	68		
	5–6 yr	119	38.95	12.978	15	86		
Noun/tokens	3–4 yr	119	54.92	19.937	17	112	F = 13.745**	5–6 yr > 4–5 yr; 5–6 yr > 3–4 yr
	4–5 yr	103	60.72	24.282	20	153		
	5–6 yr	119	73.37	35.945	23	199		
Verbs/types	3–4 yr	119	31.16	10.094	12	56	F = 8.445**	5–6 yr > 4–5 yr; 5–6 yr > 3–4 yr
	4–5 yr	103	30.83	9.300	13	69		
	5–6 yr	119	36.09	12.801	17	74		
Verbs/tokens	3–4 yr	119	67.10	25.535	28	138	F = 8.479**	5–6 yr > 4–5 yr; 5–6 yr > 3–4 yr
	4–5 yr	103	66.54	23.849	24	139		
	5–6 yr	119	80.82	37.604	28	199		
Pronouns/ types	3–4 yr	119	8.97	2.615	3	17	F = 1.359	/
	4–5 yr	103	8.74	2.574	4	16		
	5–6 yr	119	9.34	2.972	3	17		
Pronouns/ tokens	3–4 yr	119	46.62	22.453	4	107	F = 1.991	/
	4–5 yr	103	43.73	26.558	6	202		
	5–6 yr	119	50.82	30.530	4	156		
Adjectives/ types	3–4 yr	119	8.96	3.604	2	23	F = 2.381	/
	4–5 yr	103	9.83	3.628	3	19		
	5–6 yr	119	9.99	4.414	3	23		
Adjectives/ tokens	3–4 yr	119	22.16	10.499	7	55	F = .554	/
	4–5 yr	103	22.60	10.163	6	67		
	5–6 yr	119	23.73	14.221	5	104		
Adverbs/types	3–4 yr	119	10.57	4.293	2	23	F = 16.568**	5–6 yr > 4–5 yr; 5–6 yr > 3–4 yr; 4–5yr > 3–4 yr
	4–5 yr	103	12.56	4.188	2	28		
	5–6 yr	119	14.25	6.019	5	31		
Adverbs/ tokens	3–4 yr	119	27.41	13.653	6	70	F=10.487**	5–6 yr > 4–5 yr; 5–6 yr > 3–4 yr
	4–5 yr	103	31.04	13.890	10	90		
	5–6 yr	119	38.03	24.319	8	138		
Measures/ types	3–4 yr	119	3.83	1.879	1	9	F = 11.665**	5–6 yr > 4–5 yr; 5–6 yr > 3–4 yr
	4–5 yr	103	4.01	1.763	0	9		
	5–6 yr	119	5.04	2.468	1	13		
Measures/ tokens	3–4 yr	119	18.83	9.849	1	46	F = 6.203**	5–6 yr > 3–4 yr
	4–5 yr	103	20.59	13.743	0	76		
	5–6 yr	119	24.97	16.932	1	122		
Conjunctions/ types	3–4 yr	119	2.05	1.692	0	6	F = 14.316**	3–4 yr; 4–5 yr > 3–4 yr
	4–5 yr	103	2.85	1.746	0	7		
	5–6 yr	119	3.37	2.239	0	10		
Conjunctions/ tokens	3–4 yr	119	3.75	4.061	0	21	F = 14.958**	5–6 yr > 4–5 yr; 5–6 yr > 3–4 yr; 4–5 yr > 3–4 yr
	4–5 yr	103	5.87	5.231	0	29		
	5–6 yr	119	8.62	9.790	0	48		

(*continued*)

Table 4. Continued

Indicators	Group	N	M	SD	Min	Max	F	Multiple comparation
Prepositions/ types	3–4 yr	119	2.35	1.232	0	5	F = 10.550**	5–6 yr > 4–5 yr; 5–6 yr > 3–4 yr
	4–5 yr	103	2.50	1.171	0	5		
	5–6 yr	119	3.03	1.131	0	6		
Prepositions/ tokens	3–4 yr	119	6.88	6.534	0	39	F = 11.952**	5–6 yr > 4–5 yr; 5–6 yr > 3–4 yr
	4–5 yr	103	6.85	5.272	0	27		
	5–6 yr	119	10.92	9.270	0	44		

*p<.05, **p<.01, ***p < .001.

