4-2023

The Development of a Parent Report Instrument of Early Communication and Language Skills of Infants and Toddlers in Mainland China

Xueman Lucy Liu
Hainan Boao Bethel International Medical Center

Wendy Lee
Hainan Boao Bethel International Medical Center

Eric Rolfhus
Hainan Boao Bethel International Medical Center

Teresa Hutchings
Hainan Boao Bethel International Medical Center

Liqun Yao
Hainan Boao Bethel International Medical Center

See next page for additional authors

Liu, L., Lee, W., Rolfhus, E., Hutchins, T., Liqun Yao, Jinqiu Xie, Yaqing Xu, Yongmei Peng, & de Villiers. J.G., (2023) The Development of a Valid Parent Report Instrument of Early Communication and Language Skills of Infants and Toddlers in Mainland China. ***International Journal of Language and Communication Disorders,***

The Development of a Parent Report Instrument of Early Communication and Language Skills of Infants and Toddlers in Mainland China

What this paper adds

What is already known on the subject?

In China, efforts were made in recent years to develop language assessments for infants and toddlers, but limitations existed with the domains included and number of items included per age group. Many clinical practitioners also continued to rely on language subtests of general developmental scales, which were limited in depth and breadth of language skills tested and were never intended for diagnosis of language delay.

What this paper adds to existing knowledge?

This paper discusses the development of a valid caregiver report instrument for early communication and language skills of infants and toddlers in mainland China. The Diagnostic Receptive Expressive Assessment of Mandarin-Infant Toddler (DREAM-IT) includes foundational domains necessary for language and communication development in young children (receptive language, expressive language, cognitive play, and social communication domains).

The results show strong internal reliability (Cronbach's alpha) for each domain on a sample of 716 children sampled in 3-month age bands from 0 to 36 months. The external validity proved strong when tested on a group of 32 young children with Down syndrome.

What are the potential or actual clinical implications of this work?

Besides helping to inform the diagnosis of language delays in infants and toddlers in China, the caregiver report instrument has special features to support clinical practitioners in a field that is just emerging in China. The unique support features include the automatic generation of a profile of relative strengths and weaknesses of the child on the report and the recommendation of child-specific caregiver coaching videos on a companion app.

Background:

The global community and China, as a large, influential nation within the community, is making efforts to help our most vulnerable populations with disabilities beginning at the earliest age possible to prevent expanding and compounding negative consequences for individuals, families, local communities, provinces, and the nations at large. As part of the 2016-2030 Global Strategy for supporting children's health and development, the World Health Organization (WHO) and the United Nations Children's Fund (UNICEF) is promoting a Nurturing Care Framework. Besides early nutritional development and safe environment, early language and communication development is also a very key component in this preventative care framework (WHO, UNICEF & World Bank, 2018). China, in particular, has written national policies in their "Outline of

Healthy China 2030", the "Plan for Health in the 13th Five-Year Plan Period", and the "Plan for Deepening Medical and Healthcare Reform in the 13th Five Year Plan Period" to raise explicit requirements for ensuring healthcare for people with disabilities and for improving the ability of community-level institutions to deliver medical and rehabilitation services (Yang, 2020; Hacke, 2016; Tan et al., 2017; Chen et al., 2019). The "National Disability Prevention Action Plan (2016-2020)" has been promulgated across China, with measures taken to reduce and control the occurrence of disabilities (WHO, 2016). The Specifications for Disability Screening for Children Aged 0-6 Years has been formulated to realize early screening and treatment of children with five types of disabilities: hearing, vision, physical, mental, and autism (Huxia, 2019; Chen et al., 2019). Special attention has been paid to the health of children with disabilities. Following the principle of maximizing the interests of children, China is paying close attention to the health of children with disabilities. Priority has been given to preventative intervention and rehabilitation of children with disabilities aged 0-6, and in 2018, 157,000 children with disabilities aged 0-6 years were provided with basic rehabilitation services (Chen et al., 2019). It is exciting that more awareness of the efficacy of early identification and intervention is gaining attention in China and resulting in major policy changes. Besides early screenings, evidence-based language and communication assessments can assist the healthcare professionals to diagnose problems and then design individualized early intervention plans. This practice will then greatly further the initial efforts in early identification and early intervention for children with disabilities in language and communication.

Early language and communication delay is reported to have an incidence rate of between 13.5% and 17.5% for children in the first three years of life in the US and Europe (American Speech-

Language-Hearing Association [ASHA], 2022a; Tomblin et al., 1997; Norbury et al., 2016; Beitchman et al., 1986). Language issues have been identified as the most common disability of early childhood but are considered the least well detected disability (Prelock et al., 2008). It is important to understand that children not only struggle with the language impairment itself but are at greater risk than typically developing children for experiencing difficulties in academics, self-concept, peer relationships, and emotional well-being (Lindsay et al., 2010; Yew & O'Kearney, 2013, 2015; Charman et al., 2015).

It is important to note that while caregiver questionnaires have become widespread for evaluating child language, especially in the US and Europe (Fenson et al., 1993; Visser-Bochane et al., 2020), it is not a simple process to translate such a test for a different language and culture (Dale & Penfold, 2011). Strict considerations apply for the adaptation rather than the translation of such instruments, especially when the cultures and/or languages are significantly different. Some linguistic items will not translate at all, and broader questions about social interactions might be interpreted in a very different manner in another culture. For those reasons, new instruments are needed and new norming is required.

In the past couple of decades, a few communication and language assessments have been developed in China. The Putonghua MacArthur Bates Communication Development Inventories (PMCDI) (Tardif & Fletcher, 2008) was adapted from the US assessment MacArthur Bates Communication Development Inventories (MCDI) (Fenson et al., 2007). PMCDI focuses mostly on early vocabulary and was utilized in research more than clinical settings in China. Rather than just focusing on vocabulary understanding and vocabulary expression, a comprehensive

language assessment for infants and toddlers is particularly important to tap into other aspects of comprehension and expression that give a better representation of the various skills that develop within a young language learner. For instance, it would be important to know whether the child can follow different types of verbal directions involving multiple steps and increased levels of complexity. The ability to follow directions is critical for learning new skills and for the child to stay safe within the confines of rules. In addition, determining whether a child can ask or answer questions, or whether the child can describe the difference between objects, would also give insight into a child's functional use of language, using grammar as well as words, in toddlerhood and preschool. These finer grained receptive and expressive language skills are not assessed in vocabulary assessments.

The Infant and Child Language Development Screener (Zhang et al., 2003) was developed by a group of pediatricians and was normed in Shanghai, China. This assessment focuses on receptive and expressive language skills and was widely used in Chinese hospitals. The Infant and Child Language Development Screener was aimed to be used as a 15- to 20-minute language screener in a pediatric practice, therefore it has limited items per age band. The instrument provides age equivalent scores without standard scores. Therefore, it cannot serve as a diagnostic early language assessment or an assessment that provides clinicians assistance in individualized intervention.

