International Journal of Speech-Language Pathology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iasl20

Early language and communication development in Chinese children: Adaption and validation of a parent report instrument

Yiwen Zhang , Xiaojuan Xu , Yanrui Jiang , Wanqi Sun , Yan Wang , Yuanjin Song , Shumei Dong , Qi Zhu , Fan Jiang & Li Sheng

To cite this article: Yiwen Zhang , Xiaojuan Xu , Yanrui Jiang , Wanqi Sun , Yan Wang , Yuanjin Song , Shumei Dong , Qi Zhu , Fan Jiang & Li Sheng (2020): Early language and communication development in Chinese children: Adaption and validation of a parent report instrument, International Journal of Speech-Language Pathology, DOI: 10.1080/17549507.2020.1817558

To link to this article: https://doi.org/10.1080/17549507.2020.1817558

International Journal of Speech-Language Pathology, 2020; Early Online: 1–10

Early language and communication development in Chinese children: Adaption and validation of a parent report instrument

YIWEN ZHANG[1,2], XIAOJUAN XU[1,2], YANRUI JIANG[1,2], WANQI SUN[1,2], YAN WANG[1,2], YUANJIN SONG[1,2], SHUMEI DONG[1,2], QI ZHU[1,2], FAN JIANG[1,2] & LI SHENG[1,3]

[1]*Department of Developmental and Behavioural Paediatrics, Shanghai Children's Medical Centre, Shanghai Jiao Tong University School of Medicine, Shanghai, China,* [2]*Ministry of Education-Shanghai Key Laboratory of Children's Environmental Health, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China and* [3]*Department of Communication Sciences and Disorders, University of Delaware, Newark, Delaware, USA*

Abstract

Purpose: To fulfil the needs of assessment tools in the Chinese population, we adapted the LENA Developmental Snapshot, a parent survey that measures early language and communication development in English-speaking children. We reported the psychometric properties of the adapted questionnaire and evaluated the metric and functional equivalence between the adapted and the original instruments.
Method: The Snapshot was translated into Chinese and reviewed by an expert panel. English-specific items (e.g. past tense, plural) were mapped onto functionally similar Mandarin vocabulary and structures. The questionnaire was administered to 1300 families with children ages 2–48 months. The Bayley Scales of Infant Development was administered to a subsample.
Result: Scores on the adapted questionnaire showed age-related increases in the 7–36 month age range and correlated with scores on the Chinese Bayley. The questionnaire showed high internal consistency and split-half reliability. Comparison with the US norm revealed slightly lower performance in the current sample. Adaptations of English-specific items resulted in functionally equivalent targets.
Conclusion: Despite differences in linguistic roots, the Chinese adaption of the LENA Snapshot captured developmental changes in children's language and communication abilities. Additional norming and validation efforts are needed in a more representative sample.

Keywords: *questionnaire; adaptation; functional equivalence; metric equivalence; Chinese; assessment*

Introduction

China has an enormous paediatric population. As one of the most common childhood developmental disorders, developmental language disorder (DLD) is estimated to affect 7–19% of the general population (Johnson et al., 1999; Tomblin et al., 1997). Similar to Armon-Lotem, de Jong, and Meir (2015), we believe that DLD prevalence is constant across races, ethnicities, and languages. With a prevalence rate of 7%, as many as 5 million Chinese children ages 4 to 9 years could be affected by DLD (National Statistics Bureau of China, 2010). Though DLD is usually not diagnosed until 4 years of age (McKean et al., 2017), signs of language and communication delays may be noticed as early as the first year of life. DLD is frequently associated with social, emotional, and academic problems in school-age years, and reduced educational attainment and occupational status in young adulthood (Conti-Ramsden & Botting, 2008; Conti-Ramsden, Durkin, Toseeb, Botting, & Pickles, 2018; Rescorla, 2009). Nevertheless, recent longitudinal studies from Canada and the UK indicated improved longer-term outcomes (Conti-Ramsden et al., 2018) and comparable quality-of-life ratings in individuals with and without a history of DLD. Although there is no guaranteed remedy for DLD, a considerable portion of children with DLD do benefit from intervention, lending support to the importance

Correspondence: Li Sheng, Communication Sciences and Disorders, University of Delaware, 100 Discovery Blvd., Newark, Delaware 19711, USA. Email: lsheng@udel.edu; Fan Jiang, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, 1678 Dongfang Rd., Shanghai 200127, China. Email: fanjiang@shsmu.edu.cn

ISSN 1754-9507 print/ISSN 1754-9515 online
Published by Taylor & Francis
DOI: 10.1080/17549507.2020.1817558

of early detection and intervention (e.g. Robertson & Ellis Weismer, 1999).