Measures

There was a significant age effect on the types of measures children produced. Multiple comparisons showed that the five to six years age group was significantly higher than the four- to five-year old group and the three- to four-year-old group. It could be seen that five to six years old was the fastest growing period. Further analysis showed that there was also a significant age effect on the tokens of measures. Multiple comparisons showed that the five- to six-year-old group was significantly higher than the three- to four-year-old group.

Conjunctions

Analysis showed that there was a significant age effect on the types of conjunctions produced by children. Multiple comparisons showed that the five to six years old group was significantly higher than three to four years old group, and the four to five years old group was significantly higher than the three to four years old group. There was also significant age effect on the tokens of children's conjunctions. Multiple comparisons showed that the five to six age group was significantly higher than the four to five years old group and the three to four years old group, and the four to five years old group was significantly higher than the three to four years old group.

Prepositions

Results showed that there was a significant age effect on the types of prepositions. Multiple comparisons showed that the five to six age group was significantly higher than the four- to five-year-old group and the three- to four-year-old group. In addition, analysis of the tokens of prepositions showed that there was a significant age effect. The five to six age group was significantly higher than the four-to five-year-old group and the three- to four-year-old group.

Based on the above analysis, we explored the developmental characteristics of eight main categories of word classes, and found that there was a significant age effect on the types/tokens of nouns, verbs, adverbs, measures, conjunctions and prepositions produced by children. We could then use these indicators to build norm-referenced datasets as shown in Table 5.

Discussion

As an alternative approach to studying language development, language sample analysis has been used for long time in mainland China. Although most of the language samples collected before were finally stored or shared in CHILDES, none was used to establish a norm-referenced dataset for language disorder assessment. Using a greater scope of language samples, this study built a language disorder assessment system, a norm-referenced dataset that could be used for analyzing the vocabulary development of Chinese children.

Three frequently used indicators for overall lexical development were compared and analyzed through a wider spread of data. Results showed that only vocD could be used as an index demonstrating the development trend of vocabulary for children speaking Mandarin, while TTR and NDW could not. The exploration of vocD in this study found similar results to those of Jin and Jin (2008), who also confirmed the validity of using vocD in Chinese children's language analysis. Jin and Jin found that vocD increased with age for Mandarin-speaking children, and our study made further progress on dataset building based on mean and standard deviation. For TTR, She et al. (2017) found that there was a decreasing

Table 5. Norm-referenced data for six categories of Chinese word classes.