Furthermore, in order to assess early language and communication comprehensively to assist diagnosis and individualized intervention, in addition to tapping relevant skills of language reception and expression at various stages of development, it is also imperative to assess social

and cognitive play skills, as these areas not only correlate with communication and language development but are intimately tied with the purposes and functions of early communication and language (Lewis et.al., 2000). Cognitive skills often develop for infants and toddlers in play situations. These cognitive skills in very young children manifest as different types of play, such as exploratory play, constructive play, symbolic and pretend play. Developing and using certain types of play skills is correlated with cognition and language ability. Symbolic and pretend play involves using objects to represent other objects and assigning functions to inanimate objects like dolls or stuffed animals. This type of early play has been found to be present in normally developing children, and less likely in children who demonstrate developmental delays (Short, et al., 2020; Lewis et al., 2000; Trawick-Smith, 2018). Research has shown an association between the beginning of pretend play and the beginning of vocabulary development, and the beginning of combining pretend play events with the onset of combining words (McCune, 1995). Similarly, an association between 14-month-old symbolic play ability, vocabulary production, and overall verbal expression at 2 years was also observed (Lyytinen et al., 1999). In another study, children who initiated more doll play or other-directed play acts at 13.5 months exhibited better language scores both concurrently and predictively at 22 months (Ungerer & Sigman, 1984). Findings from Tamis-LeMonda and Bornstein (1993, 1994) showed a positive correlation between symbolic play and language comprehension at 13 months and between symbolic play and semantic diversity at 20 months. As prerequisites to symbolic play, infants experiment with objects, recognize the variants in a category by learning to handle them, build with them, experiment with their properties, and use them in conventional ways.

Many preverbal social skills are foundational to language development and are observed prior to children's demonstration of language comprehension and use (Bruner, 1975; Baldwin, 1991; Carpenter et al., 1998; Bates, 2014; Moore et al., 2014). In one study, sharing attention at 8 months, following attention within visual field at 10-11 months, and following attention outside the visual field at 14 months showed a positive correlation with receptive vocabulary at 18 months (Beuker et al., 2013). As important as following the attention of others, those children who started initiating joint attention earlier also demonstrated greater receptive vocabulary acquisition between 10-15 months and greater expressive vocabulary at 14-18 months, which could then potentially also positively influence social skills and cognition (Beuker et al., 2013). Joint attention to an object or event with his/her caregiver is critical for the child to grasp the meaning of the language input that the caregiver provides. Likewise, when a child directs attention, he/she prompts the caregiver to engage and provide language input. In another study, it was found that when caregivers engage with their children in a responsive back and forth vocal social interaction, a child's production of syllabic, speech-like vocalizations are produced at a higher rate, paving the way for later language (Bloom et al., 2009).

Both early social and cognitive play skills are highly correlated with and to some degree intertwined with early language ability. Assessing early language and communication comprehensively and providing intervention for early language and communication warrants the inclusion of both domains in an assessment of early language ability in infants and toddlers.

The measure should be designed to be broader, in including all of these aspects that contribute to the child's communicative development. The goal is to create a more holistic picture of the

child's language development to identify potential weaknesses that could put the child as risk. In addition, by assessing these particulars, precise guidance can be offered for intervention or increased stimulation.

Having determined the importance of assessing crucial detailed skills in the areas of receptive, expressive, cognitive play, and social domains, another question to answer is how a communication and language assessment should be administered. When designing a communication/language assessment for infants and toddlers, many factors must be considered, including ease of administration, appropriateness for the population, and reliability & validity of the assessment (Xue et al., 2015). It is difficult to elicit specific communication behaviors from infants and toddlers to allow direct observation in a limited time frame typical of an efficient administration of a language assessment. Other plausible ways of assessing infant and toddler's language must be considered. At the younger ages when it is almost impossible to directly test an infant or toddler's language, reliance on caregiver report of child's ability is often necessary. Caregivers are most likely to have observed a child's behavior over time to enable them to make accurate, representative reports. In addition to being relatively easy to administer with relatively lower cost, use of caregiver report has also been shown to yield good reliability and validity. A recent study compared the parent-report screening tool ASQ-3 and the parent-report assessment tool MCDI against two direct observational assessments, the Preschool-Language Scales-4 (PLS-4) (Zimmerman, Steiner, & Pond 2002) and the Peabody Picture Vocabulary Test-4 (PPVT-4) (Dunn & Dunn, 1997). The study found the internal consistency reliability for both parent-report instruments was rated as acceptable to excellent (Xue et al., 2015). There was also moderate concurrent convergent and discriminant validity between the parent-reported ASQ-3 and MCDI.

While predictive validity of caregiver report was harder to achieve with the direct behavioral measures, particularly for infants younger than 12 months, there was some evidence of predictive validity; for instance, MCDI scores at age 2 significantly predicted PLS-4 and PPVT-4 scores at age 3 (Xue et al., 2015). In another study comparing direct language assessment with caregiver report at 24-26 months, the concurrent ability of caregivers judging language skills was high, and there was no difference between the direct language assessment and caregiver report when predicting language delay at age 3 (Sachse & Von Suchodoletz, 2008). In a meta-analysis of available language and behavioral screeners, Sim et al., (2019) found that caregiver report measures, for children 2 to 6 years, particularly Rescorla's Language Development Survey (Rescorla & Alley, 2001), had better specificity and sensitivity for detecting later language delay than behavioral measures. In addition, they found that surveys that tapped broader behaviors as well as language were more accurate than those that surveyed language or behaviors alone.

The assessment was designed to be a caregiver report measure for reasons highlighted in the review, namely, because caregivers can report reliable language information, such an assessment is easy and efficient to administer., and no other kind of assessment is yet available for the period of infancy. However, some positive modifications were added: in particular, a trained professional guide assisted with the administration of the caregiver-reported assessment. Caregivers could ask the professional administrator for clarification and further illustrations, which were prepared and standardized in advance. By including this assistance, we hoped to mitigate the effects of educational level of caregivers, which has previously been shown to impact raters' responses (Xue et al., 2015).

We describe the method by which we selected, piloted and normed the assessment in China, and detail how the tool might contribute to the detection and amelioration of risk for communication disorders in this young age group of children.

Method

Item Selection

A wide net was cast to find suitable items for the assessment. A good standard is to begin with twice the number of items desired in the final assessment. The research literature across four domains of development were consulted: receptive language, expressive language, early play, and social development. Parental questionnaires in other languages were studied to see how they might have tapped parental observation on the skills in question. The team of researchers assembled evidence of phenomena that are typical at different ages and discriminating of delay or disorder, consulting other instruments from studies mostly in English but also in Mandarin. Then we conducted multiple discussions with a team of experts including developmental psycholinguists, speech-language pathologists, experts on autism, and developmental psychologists to refine domain contents and items likely to cover the broad range from 0 to 36 months. Adaptations were made in consultation with experts on likely play and cultural activities in Chinese homes, such as what kinds of games caregivers play with babies, the topics of early conversations and typical communicative routines. Work on phonological development provided evidence of the order of consonant development in Mandarin (Zhu, 2002), and attention paid to the foods and toys with which young children were familiar. Ultimately, the choice of good

items comes not from experts' intuition but from empirical results, which is how the final set of items was decided.