In the Chinese setting, there is currently a severe shortage of qualified personnel and accessible standardised measures to support the early detection of DLD (Salas-Provance, 2011; Sheng, Shi, Wang, Hao, & Zheng, 2020). DLD detection is mostly achieved through self-referral by concerned parents to local hospitals, some of which are equipped with developmental and behavioural paediatricians who have knowledge of early language development and may have protocols to diagnose DLD. Follow-up treatment is even more variable than the state-of-affairs for diagnosis as there is no legislature to stipulate service for this disorder. With this backdrop in mind, the goal of the current study is to describe the Chinese adaptation of a parent survey to assess early language and communication development, as an initial attempt to fill the significant gap of instrument development for this population.

Numerous assessment tools exist for infants, toddlers, and preschoolers in the English language (e.g. the *Communication and Symbolic Behaviour Scales Developmental Profile*, Wetherby & Prizant, 2002; the *MacArthur-Bates Communicative Development Inventory*, Fenson et al., 2007; the *Language Development Survey*, Rescorla & Alley, 2001; the *Rossetti Infant-Toddler Language Scale*, Rossetti, 1990; the *Pre-school Language Scales*, Zimmerman, Steiner, & Pond, 2011). These tools provide clinicians and language specialists with precise information of the child's language and prelinguistic communication behaviours and enable timely identification of language delays and disorders. However, diagnosis and intervention of language delay in China is extremely constrained due to the lack of well-designed, efficient tools that are culturally and linguistically appropriate.

To the best of our knowledge, there are only two standardised language assessment tools normed on children in the People's Republic of China. The *MacArthur Bates Communicative Development Inventory* (Fenson et al., 2007), a parent report measure for children between 8 and 30 months of age, was adapted into Mandarin and Cantonese (Tardif & Fletcher, 2008) and has been found to be a sound tool for measuring early language development in Chinese children (the *Chinese Communicative Development Inventory*, CCDI, Tardif, Fletcher, Liang, & Kaciroti, 2009). Moreover, a simplified short form of the CCDI is available (Soli, Zheng, Meng, & Gang, 2012), which reduces the administration time and linguistic expertise needed to interpret the scores. However, oral administration of the CCDI by a trained clinician is recommended due to dialectal variations of vocabulary words, which presents a potential challenge in the busy clinical setting. The Mandarin norms of the CCDI were collected in Beijing only and the Cantonese norms in Hong Kong only. These norms may not be representative of the entire Mandarin- and/or Cantonese-speaking populations. The *Diagnostic Receptive and Expressive Assessment of Mandarin* (DREAM, Ning, Liu, & de Villiers, 2014) is a comprehensive language test normed for Chinese children, providing composite scores for semantics, syntax, receptive language, and expressive language. However, the test is very costly; its administration requires extensive training and credentialing, and it cannot be used with children younger than 30 months.

Taken together, there is a pressing need for new assessment tools for Chinese children in the 0–3 years age range that are time-efficient, easy to administer and score, and not constrained by regional dialectal variations. In this article we report the process of adapting an existing parent questionnaire from English to Chinese as part of an ongoing effort to develop a standardised norm-referenced early language and communication assessment tool for the Chinese population.

Issues in test adaptation

Translating an established tool from English to another language is common in child development research and is an effective solution when there is a lack of available instruments (van Widenfeld, Treffers, de Beurs, Siebelink, & Koudiks, 2005). Oftentimes, equivalence is assumed between the original and the translated instrument but in most cases, little information is provided about the translation and adaptation process (van Widenfeld et al., 2005). Peña (2007) described four types of equivalence when adapting an instrument cross-linguistically. *Linguistic equivalence* refers to the equivalent use of words and linguistic meanings across the original and the translated versions and is typically ensured through methods such as back-translation and expert review. *Functional equivalence*, or *conceptual equivalence* (Hilton & Skrutkowski, 2002), means that the instruments will elicit the same target behaviour, even though there are differences in wording and instruction. To achieve functional equivalence, the adaptation may have to shift away from the source instrument's wording to represent the construct in a way that is conceptually similar or linguistically more familiar in the target language. In other words, linguistic equivalence may have to give way to functional equivalence to obtain equivalence in meaning and salience. *Cultural equivalence* concerns how respondents from different cultures and linguistic groups may interpret instructions and test items in culturally disparate ways, which may lead to response biases that threaten the functional equivalence of the instrument. Finally, *metric equivalence* refers to equivalence in item difficulty. This type of equivalence is essential for developing a tool that assesses language ability.