1)	Group	M	SD	Min	Max
Noun/types	3–4 yr	30.83	9.463	10	55
	4–5 yr	33.90	10.178	15	68
	5–6 yr	38.95	12.978	15	86
Noun/tokens	3–4 yr	54.92	19.937	17	112
	4–5 yr	60.72	24.282	20	153
	5–6 yr	73.37	35.945	23	199
Verbs/types	3–4 yr	31.16	10.094	12	56
	4–5 yr	30.83	9.300	13	69
	5–6 yr	36.09	12.801	17	74
Verbs/tokens	3–4 yr	67.10	25.535	28	138
	4–5 yr	66.54	23.849	24	139
	5–6 yr	80.82	37.604	28	199
Pronouns/types	3–4 yr	8.97	2.615	3	17
	4–5 yr	8.74	2.574	4	16
	5–6 yr	9.34	2.972	3	17
Pronouns/tokens	3–4 yr	46.62	22.453	4	107
	4–5 yr	43.73	26.558	6	202
	5–6 yr	50.82	30.530	4	156
Adjectives/types	3–4 yr	8.96	3.604	2	23
	4–5 yr	9.83	3.628	3	19
	5–6 yr	9.99	4.414	3	23
Adjectives/tokens	3–4 yr	22.16	10.499	7	55
	4–5 yr	22.60	10.163	6	67
	5–6 yr	23.73	14.221	5	104
Adverbs/types	3–4 yr	10.57	4.293	2	23
	4–5 yr	12.56	4.188	2	28
	5–6 yr	14.25	6.019	5	31
Adverbs/tokens	3–4 yr	27.41	13.653	6	70
	4–5yr	31.04	13.890	10	90
	5–6yr	38.03	24.319	8	138
Measures/types	3–4yr	3.83	1.879	1	9
	4–5yr	4.01	1.763	0	9
	5–6 yr	5.04	2.468	1	13
Measures/tokens	3–4 yr	18.83	9.849	1	46
	4–5 yr	20.59	13.743	0	76
	5–6 yr	24.97	16.932	1	122
Conjunctions/types	3–4 yr	2.05	1.692	0	6
	4–5 yr	2.85	1.746	0	7
	5–6 yr	3.37	2.239	0	10
Conjunctions/tokens	3–4 yr	3.75	4.061	0	21
	4–5 yr	5.87	5.231	0	29
	5–6 yr	8.62	9.790	0	48

I)	Group	M	SD	Min	Max
Prepositions/types	3–4 yr	2.35	1.232	0	5
	4–5 yr	2.50	1.171	0	5
	5–6 yr	3.03	1.131	0	6
Prepositions/tokens	3–4 yr	6.88	6.534	0	39
	4–5 yr	6.85	5.272	0	27
	5–6 yr	10.92	9.270	0	44

trend when analyzing language samples from conversation, free play and story-retelling, indicating that TTR was not appropriate for lexical analysis in the context of Mandarin. After analysis of more samples, we found similar results for TTR and NDW, which also indicated cross-cultural similarity when using TTR and NDW, as both indicators lacked validity for representing the developmental trend of vocabulary skills.

However, it was not adequate to include vocD for vocabulary assessment, as Mandarin was quite different from English in word class, such as measures. In order to get the whole picture of vocabulary development, we need to further explore and analyze indicators that reflect the developmental trend of different word classes. Through analysis, we found that the types and tokens of nouns, verbs, measures, adverbs, conjunctions and prepositions gradually increased and enriched with age, while the types and tokens of pronouns and adjectives were not. The measures were more unique in Mandarin than other indicators. Those results respectively corresponded to the developmental studies of word classes by Chinese researchers. TX Zhang (2010) found that the frequencies of children's nouns, verbs and adverbs all showed an increasing trend with age. For measures, TX Zhang (2010) and Yang and Zhou (2010) found that the children produced more measures from three to six years old. The development of conjunctions was similar to Zhu's (1987) findings that three- to six-year-old children's use of conjunctions showed a significant age effect. The age effects of these six categories of word classes above was a basic requirement for datasets developed in the vocabulary assessment system.

Through the quantitative analysis of lexical diversity and word classes, we finally constructed a vocabulary assessment system for Mandarin-speaking children, which covered the first-level indicators: vocD and secondary indicators: the types and tokens of nouns, verbs, measures, prepositions, adverbs, and conjunctions. By comparing with the reference data of the first-level indicator, vocD, the basic status for children's vocabulary development could be predicted and estimated. When children significantly lagged behind that of normally developing children's referencing data in vocD, we could further refer to the secondary indicators of the vocabulary assessment system, the types and tokens of nouns, verbs, measure, prepositions, adverbs, conjunctions respectively, for further detailed assessment. This route for vocabulary assessment was as below (Figure 1).