The test items were designed to tap four domains relevant to early communication: *Receptive language, Expressive language, Cognitive play, and Social Behavior*. There are vast differences in terms of the content and skills that children command in each of these domains across even six-month spans of infancy and toddlerhood, but most existing work is from samples studied in the US and Europe. It was important not to make assumptions about how items will behave in the context of growing up in China, or in Mandarin and other languages spoken there. In making a new assessment, it seemed necessary to ask caregivers about *all* the items before we could be sure what the age span was for a given item.

However, there were practical limits to this. It did not seem reasonable to ask caregivers of a three-month-old what level of pretense their baby could engage in. Likewise, it seemed unnecessary to ask caregivers of a 34-month-old what sounds they could now babble. Child development is more similar than different across cultures. Nevertheless, it is an empirical question to determine the cutoff age for a given item to be considered unlikely. For most of the items intended for assessing babies, it seemed probable that the range for which the questions become irrelevant was between 12 and 24 months. Similarly, the age at which a broad range of typically developing children achieve more sophisticated communication and social skills could be as early as 12 months. Therefore, the first iteration for First Administration used two slightly overlapping sets of items: 191 items were administered to caregivers of children 0-24 months, and another 243 items to caregivers of children 12-36 months. The design required

oversampling, that is, doubling the sample of children aged 12-24 months where items were likely to either be at ceiling (for the infant items) or at floor for the items for older children. Most items were presented as statements in Mandarin to which a caregiver could reply “not yet” (还没有, 0), “sometimes” (有时, 1) or “always” (总是, 2). A few questions provided lists, for example of speech sounds heard from the child. Table 1 provides sample items for each of the four domains, from both the lower and upper item-sets.

Table 1. Sample Questions for 0.0 –17.9 months and 18.0 – 36.0 months.

Sample items for 0.0 – 17.9 Months.				
Domain	Question in Chinese	Question in English	Answer Choices in Chinese	Answer Choices in English
Receptive Language	您的的宝宝理解简单的问题(例如: 爸爸在哪儿？他是谁？）吗？	Does your baby understand simple questions (ex: where's daddy?, who's that?)	还没有 有时 总是	Not Yet Sometimes Always
Expressive Language	您的孩子会用语言来引导您对某物某事的注意吗？（“看”, “那儿”等	Does your child use words to direct your attention to something? (Look, there, etc.)	还没有 有时 总是	Not Yet Sometimes Always

Cognitive Play	您的孩子能用不同的方法玩一个玩具吗？例如，用一个堆高高的杯子喂娃娃喝水，或用一个长方形的积木当电话。	Does your baby play with a toy in different ways (ex: using stacking cup to give doll a drink, using small rectangular block as a phone, etc.) ?	还没有 有时 总是	Not Yet Sometimes Always
Social Interaction	您的宝宝会不会一直重复同样的动作或说同样的东西来引您发笑？（5比如，孩子会不会把玩具拿在头上然后故意让玩具掉下来，如果你笑了，他就重复做这个动作来逗你）。	Does your baby keep doing the same action or say the same thing to get you to laugh? (For example, does baby keep pushing a toy off his/her head and look at person to get/maintain a reaction?)	还没有 有时 总是	Not Yet Sometimes Always

Sample Questions for 18.0 – 36.0 months

Domain	Question in Chinese	Question in English	Answer Choices in Chinese	Answer Choices in English
Receptive Language	您的孩子明白 ”在床底下“，”在玩具盒里“，”在桌子上“这样的短语吗？	Does your child understand phrases like, 'on the shelf, under the bed, in the toybox, on the table?'	还没有 有时 总是	Not Yet Sometimes Always

Expressive Language	如果您和孩子去了某个地方买东西或出门去玩，您的孩子能谈论吗？(比如，"我们去了动物园"。)	When you go with your child on an errand, or a trip somewhere, can your child later talk about it e.g., 'go zoo'	还没有 有时 总是	Not Yet Sometimes Always
Cognitive Play	在玩过家家时，您的孩子会假装让娃娃/动物互相说话吗？（比如："小熊，你累了吗？""小兔，我不累。"）	During pretend play, he/she makes the dolls or animals talk to each other ex. Bear, are you tired? Rabbit, I am not tired.	还没有 有时 总是	Not Yet Sometimes Always
Social Interaction	当您伸手去拿某物时，您的孩子知道您要什么吗？他会不会试图把那样东西拿给你？	Does your child seem to know when you are reaching for something what you want? Will he/she try to give you that thing?	还没有 有时 总是	Not Yet Sometimes Always

To lessen the time burden on the caregiver, the items were divided into sections, the first question of which, if answered in the negative, would preclude the need to ask the rest of the questions in that section. For instance, if a caregiver answered the "trigger" question "Does your child speak *any* words" with "Not yet", then many other questions about the types of words spoken would not need to be asked. The trigger questions were all *logical*, that is, no assumptions were made about probable orderings of abilities.

Sample

We conducted an extensive pilot study to select the best items across ages, and to ensure that caregivers could understand the wording of the questions with a brief example. The sample was balanced for gender and divided into 3-month age bands from 0 to 36 months as shown in Table 2. Social class is another important variable that can affect child development, and in particular, the educational level of the caregiver (Liu et al., 2017; Chaney, 1994). We ensured that the sample was adequately representative of different groups of educational achievement. Social class was defined by the highest level of education achieved by the primary caregiver, as this has been shown to be the index most closely predictive of a child's language development (Hoff, 2006).

Table 2. Sociodemographic Characteristics of Respondents/Caregivers and their Children

Characteristic	Pilot Sample *N*=416 total		Norming Sample *N*=716 total	
Respondent/Caregiver Characteristics	**N**	**%**	**N**	**%**
Primary Caregiver Education Level	297*		376*	
Low	72	24.2	11	8.0
Medium	141	47.5	19	41.5
High	84	28.3	22	50.5
Primary Caregiver Relationship to Child	414*		247*	
Mother	304	73.4	134	54.3
Grandmother	58	14.0	86	34.8
Father	28	6.8	14	5.7
Grandfather	7	1.7	10	4.0
Other	17	4	3	1.2
Child Characteristics	**N**	**%**	**N**	**%**

Gender – Female	207*	49.8	519*	50.1
Birth Weight(grams)	409*		519*	
< 1500 (<3 lbs.)	1	0.2	1	0.2
1500 – 2267	16	3.9	5	1.0
2268 – 4081	379	92.7	496	95.6
> 4081	13	3.2	17	3.3
Delivery Time (weeks)	410*		519*	
25-27	0	0.0	0	0.0
28-30	3	0.7	1	0.2
31-33	9	2.2	1	0.2
34-36	36	8.8	21	4.0
37-38	76	18.5	165	31.8
39-40	220	53.7	292	56.3
41	49	12.0	39	7.5
42+	17	4.15	0	0.0

*Number of cases with non-missing demographic data for that element.