The current study reports the translation and adaptation process of the LENA Development Snapshot (Gilkerson, Richards, Greenwood, &

Montgomery, 2017), a parent questionnaire designed for English, into Mandarin Chinese. Our goal was to develop a tool that is functionally and metrically equivalent to the original version.

The LENA developmental snapshot (snapshot)

The Snapshot is a norm-referenced standardised parent report instrument used to assess communication and language skills among children aged 2 to 36 months. The final version of the instrument contains 52 age-ordered items. Parents read short questions (e.g. When you talk to your child, does he/she look in the direction of your voice?) and have to check "yes" or "not yet". Parents stop the questionnaire after five consecutive "not yet" answers. A developmental age score and age-referenced standard score are provided to give parents information about the child's language skills relative to his/her chronological age.

Questions on the Snapshot were selected based on the expertise of a team of speech-language pathologists (SLPs), linguists, and statisticians and review of 10 commonly used paediatric language and cognitive assessments. In the pilot phase, a 54-item survey was given to 15 families. A qualitative evaluation was conducted using parent interview responses to refine and modify the survey. In the normative data collection phase, a revised 63-item survey was given to 308 families. Item analysis was used to eliminate 11 questions that did not provide consistent developmental information and re-order items according to developmental stages. Internal consistency was very high (α = .98). Test-retest reliability in the form of correlations between adjacent monthly scores over a 12-month period for a subsample ($n = 59$) ranged from .93 to .98. Criterion validity was reported as substantial to almost perfect correlations (r ranged from .67 to .97) with 11 other indices gathered from standardised language and cognitive tests. Limitations of the instrument include the relatively small size of the US norm ($n = 308$) and lack of consideration of social or pragmatic skills. Nevertheless, the Snapshot has advantages such as quick administration (i.e. takes 10–15 min to give), ease of use (i.e. can be completed, scored, and interpreted by parents without the help of trained professionals), and amenability to repeated administration to monitor progress (Gilkerson et al., 2017).

The current study

The objective of the current study was to investigate the feasibility of adapting LENA Snapshot into Mandarin.[1] First, the test adaptation process was described. Then, we evaluated the extent of age-related changes in scores in the adapted instrument using a large sample of respondents and compared the results from our sample to those from the Snapshot normative sample. In addition, we reported on the reliability and concurrent criterion-referenced validity of the instrument, as well as the functional equivalence and metric equivalence between the adapted instrument and the source instrument. This allowed us to assess each item and make recommendations for a refined version. Along with others (van Widenfeld et al., 2005), we believe that translation and adaptation of test instruments have not received adequate attention in paediatric developmental research. Through our report, we hope to demonstrate that adapting an instrument into a new language is a major undertaking involving many steps and considerations, but that despite some errors and pitfalls, this objective is achievable.

Method

Participants

1300 typically developing children between 2 and 48 months of age and their primary caregivers were recruited from 17 hospitals or community health centres across eight different provinces (Guangxi, Hebei, Hunan, Inner Mongolia, Jiangsu, Shandong, Yunan, Zhejiang) and the city of Shanghai in Mainland China. These sites were selected to represent different regions of the country with variable population size and levels of socio-economic development, and to achieve balanced urban/rural distribution. Parents who brought in their children for health check-ups at these sites were invited to participate. Children with physical disabilities, cerebral palsy, Down Syndrome, autism, and known congenital organic diseases were not included. The average age of the sample was 17.77 months (SD = 12.93). There were 711 (54.7%) boys and 589 girls (45.3%). Among the participants who reported their place of residence ($n = 1208$), 54.1% were urban dwellers and 45.9% lived in rural areas. The average birth weight of the children was 3339g (SD = 454, $n = 1281$). Among those who reported mother's educational level ($n = 1284$), 29.2% had 9 years of education or less; 28.1% had a high school education, 36.7% had an associate or college degree, and 6% had a graduate degree. Of the fathers ($n = 1261$), 26.4% had 9 years of education or less, 27.9% had a high school education, 38.9% had an associate or college degree, and 6.7% had a graduate degree. Average household size was 4.25 persons (SD = 1.24, $n = 1295$). The majority of the families had grandparents living with them (53.5%, $n = 1287$), 45.8% of the sample were nucleus families with both parents present, and 0.7% were single-parent families. Table I presents demographic information of the current sample as well as relevant information from the most recent census (National Statistics Bureau of China, 2010). Even though the gender distribution of our sample closely mimicked the national distribution, the current Chinese sample had higher parental education than the average education of the broader Chinese adult

population. Compared to the Snapshot US normative sample (Gilkerson et al., 2017), our sample had somewhat lower education. Specifically, 56% of the US sample included mothers who had post-secondary education, whereas this percentage was 42.7% for our sample. Only 11% of the US sample were mothers who had "some high school", whereas 29.2% of our sample were mothers who had 9 years or less education.