There were definitely some limitations in this study. The dataset was built on language samples collected in context of free play, which may not track more complicated lexical usage as narrative tasks. In future, samples from different

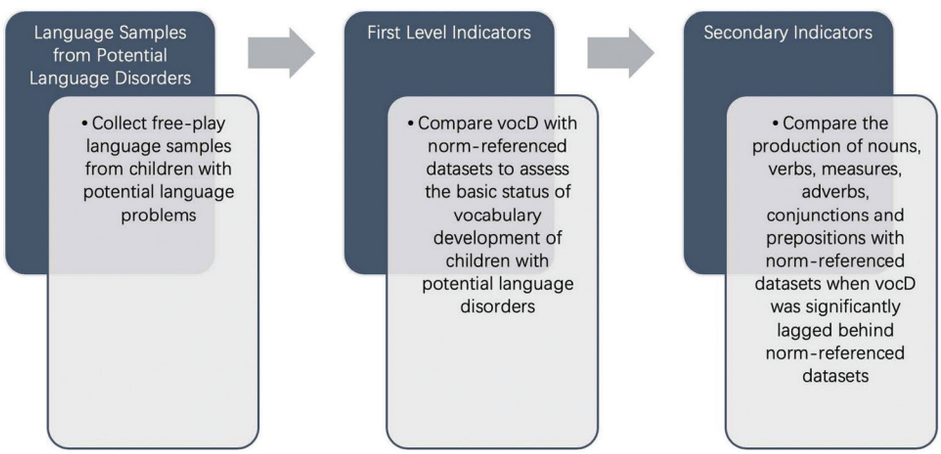

Figure 1. The route for vocabulary assessment for Mandarin-speaking children with potential language disorders.

contexts should be included and compared. In conclusion, this study was the first attempt to establish a norm-referenced dataset based on language sample analysis, a non-standardized approach, and we did build it up. This dataset could be used further for language assessment using KIDEval command in CLAN, offering a free and more automatic way to assessment children's language development based on computers, which would also save a lot of time in language samples analysis. We could compare data from children with potential language disorders with norm-referenced datasets to discover the specific vocabulary deficits, and therefore find a direction for later intervention in vocabulary.

Funding

This research was supported by the National Social Science Foundation (grant number 17BYY093).

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

References

Che Y (2009) *A comparative study on English and Chinese-speaking children's early acquisition of adverbs*. Masters Thesis, Changsha University of Science & Technology, China.

Chen GM (2009) *Early Childhood Psychology*. Beijing: Beijing Normal University Press.

Chen HW (2005) *Collecting language samples in clinical settings: A feasibility study*. Masters Thesis, National Taipei Health and Wellness University, Taiwan.

Erbaugh MS (1992) The acquisition of Mandarin. *Crosslinguistic Study of Language Acquisition* 3: 373–455.

Hao Y, Sheng L, Zhang Y, et al. (2018) A narrative evaluation of Mandarin-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research* 61(2): 1–15.

Jin XM and Jin ZJ (2008) The research on mean length of utterance and lexical. diversity in preschool Mandarin-speaking children. *Chinese Journal of Evidence-Based Pediatrics* (4): 261–266.

Kan PF and Windsor J (2010) Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research* 53(3): 739–756.

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

Kong LD, Hu DM, Ouyang JL, et al. (2004) *The Acquistion of Real words for Mandarin-Speaking Children*. Anhui: Anhui University Press.

Lai SA and Schwanenflugel PJ (2016) Validating the use of d for measuring lexical diversity in low-income kindergarten children. *Language Speech and Hearing Services in Schools* 47(3): 225–235.

Li K (2015) *The acquisition of the bei-construction in Mandarin-speaking children with specific language impairment*. Masters Thesis, Guangdong University of Foreign Studies, China.

Li L (2014) *Study on development of preschool children's narrative abilities in a Mandarin-speaking context*. PhD Thesis, Shanghai International Studies University, China.