Sample Size	Pilot Sample			Norming sample			
Age Band	Female	Male	Total	Female	Male	Unknown	Total
0.0 - 5.9	27	25	52	24	22	3	49
6.0 - 11.9	20	25	45	19	17	1	37
12.0 - 17.9	55	56	111	71	69	1	141
18.0 - 23.9	51	46	97	57	69	1	127
24.0 - 29.9	26	28	54	68	64	0	132
30.0 - 36.0	27	25	52	111	119	0	230
> 36.0	1	4	5	–	–	–	–
Total	207	209	416	350	360	6	716

Caregivers participating in this study were chosen at random from a group of parents and others taking their children for regular “well baby checks” with their community doctors at a district level maternal and child community hospital in Chengdu, China. The group of children whose caregivers participated were neither included or excluded based on ASD, ID, global delay or other diagnoses or suspicion thereof. Caregivers of children from preschools and other maternal and child community hospitals in the region that have collaborative relationships with the district

level maternal and child community hospital were also recruited for age groups >= 2 years old. All caregivers were informed about the nature of the research and given the free choice to participate or not, provided signed informed consent, and ethical clearance was provided by the committees at the participating hospital.

The sampling plan for the data collection was to seek questionnaires from ~25 caregivers of children for every 3-month interval, between 0-12 months, ~50 caregivers of children for every 3-month interval between 12-24 months, and ~25 caregivers for every 3 months between 24-36 months. In sampling, we attempted to collect twice the number of children from 12-24 months because the greatest growth of language and communication skills occurs in this age range and also because we wanted to test the same aged children on different questionnaire forms (one for 0-17 month skills and one for 18-36 month skills) to determine what the appropriate break point was for each questionnaire. A total of 416 children were tested using the questionnaires in the pilot testing. Gender of the children was balanced.

Since the educational level of the primary caregiver has been shown to be the index most closely predictive of a child's language development, we ensured that the sample was adequately representative of different groups of educational achievement (Hoff, 2006). Table 2 provides a chart showing the distribution by education of the caregiver. The lowest level of primary education included Primary to Junior High School, the middle level included High School to Associate Degree education, and the highest level included a bachelor's degree or above.

After the piloting, an analysis of the item level data was conducted and those items that were most discriminating were retained for the final form. Approximately 6 months later 193 of the

children from the previous administration returned to complete the questionnaire a second time. These items were then arranged into 1 form for each of the six, 6-month age bands. 44 additional children were recruited in the 0-6 month age range and 149 additional children were recruited throughout the age ranges to increase the total number of participants whose caregivers received the questionnaire a single time. Data from the 6-month interval repeat administration from the pilot caregivers was not used for norming, though it was used to check that systematic growth occurred across the items.

To broaden the geographic distribution for the norming sample, a further group of caregivers were recruited from a different urban area, through a city-level maternal and child hospital in Shanghai. Four different district sites were included in the recruitment. Again, all caregivers gave signed informed consent and the study was approved by the hospital's ethics committee.

The final normative sample was comprised of N=716 first- time administrations of the questionnaire to caregivers.

Procedure

At the stage of piloting, the items were programmed as individual fields in FileMaker Pro on an iPad held by the Examiner as s/he questioned the caregiver. Caregivers were invited to participate during their "well baby checkup", and testing took place in a designated testing room in the health center. If a caregiver did not understand a question and asked for clarification, the Examiner had a prepared list of possible examples to add context to the prompt. All examiners

received a training conducted by American Speech-Language Hearing Association certified bilingual speech-language pathologists to go through all items and prompt examples. Before answering communication specific items, caregivers received a detailed health questionnaire with questions covering prenatal health, infant and child health, hearing status, and prematurity, as well as detailed demographic questions about languages spoken in the home, caregiver education, and so forth.

Before the second administration to the pilot group, and subsequent testing of the norming sample, data analysis was conducted to identify and eliminate less discriminating and redundant items that did not contribute to scale reliability at each 6-month age-band. As shown in Table 3, this process allowed us to reduce the number of candidate items by over 50% from the initial pool for the final item-set.

After piloting, we made a change by providing checklists for types of lexical items. In piloting, the tester would ask, for example, “How many names of foods does your child say?”, and then prompt if the caregiver could not think of any. This was a useful step in piloting for choosing items, but it introduced too much variability and relied on caregiver free recall. To avoid unreliability, we decided to provide a uniform checklist that caregivers could use for such items. The opportunity arose for caregivers to add additional examples if they were not on checklist. As before, the examiner recorded responses onto the iPad but shared the view of the checklist with the caregiver to respond, reading the words as well.

Results

We sought to develop an age-appropriate item-set for each six-month age-band using a classical test theory approach. We selected items within each set to achieve the following characteristics: (1) item-means ranging from 0 to 2.0 to reflect the full range of "difficulty" within the set, (2) all items exhibited item-total correlations above 0.30 with the raw total score [except for items with very low or high mean scores due to restriction of range], (3) item-set total raw scores exhibited normal distributions without floor or ceiling restrictions, and (4) removal of items that were redundant in terms of mean endorsement (i.e. difficulty). In order to cross-validate decisions, Rasch modeling was conducted on candidate (i.e., interim) item sets constructed for each 6-month band to check fit statistics.

After the final data collection, items were examined within each of the four scales in terms of item-total correlations and mean caregiver ratings (endorsement). The items with lower average caregiver endorsement rates comprising the lower-level questionnaire showed good item-total correlations (i.e., typically > .30) until approximately 18 months. The items with high endorsement rates from the upper-level questionnaire performed well down to about 18 months. Items which demonstrated similar caregiver endorsement-rates within the same domain, or for which caregivers had difficulty understanding the question, were eliminated. Table 3 shows the initial number of items tested in Piloting, and the number of items retained in the final instrument. Most retained items were used in more than one band.

Table 3. Item counts for initial administration and retained in final instrument.

Age Band (months)	0.0-5.9	6.0-11.9	12.0-17.9	18.0-23.9	24.0-29.9	30.0-36.0
Initial Total	186	201	202	248	247	247
Final Total	68	91	88	89	96	98
Receptive	13	18	24	16	22	20

Expressive	17	24	27	32	27	28
Play	18	25	17	17	23	20
Social	19	18	16	22	22	28
Additional	1	6	4	2	2	2

Scaling and Norming

In the Pilot we experimented with different ranges of item-level scores (e.g., 0-4, 0-6). In the final instrument each item is scored 0, 1, or 2 points. Item-level data were transformed into item-response theory (IRT) ability scores using the generalized partial credit model in Stata version 17. There was extensive overlap of items within adjacent 6-month bands (e.g., 0.0-5.9 and 6.0-11.9), allowing scaling to be done within the 18-month lower and upper item-sets separately. As the lower (0.0-17.9 months) and upper (18.0 -36.0 months) item sets were mostly different, and there were not enough common items between the upper and lower sets to scale all ages together, ability scores were estimated separately for the lower and upper age-bands. To align the lower and upper sets, equipercentile equating was used to align the 12.0-17.9 month and 18.0-23.9 month bands.