Fifteen age bands were formed using a one-month interval for the 2- to 6-month-olds, a three-month interval for the 7- to 24-month-olds, and a six-month interval for the 25- to 48-month-olds. The age segmentation was based on the *Bayley Scales of Infant Development – Chinese* (Yi, 1995) which uses a similar age division. Table II shows the sample size, mean age, and gender distribution of each age band.

Table I. Demographic information of the current sample and the population in general.

	Our sample	National[a]
Total *N*	1300	
Female	589 (45.3%)	45.6%[b]
Male	711 (54.7%)	54.3%[b]
N for Place of residence	1208	
Urban dwellers	54.1%	n/a[c]
Rural dwellers	45.9%	n/a[c]
N for mother's educational	1284	
≤9 years	29.2%	71.0%[c]
High school	28.1%	15.3%[c]
Associate/college degree	36.7	13.2%[c]
Graduate degree	6%	0.5%[c]
N for father's educational	1261	
9 years	26.4%	66.2%[d]
High school	27.9%	18.3%[d]
Associate or college degree	38.9%	14.5%[d]
Graduate degree	6.7%	0.6%[d]
Average household size	4.3	3.1[a]

Note. [a]The national information was extracted from the 2010 Census published by the National Statistics Bureau of China. [b]Percentages were calculated based on the national population between 0 and 4 years of age. [c]Percentages were calculated based on females between 20 and 49 years of age. [d]Percentages were calculated based on males between 20 and 49 years of age. [e]N/a = not available. The 2010 national Census reported the percentage of the population who held an agricultural (70.86%) and non-agricultural (29.14%) residency status, according to the "hukou" registration system. This registration system is tied to the birthplace of individuals. Because migration from rural to urban areas is common, this registration system does not provide accurate information of the current dwelling status of individuals. For relevant information, see Chan and Zhang (1999).

Instruments

The Chinese Early Language and Communication Questionnaire (CELCQ) (Xu, 2014) was translated and modified from the LENA Developmental Snapshot (Gilkerson et al., 2017). A fluent Mandarin-English bilingual translated the Snapshot into Mandarin Chinese. Three individuals with native Mandarin proficiency and high English proficiency reviewed the translation and resolved any disagreements by consensus. The three individuals included two developmental paediatricians, one with over 30 years of clinical experience, the other with over 15 years of clinical experience, and one linguist who had over 12 years of research experience on child language acquisition. Of the 52 Snapshot items, 46 are general statements about the child's functioning. For example, item 1 asks "When you talk to your child, does he/she look in the direction of your voice?" These items were translated literally. The remaining six items focus on English-specific vocabulary items or grammatical structures such as articles, the plural –s, and past tense. These items were mapped onto concepts and grammatical structures that tap onto similar underlying constructs (see Table III for the original wording and the Mandarin translation).

The Bayley Scales of Infant Development (BSID) – *Chinese* (Yi, 1995) was used as the criterion measure to assess the concurrent validity of the CELCQ. The Chinese BSID was modified from the English version of the BSID (Bayley, 1969). The Chinese norms were developed using 2409 normal children between 2 and 30 months from urban and rural areas across China. The modification resulted in some shifting of item order but all questions in the English version were retained. In a validation study with 206 children, Sun, Ren, and Su (1996) found that the test had good to excellent test-retest (0.91–0.94), split-half (0.82–0.90), and inter-rater (0.98–0.99) reliability. The test score distribution closely paralleled the distribution of the US norms and correlated significantly

Table II. Age groups and mean total scores on the Chinese Early Language and Communication Questionnaire (95% confidence interval).

Age group	Total (*n*)	Mean Age (months)	Male (%)	Total
2∼	75	2.61 (0.36)	56.0%	5.67 (5.26–6.07)
3∼	92	3.37 (0.30)	53.3%	6.39 (5.90–6.88)
4∼	75	4.51 (0.31)	58.8%	7.35 (6.91–7.78)
5∼	83	5.50 (0.33)	63.9%	8.08 (7.56–8.61)
6∼	71	6.35 (0.29)	49.3%	8.32 (7.77–8.88)
7–9	95	8.59 (0.82)	56.8%	10.77 (10.08–11.46)
10–12	95	11.73 (0.83)	55.8%	16.17 (15.10–17.24)
13–15	83	14.51 (0.92)	54.2%	19.03 (17.85–20.22)
16–18	118	17.71 (0.78)	56.8%	23.29 (21.99–24.58)
19–21	73	20.46 (0.79)	58.9%	25.22 (23.43–27.00)
22–24	98	23.41 (0.92)	46.9%	32.11 (30.37–33.86)
25–30	102	28.37 (1.74)	59.8%	35.72 (33.67–37.76)
31–36	96	34.30 (1.77)	51.0%	43.77 (42.48–45.06)
37–42	73	39.93 (1.84)	50.7%	45.07 (43.77–46.37)
43–48	71	45.62 (1.60)	46.5%	47.56 (46.67–48.46)