Li XN, Zhou GG and Kong LD (1990) A pilot study on the use of ba sentences for children aged 2–5 years old. *Linguistic Researches* (4): 43–50.

Lin JY and Liu HM (2017) Expressive Mandarin vocabulary development in late talkers and typical language development children. *Bulletin of Special Education* 42(1): 27–50.

Liu Y (2016) A case study on the acquisition order and errors for six categories of real words in Mandarin. *Modern Chinese* (3): 141–143.

Liu YH, Pan WY and Gu H (2001) *Practical Modern Mandarin Grammar*. Beijing: Shangwu Press.

MacWhinney B (2000) *The CHILDES Project: Tools for Analyzing Talk. 3rd Edition*. Mahwah, NJ: Lawrence Erlbaum Associates. .

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

Ou LL (2009) *On English and Chinese-speaking children's early acquisition of nouns*. Masters Thesis, Changsha University of Science & Technology, China.

Owen AJ and Leonard LB (2002) Lexical diversity in the spontaneous speech of children with specific language impairment. *Journal of Speech, Language, and Hearing Research* 45: 927–937.

Peng XH (2009) On noun bias in children's early lexicon. *Journal of Hunan University (Social Sciences*) 23(6): 81–84.

Ratner NB and MacWhinney B (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.

Richards B (1987) Type/token ratios: What do they really tell us? *Journal of Child Language* 14(2): 201–209.

She L, Yang ZK, Liao YK, et al. (2017) The study of validity in language sampling indicators for Chinese children in Guangzhou. *Bulletin of Eastern-Taiwan Special Education* 106(19):129–155.

Silverman S and Ratner NB (2002) Measuring lexical diversity in children who stutter: Application of vocd. *Journal of Fluency Disorders* 27(4): 289-304.

Tardif TZ (1995) *Adult-to-child speech and language acquisition in Mandarin Chinese*. PhD Thesis, Yale University, USA.

Templin M (1957) *Certain Language Skills in Children: Their Development and Interrelationships*. Monograph Series No. 26. Minneapolis: University of Minnesota, Institute of Child Welfare.

Tian WW (2008) *Study on the acquisition of passive sentences*. Masters Thesis, Capital Normal University, China.

Weng CQ (2016) *Syntactic characeristics of preschool children in the context of antonomous expression context*. Masters Thesis, Shanxi Normal University, China.

Wong AM-Y, Klee T, Stokes SF, et al. (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.

Wright L, Pring T and Ebbels S (2018) Effectiveness of vocabulary intervention for older children with (developmental) language disorder. *International Journal of Language & Communication Disorders* 53(3): 480–494.

Xia Y (2008) *A longitudinal study of word class development in autistic children during parent–child interactions*. Masters Thesis, East China Normal University, China.

Yan XJ (2010) *A longitudinal study of the acquisitions of classifier in preschool chidlren*. Masters Thesis, University of Taipei, Taiwan.

Yang XL and Zhou J (2010) On the influence of maternal educational background on children's classifier using. *Early Childhood Education* (3): 43–48.

Zhang J (2013) *A comparative analysis of English and Chinese-speaking children's acquisition of degree adverbs*. Masters Thesis, Changsha University of Science & Technology, China.

Zhang TX (2010) *Lexical research based on a 3–6-year-old Mandarin-speaking children spoken corpus*. PhD Thesis, Shandong University, China.

Zhang WJ and Zhou J (2016) The development of the use of conjunctions in 4- to 6-year-old children's conversation. *Journal of Educational Development* (3): 23–27.

Zhang YB, Jiang ZX, Zhang HQ, et al. (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* (10): 19–25.

Zhang YW, Jin XM, Shen XM, et al. (2002) Effect of social and environmental factors on vocabulary and grammatical development in 2- to 3-year old children. *China Journal of Pediatrics* 40(11): 646–649.

Zhu MS (1987) *Study on Child Language Development*. Shanghai: East China Normal Universtiy Press.