Standardized assessments often present normative information in the form of age-based scaled scores based on an arbitrary scale (e.g., a mean of 100, standard deviation of 15). Caregivers often find this difficult to interpret. To facilitate interpretation by medical practitioners in China, and communication with caregivers, we elected to develop normative growth curves from ages 0.0 to 36.0 months. Growth charts have a long history of use in medicine dating from the 18th century (Cole, 2012), and charting height and weight growth are a regular part of pediatrician's communication with caregivers. This representation of growth over time fits with our recommendations for use of DREAM-IT, and how results can be reported. DREAM-IT can be

completed regularly during periodic wellness checks, and results from multiple administrations plotted over time relative to normative growth curves to track developmental trends.

Once ability scores were placed on the same scale, the normative data set was processed by RefCurve (Winkler et al., 2020) software to create smoothed percentile growth curves. These curves were based on each student's age in decimal months and their IRT ability score. Although RefCurve allows the user to adjust the smoothing parameters (i.e., sensitivity to the moments of the raw data), little adjustment was done as the distribution of ability scores and the resulting percentile growth curves increased smoothly and monotonically (i.e., examined through visual inspection) as expected across the entire age-range.

Internal Validity

As evidence of the internal validity of each scale, Cronbach's Alpha was calculated for each domain scale overall and by six-month age-band (Table 4). Only cases used for norming are included in this analysis.

Table 4. Internal consistency reliability by domain.

Age Band (months)	0.0-5.9	6.0-11.9	12.0-17.9	18.0-23.9	24.0-29.9	30.0-36.0
Receptive	.79	.92	.91	.88	.91	.88
Expressive	.81	.85	.91	.96	.97	.95
Play	.86	.93	.79	.80	.90	.91
Social	.83	.85	.73	.86	.93	.94

Note: Nearly all alphas (22 of 24) are above 0.80, ranging from .73 to .97. This is evidence of excellent internal consistency of the four domains.

To examine the relationship among the four domains, correlations among the scales for the lower and upper items sets are provided in Table 5. Each band includes three 6-month item-sets and a

range of student ages of 18 months. Again, only cases used for norming are included in this analysis.

Table 5. Scale intercorrelations based on ability scores for lower and upper age-bands.

Lower Band (0.0 - 17.9, N=227)	Receptive	Expressive	Play	Social
Receptive	1.0	–	–	–
Expressive	.75	1.0	–	–
Play	.82	.77	1.0	–
Social	.84	.82	.93	1.0
Upper Band (18.0 – 36.0, N=488)	**Receptive**	**Expressive**	**Play**	**Social**
Receptive	1.0	–	–	–
Expressive	.54	1.0	–	–
Play	.62	.89	1.0	–
Social	.57	.95	.91	1.0

*All correlations are statistically significant at $p < .05$.

We anticipated that the scales would exhibit moderate to high intercorrelations – which are observed in the range 0.54 - 0.95. The rather high correlation ($r = 0.93$) between Social and Play in the lower band is likely due to the dependence of the assessed Play behaviors on Social skills. In the upper band many of the Social and Play behaviors assessed require expressive language skills – and this is observed in the high correlations each domain has (*rs* = .89 and .91 respectively) with Expressive skills. We also observe that the correlations between the Receptive and other domains are not as high in the upper band. Individual differences in developmental trajectories across the four domains are more often observed and salient at older ages.

External Validity – Down Syndrome Sample

The external validity of a caregiver report instrument for this age range is difficult to assess. Children with language impairments are not usually identified as having delays or disorders until

later, so it is not feasible to select a clinical group based on independent criteria and compare their caregiver reports to those of a typically developing group. One alternative is to take a clinical group that has a high probability of developmental delays and can be identified by some marker independent of those skills. Children with Down syndrome fit this description. Down syndrome is a chromosomal difference identified by many clear signs at birth or in utero (Bull, 2020), and though behavioral outcomes vary enormously (Bull, 2020; d'Souza et al., 2017), on average the children have significant developmental delays in all the domains assessed by the current measure: receptive, expressive, cognitive play and social communication (Chapman & Hesketh, 2000; Yoder & Warren, 2004). Some studies have found that receptive language is less affected than expressive language for children with Down syndrome, however receptive skills are often below those of typically functioning children. The evidence on social communication skills such as joint attention for this group of children is mixed (Seager et al., 2018). For these reasons, the sample chosen for comparison was children with Down syndrome between ages 19 and 36 months, when some language skills typically emerge.

Table 6. Sociodemographic description of the Down Syndrome sample

Respondent/Caregiver Characteristics		
Primary Caregiver Education Highest Level	*n*	%
Primary school	6	21.4%
Junior High School	9	32.4%
High School Graduate	4	14.3%
Associates degree	5	17.9%
Bachelor's degree	3	10.7%
M.S. or Ph.D.	1	3.6%
Unreported	4	-

Primary Caregiver Relationship to Child	*n*	%
Mother	14	70.0%
Grandmother	5	25.0%
Father	1	5.0%
Grandfather	0	0.0%
Unreported	11	-

Child Characteristics	*n*	%
Gender – Male	23	71.8%

Birth Weight (grams)	*n*	%
< 1500 (<3 lbs.)	2	9.1%
1500 – 2267	0	0.0%
2268 – 4081	19	86.4%
> 4081	1	4.5%
Unreported	10	-

Delivery Time (weeks)	*n*	%
31-33	1	3.1%
34-36	2	6.3%
37-38	9	28.1%
39-40	9	28.1%
41	2	6.3%
42+	9	28.1%

Age Band	Female *n*	Male *n*	Total *n*
18.0-20.99	2	0	2
21.0-23.99	1	1	2
24.0-26.99	1	2	3
27.0-29.99	3	5	8
30.0-32.99	1	8	9
33.0-35.99	0	8	8
Total	8	24	32

Table 7 provides the mean percentile rank (standard deviation and range) of the clinical sample of children with Down syndrome for each domain. This is already age-adjusted, as the percentile growth curve norms are inherently age-adjusted. In addition, propensity-score matching was used to create a typically performing comparison group based on age in months, caregiver education level and gender. The standardized effect size was calculated on differences in ability scores between the identified children with Down syndrome and their community sample-matched comparisons.

Table 7. Dream-IT mean percentile rank by domain, and matched-sample differences.

Domain	Down Syndrome sample percentile rank **Mean** (*sd*)	Range of percentile scores (min – max)	Difference in IRT ability scores between matching Down syndrome and typically developing group expressed as effect size[a].
Receptive	**10.2** (16.0)	(1 - 82)	z = 10.51 (p<.000); *d'* = 1.05
Expressive	**4.7** (8.3)	(1 - 34)	z = 18.69 (p<.000); *d'* = 1.86
Play	**1.4** (0.9)	(1 - 4)	z = 16.84 (p<.000); *d'* = 1.68
Social	**1.7** (2.1)	(1 - 13)	z = 13.24 (p<.000); *d'* = 1.32

d' = standardized effect-size difference, expressed as Cohen's *d* - typically developing children scoring higher.
[a] This analysis used propensity scores to match Down syndrome cases with typical cases on caregiver education, gender and age in months.

Predictive Validity

Another form of external validity is whether the DREAM-IT performance at one time exhibits predictive validity to DREAM-IT scores of typically developing children at a future point in

time. To establish predictive validity of the DREAM-IT, a subset of caregivers who completed the instrument in the Pilot data-collection were asked to return approximately 6-7 months later to complete a 'retest.', using only the common items that were retained. N=107 caregivers did so with a mean interval of 6.7 months. These ratings showed high predictive validity, correlations between IRT ability scores were high and remarkably consistent across domains: Receptive r = .82, Expressive r = .82, Social r = .84, Play r = .82. Therefore, in a community sample of caregivers returning nearly 7-months later, rank order of children on each of these domains was primarily maintained.

Discussion

This paper describes the development of a novel instrument: a clinician-administered caregiver report for children in Mandarin-speaking households in China. Caregivers acted as in-home experts to rate their observations of their child's behaviors in four domains known to be associated with typical development of language. One interesting question is whether these caregivers in China report similar developments at the same ages as have been reported in samples collected elsewhere, in English-speaking countries. This is not an easy question to answer given two cautionary notes: first, that adaptation of questionnaires both cross-linguistically and cross-culturally is never exact, and second, that the holistic assessment described here, that includes four domains, is not well matched to any existing instruments. The closest equivalent study is that of Reilly et al. (2007), describing a large sample of English-speaking children (N=1720) in the province of Victoria, Australia, studied at three time points: 8, 12 and 24 months. For parental report measures they added the Communication and Symbolic

Behavior Scales (Wetherby & Prizant, 2002) to the MacArthur -Bates CDI (Fenson et al., 1993), though by 24 months the CBCS scores were at ceiling.

Despite these difficulties of comparison, we can consider the broad milestones that have been reported before for children outside of China. The age of initial babbling, first words, and early sentences, the understanding of basic questions, as well as the onset of pretend play and dyadic social attention, were roughly as expected given norms from other countries such as Australia Reilly et al., 2009), the US (ASHA, 2022b, 2022c; National Institute of Deafness and Other Communication Disorders [NIDCD], 2022; Fenson, et al., 1993) and the Netherlands (Visser-Bochane et al., 2020). Usually the data in such reports are reported as falling with a limited window, as broad as 12-24 months (Reilly et al.,2007) but sometimes matching the ranges in this study (e.g., 9-12 months, 21-24 months.) In general, the rough ages are confirmed in our sample in China for these general developments.

However, one must be wary of comparing broad milestones across cultures. For example, in the Visser-Bochane et al. study in the Netherlands, there is a window of 3 months (from less than 12 months to 15 months) provided for the supposed milestone “understands three word sentences”. However, the fact is that it can make a huge difference what the content of those sentences are, and even the definition of a word can vary across languages (for extensive discussion, see Allen & Dench, 2015). It is more valuable to get the caregivers’ response to more specific descriptions in an item for which the score of a child can be then counted up for comparison to their peers. Failing to achieve broad milestones is indeed a red flag, but it is insufficient for diagnosis and intervention. Consider an analogy with other areas, such as physical development. A child might

walk on target at 13 months, but a pediatrician or therapist might note if they walk favoring one leg, or on their toes, or bow-legged. Language is even more precise and deserves careful reporting. For example, in expressive language, “Baobao (the nickname parents call the child wants bubbles”, “I want bubbles”, and “Mom wants bubbles” are all “three-word sentences” but reflect three different levels of language and pragmatics skills. Only by attending to specific achievements will we achieve diagnostic tools that go beyond red flags to instruments with the appropriate psychometric properties: sufficient validity, reliability, specificity and sensitivity.

A large pool of these behavioral observation questions was developed after an exhaustive review of appropriate measures developed for other language contexts. An initial wave of assessment enabled redundant or inappropriate items to be removed and developmentally appropriate item-sets developed for each 6-month band, for each domain, without sacrificing scale reliability. All domains at each 6-month band exhibit high-levels of internal consistency, after reducing the items sets by more than 50%.

Due to the early state of knowledge for identifying atypical language development in China, we did not have the opportunity to validate the DREAM-IT externally using well-specified diagnostic categories of language delay as exist in the United States. However, the initial study of children diagnosed as having Down syndrome reveals the expected pattern of results on the DREAM-IT: Significantly lower scores on age-appropriate Receptive language behaviors, and very little age-appropriate behavior reported for the Expressive, Play, and Social domains. These data are compatible with other reports on children with Down syndrome in other languages and cultures (Yoder & Warren, 2004; Seager et al., 2018).

The study, diagnosis, and treatment of language delays and disorders in China is a relatively new field, and those practicing in a clinical capacity to diagnose are often rehabilitation generalists with a general but limited knowledge of speech-language pathology. Doctors, nurses, or teachers may have to assume an additional role in the diagnosis and treatment of language disorders because of a lack of already trained and specialized staff in the institution. Because the base of professionals is still mostly developing, it was important to design and integrate special features to support the intervention that must occur following the test administration. In the case of the DREAM-IT assessment, one such feature is the automatic generation of a profile of relative strengths and weaknesses in the skills of the child, based on the assessment's proprietary algorithms developed from analysis of related groups of skills. For instance, the assessment could be used to identify that an older infant has difficulties in knowing how to use varied objects. Likewise, the test could be used to as an indication that a two-year-old has considerable difficulty in using language about simple emotions and desires in self and others. Identifying these areas can help professionals know which skills to further probe and then target.

One of the purposes of developing an early language assessment is to facilitate early identification and early intervention, which is in line with the Nurturing Care initiatives in China. Early exploration for an implementation model was made with an early language screening paired with a caregiver training course (Camarata et al., 2022). As a caregiver report type of assessment, Dream-IT can be administered on site or online with a caregiver. Caregiver-implemented communication and language intervention for young children under the guidance of professionals has been found an evidence-based and cost-effective means of assisting children, families, and communities overcome communication and/or language deficits (Baxendale &

Hesketh, 2003, Roberts & Kaiser, 2011) so the assessment that can be delivered online and can be directly pair with caregiver intervention resources through a parent app[1] may be an effective strategy for preventative intervention and early intervention, especially where professional resources are limited.

The current assessment is designed to be comprehensive, with large numbers of items targeting very specific easily observed skills in four different domains of language and early communication. It provides normative scores in 6-month age ranges. Nevertheless, it is a parent report measure, and so to further ensure the proper diagnosis of a child's language difficulties, it is recommended that test administrators also collect three samples of ~5-minute video clips of the child engaged in various activities with caregivers, such as reading a book together, completing a puzzle or slightly challenging task, spontaneously playing together, and sharing a snack. These video clips can be analyzed as language samples to further assess communication in areas such as length of utterance, number of different words used, complexity of syntax used, imitation skills, and ability to follow directions (Prath, 2018) and to cross-check and supplement caregiver's reported responses. It is very challenging to make a comprehensive, norm-referenced, behavioral language instrument for children in this age range. It is common practice to use

[1] In order to provide more individualized parent coaching for children with a language delay, the development team has also paired key skill areas with specially designed short instructional caregiver coaching videos to help both clinicians and caregivers learn how to effectively use evidence-based strategies to target specific goal areas identified for a specific child. These are available through companion software, the Bethel parent app (Bethel HSTC, 2020). The use of the coaching videos to help caregivers learn and use strategies in effective caregiving is in line with the World Health Organization (WHO) and the United Nations Children's Fund (UNICEF) endorsement of a Nurturing Care Framework (WHO, UNICEF & World Bank, 2018), as part of their 2016-2030 Global Strategy to support children's health and development through training to increase nurturing parenting (Britto et al., 2017, 2018).

informal behavioral observation as a qualitative measure to enhance a quantitative parent questionnaire assessment.

The DSM-V diagnostic standard did not limit the Language Disorder diagnosis to children after kindergarten ages. The diagnostic criteria only cautioned the clinicians to use the Language Disorders diagnosis carefully with children younger than 4 years of age due to the fact that the language delays observed in very young children will become stabilized after 4 years of age (APA, 2013; Rice, 2014). Future research will be needed to trace children tested with the DREAM-IT instrument over time, as the differentiation of "late talkers" and persistent language delay is a significant challenge still being addressed in US and European research despite a much longer tradition of diagnosis of language delay in children (Leonard, 2014). Nevertheless, early diagnosis of language delay is important for the reason that early intervention is a crucial factor for good prognosis.

References

Allen, S. E., & Dench, C. (2015). Calculating mean length of utterance for eastern Canadian Inuktitut. *First language*, *35*(4-5), 377-406.

American Academy of Pediatrics. (2001). Committee on Children with Disabilities. Developmental surveillance and screening of infants and young children. *Pediatrics*, *108*, 192-196.

American Psychiatric Association, D., & American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (Vol. 5, No. 5). Washington, DC: American psychiatric association.

American Speech-Language-Hearing Association. (2022a, June 23). *The practice portal.* Retrieved June 23, 2022, from https://www.asha.org/practice-portal/

American Speech-Language-Hearing Association. (2022b, June 23). *How does your child hear and talk?* Retrieved June 23, 2022, from https://www.asha.org/public/speech/development/chart/

American Speech-Language-Hearing Association. (2022c, June 23). *Developmental norms for speech and language*. American Speech-Language-Hearing Association. Retrieved June 23, 2022, from https://www.asha.org/slp/schools/prof-consult/norms/

Bates, E. (2014). *The emergence of symbols: Cognition and communication in infancy*. Cambridge, MA: Academic Press.

Baxendale, J., & Hesketh, A. (2003). Comparison of the effectiveness of the Hanen Parent Programme and traditional clinic therapy. *International Journal of Language & Communication Disorders*, *38*(4), 397-415.

Beitchman, J. H., Nair, R., Clegg, M., Patel, P.G., Ferguson, B., Pressman, E. & Smith, A. (1986). Prevalence of speech and language disorders in 5-year-old kindergarten children in the Ottawa-Carleton region. *Journal of Speech and Hearing Disorders, 51*(2), 98-110.

Bethel Hearing & Speaking Training Center. (2020). *Bethel Parent Application.* Bethel Hearing and Speaking Training Center: Dallas, TX.

Beuker, K. T., Rommelse, N. N., Donders, R., & Buitelaar, J. K. (2013). Development of early communication skills in the first two years of life. *Infant Behavior and Development*, *36*(1), 71-83.

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of child language*, *14*(2), 211-227.

Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., ... & Lancet Early Childhood Development Series Steering Committee. (2017). Nurturing care: promoting early childhood development. *The Lancet*, *389*(10064), 91-102.

Britto, P. R., Singh, M., Dua, T., Kaur, R., & Yousafzai, A. K. (2018). What implementation evidence matters: scaling‐up nurturing interventions that promote early childhood development. *Annals of the New York Academy of Sciences*, *1419*(1), 5-16.

Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of child language*, *2*(1), 1-19.

Bull, M. J. (2020). Down syndrome. *New England Journal of Medicine*, *382*(24), 2344-2352.

Camarata, S., Liu, L., Lee, W., Li, T., Jiang, F., & Simms, M. (2022). Adapting the UNICEF/WHO NurturingCare Framework for speech-language pathologists to support Sustainable Development Goal 4. *International Journal of Speech-Language Pathology*. https://doi.org/10.1080/17549507.2022.2141327

Carpenter, M., Nagell, K., Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development*, *63(4)*, i-iv, 1-143.

Chaney, C. (1994). Language development, metalinguistic awareness, and emergent literacy skills of 3-year-old children in relation to social class. *Applied psycholinguistics*, *15*(3), 371-394.

Chapman, R. S., & Hesketh, L. J. (2000). Behavioral phenotype of individuals with Down syndrome. *Mental retardation and developmental disabilities research reviews*, *6*(2), 84-95.

Charman, T., Ricketts, J., Dockrell, J. E., Lindsay, G., & Palikara, O. (2015). Emotional and behavioural problems in children with language impairments and children with autism spectrum disorders. *International Journal of Language & Communication Disorders*, *50*(1), 84-93.

Chen, P., Li, F., & Harmer, P. (2019). Healthy China 2030: moving from blueprint to action with a new focus on public health. *The Lancet Public Health*, *4*(9), e447. Chinese translation: https://nejmqianyan.cn/article/YXQYra1706537?sg=AbW1NGsHw3NxPd6F

Cole T. J. (2012). The development of growth references and growth charts. *Annals of human biology*, *39*(5), 382–394. https://doi.org/10.3109/03014460.2012.694475

Dale, P. S., & Penfold, M. (2011). *Adaptations of the MacArthur-Bates CDI into non-US English languages.* Retrieved June 23, 2022 from http://mb-cdi.stanford.edu/documents/AdaptationsSurvey7-5-11Web.pdf

d'Souza, D., d'Souza, H., & Karmiloff-Smith, A. (2017). Precursors to language development in typically and atypically developing infants and toddlers: The importance of embracing complexity. *Journal of Child Language*, *44*(3), 591-627.

Dunn, L. M., & Dunn, L. M. (1997). Peabody Picture Vocabulary Test— Third Edition. Circle Pines, MN: American Guidance Service.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., & Reznick, J. S. (2007). *MacArthur-Bates CDI (Communication Developmental Inventory) Words and Sentences.* Baltimore, MD: Paul H. Brookes Publishing Company.

Fenson, L., Dale P.S. & Reznick J.S. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual.* San Diego, CA: Singular Publishing Group.

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry*, *54*(3), 280-290.

Haacke, O. (2016, March 18). Official 13th five-Year plan outline released. US. Retrieved June 27, 2022, from https://www.uschina.org/policy/official-13th-five-year-plan-outline-released

Hoff, E. (2006). How social contexts support and shape language development. *Developmental review*, *26*(1), 55-88.

Huxia (2019). *China makes headway in disability prevention.* Xinhuanet. http://www.xinhuanet.com/english/2019-07/25/c_138256819.htmt

Leonard, L.B. (2014). *Children with Specific Language Impairment (2nd ed).* Cambridge, MA: MIT Press.

Lewis, P, Boucher, J., Lupton, L. & Watson, S. (2000) Relationships between symbolic play, functional play, verbal and non-verbal ability in young children. *International Journal of Communication Disorders*, *35*(*1*):117-127.

Lindsay, G., Dockrell, J., Desforges, M., Law, J., & Peacey, N. (2010). Meeting the needs of children and young people with speech, language and communication difficulties. *International Journal of Language & Communication Disorders*, *45*(4), 448-460.

Liu, X., de Villiers, J., Ning, C-Y., Rolfhus, E., Lee, W., Hutchings, T., Hsiang, F. & Ziang, Y. (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.

Lyytinen, P., Laakso, M. L., Poikkeus, A. M., & Rita, N. (1999). The development and predictive relations of play and language across the second year. *Scandinavian Journal of Psychology*, *40*(3), 177-186.

McCune, L. (1995). A normative study of representational play in the transition to language. *Developmental psychology*, *31*(2), 198-206.

The Mobile Economy China 2021 - GSMA. (2021, February). Retrieved June 27, 2022, from https://www.gsma.com/mobileeconomy/wpcontent/uploads/2021/02/GSMA_MobileEconomy2021_China_Eng-1.pdf

Moore, C., Dunham, P. J., & Dunham, P. (Eds.). (2014). *Joint attention: Its origins and role in development*. New York , NY: Psychology Press.

National Institute of Deafness and Other Communication Disorders. (2022, June 23). *Speech and language developmental milestones*. Retrieved June 23, 2022, from https://www.nidcd.nih.gov/health/speech-and-language

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., ... & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of child psychology and psychiatry*, *57*(11), 1247-1257.

Prath, S. (2018). The How and Why of Collecting a Language Sample. Retrieved July 2nd, 2022 from https://leader.pubs.asha.org/do/10.1044/the-how-and-why-of-collecting-a-language-sample/full/

Prelock, P. A., Hutchins, T., & Glascoe, F. P. (2008). Speech-language impairment: How to identify the most common and least diagnosed disability of childhood. *The Medscape Journal of Medicine*, *10*(6), 136.

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., ... & Ukoumunne, O. C. (2007). Predicting language at 2 years of age: a prospective community study. *Pediatrics*, *120*(6), e1441-e1449.

Rescorla, L., & Alley, A. (2001). Validation of the language development survey (LDS). *Journal of Speech, Language, and Hearing Research, 44(2), 424-445*.

Rice, M. (2014, November). *Navigating the diagnostic complexities of DSM-5 language and communication disorders: Keep your eye on SLI*. Oral seminar at the American Speech-Language Hearing Association Annual Convention in Orlando, FL.

Roberts, M. Y., & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech-Language Pathology*, *20*(3), 180-199.

Sachse, S., & Von Suchodoletz, W. (2008). Early identification of language delay by direct language assessment or parent report? *Journal of Developmental & Behavioral Pediatrics*, *29*(1), 34-41.

Seager, E., Mason-Apps, E., Stojanovik, V., Norbury, C., Bozicevic, L., & Murray, L. (2018). How do maternal interaction style and joint attention relate to language development in infants with Down syndrome and typically developing infants? *Research in Developmental Disabilities*, *83*, 194-205.

Short, E., Schindler, R., Obeid, R., Noeder, M., Hlavaty, L., Gross, S., Lewis, B., Russ, S., and Manos, M., (2020). Examining the role of language in play among children with and without developmental disabilities. https://doi.org/10.1044/2020_LSHSS-19-00084

Sim, F., Thompson, L., Marryat, L., Ramparsad, N., & Wilson, P. (2019). Predictive validity of preschool screening tools for language and behavioural difficulties: A PRISMA systematic review. *Plos one*, *14*(2), e0211409.

Tamis-LeMonda, C. S., & Bornstein, M. H. (1993). Play and its relations to other mental functions in the child. *New Directions for Child and Adolescent Development*, *1993*(59), 17-28.

Tamis-LeMonda, C. S., & Bornstein, M. H. (1994). Specificity in mother-toddler language-play relations across the second year. *Developmental psychology*, *30*(2), 283.

Tan, X., Liu, X., & Shao, H. (2017). Healthy China 2030: a vision for health care. *Value in health regional issues*, *12*, 112-114.

Tardif, T., & Fletcher, P. (2008). *Chinese Communicative Development Inventories: user's guide and manual*. Beijing: Peking University Medical Press.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. Journal of Speech, *Language, and Hearing Research, 40*, 1245–1260.

Trawick-Smith, J. (2018). *Early Childhood Development: A Multicultural Perspective* (7th ed.). Pearson.

Ungerer, J. A., & Sigman, M. (1984). The relation of play and sensorimotor behavior to language in the second year. *Child development*, 1448-1455.

Visser-Bochane, M. I., Reijneveld, S. A., Krijnen, W. P., Van der Schans, C. P., & Luinge, M. R. (2020). Identifying milestones in language development for young children ages 1 to 6 years. *Academic pediatrics*, *20*(3), 421-429.

Wetherby A. & Prizant, B. (2002). *Communication and Symbolic Behaviour Scales.* Baltimore, MD: Paul H. Brookes.

Winkler, C., Linden, K., Mayr, A., Schultz, T., Welchowski, T., Breuer, J. & Herberg, U. (2020). RefCurv: A software for the construction of pediatric reference curves. *Software Impacts, (6), 100040.* https://doi.org/10.1016/j.simpa.2020.100040

World Health Organization. (2016). *China-WHO country cooperation strategy 2016-2020.* Retrieved July 2, 2022 from https://www.who.int/publications/i/item/WPRO-2016-DPM-003

World Health Organization, United Nations Children's Fund & World Bank Group. (2018). *Nurturing care for early childhood development.* Retrieved July 2, 2022 from https://nurturing-care.org/

Xue,Y., Bandel, E., Vogel, C. A. & Boller, K. (2015). *Measuring Infant/Toddler Language Development: Lessons Learned about Assessment and Screening Tools* (OPRE Brief 2015-52). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Yang, X. (2020, June 27). *Scio briefing on Reform and development of health sector in 13th five-Year Plan Period (2016-2020).* The State Council Information Office of the People's Republic of China. Retrieved June 27, 2022, from *http://english.scio.gov.cn/pressroom/2020-11/03/content_76875209.htm*

Yew, S. G. K., & O'Kearney, R. (2013). Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: Meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry*, *54*(5), 516-524.

Yew, S. G. K., & O'Kearney, R. (2015). The role of early language difficulties in the trajectories of conduct problems across childhood. *Journal of Abnormal Child Psychology*, *43*(8), 1515-1527.

Yoder, P. J., & Warren, S. F. (2004). Early predictors of language in children with and without Down syndrome. *American journal on mental retardation*, *109*(4), 285-300.

Zimmerman, I.L., Steiner, V.G., & Pond, R.E. (2002a). *Preschool Language Scales, Fourth Edition (PLS-4).* San Antonio, TX: Pearson.

Zhang, Y., Jin, X., & Shen, X. (2003). The establishment of screening standard for the language development of 2-3-year old children in Mainland China. *Zhong Guo Er Tong Bao Jian Za Zhi*, *11*, 308-311.

Zhu, H. (2002). *Phonological Development in Specific Contexts: Studies of Chinese-Speaking Children.* New Sommerset, UK: Multilingual Matters Ltd.