Note: 7–9 months <10–12 months, 13–15 months <16–18 months, 19–21 months <22–24 months <25–30 months <31–36 months. All significant comparisons were at the adjusted α level of 0.0036.

Table III. English-specific items, Mandarin targets, and English gloss of the Mandarin targets.

Item # in Snapshot	Item # in CELCQ (% yes)	English target	Mandarin target	English gloss of Mandarin target
39	44 (21.6)	Is your child adding "-s" to words to indicate "more than one"? For example: Does your child say "cats" for more than one cat, or "spoons" for more than one spoon?	你孩子会正确表达一个和多个（一个以上）的物品吗? 例如:您孩子会说"很多猫"或"几个勺子"吗?	Does your child correctly indicate one versus multiple (more than one) objects? For example: Does your child say "many cats" or "a few spoons"?
41	41 (23.7)	Is your child adding "-ing" to the end of verbs to indicate ongoing action? For example: Does your child use words like "eating", "jumping", and "running"?	您孩子会表示正在发生的动作吗？ 例如:您孩子会这样说"正在吃"、"正在跳"、"正在跑"吗?	Is your child able to indicate ongoing actions? For example: Does your child say things like "zhengzai-eat", "zhengzai-jump", "zhengzai-run"? Note. "Zhengzai" is an imperfective aspect marker.
42	39 (25.1)	Does your child use the words "a", "an", and "the"? For example: Does your child say things like "a bed", "an apple", and "the ball"?	您孩子可以使用指示代词，如"这""一"吗? 例如:您孩子会说一些像"一张床", "一个苹果",和"这个球"吗?	Does your child use demonstrative pronouns such as "this" or numerals like "one"? For example: Does your child say things like "one-classifier-bed", "one-classifier-apple", and "this –classifier-ball"?
44	49 (10.7)	Does your child understand concepts like "least", "most", and "first"?	您孩子理解"至少"、"最"、"第一"的意思吗？	Does your child understand "least", "most", and "first"?
46	47 (16.7)	Does your child use the plural pronouns "we", "they", "them", and "us"?	您孩子会使用复数代词"我们","他们"吗?	Does your child use plural pronouns "wo-men", and "ta-men"? Note. "Men" is a plural marker used with certain nouns and pronouns
47	42 (23.4)	Is your child adding "-ed" to the end of verbs to indicate an action that happened in the past? For example: Does your child say things like "jumped" or "played"?	您孩子会表达过去发生的动作吗？ 例如:您孩子会说一些像"跳过了"或"玩过了"的词语吗?	Is your child able to indicate actions that were completed in the past? For example: Does your child say things like "jump-guole" or "play-guole"? Note. "Guole" is a perfective aspect marker.

with children's scores on the Gesell Developmental Schedules (Gesell, 1949), a commonly used assessment tool among Chinese paediatricians. Like the original English test, the Chinese BSID has three components: a 163-item Mental Scale that yields a Mental Development Index (MDI), an 81-item Motor Scale that yields a Psychomotor Development Index (PDI), and a 30-item Infant Behaviour Record. Only the first two scales were administered in the current study. Although newer editions of the BSID have been published, they have not been normed for the Chinese population. Two researchers independently reviewed the items on the mental scale of the Chinese BSID and coded each item as measuring language and communication behaviours, or nonverbal cognition. Item-by-item agreement was 99% and disagreements were resolved through discussion. In the end, 39 items were coded as measuring prelinguistic communication and expressive or receptive language, and the remaining 124 items as measuring nonverbal cognitive skills.

Procedures

The study procedures were approved by the ethics board of a major paediatric hospital in China. The primary caregiver of the child signed an informed consent, filled out a demographic information questionnaire, and then completed the CELCQ. A trained paediatrician at each site was available to answer any questions the parents may have. Similar to the Snapshot, parents were asked to check the "yes" box if their child had already displayed a certain behaviour or skill and the "not yet" box if the child had not. The item order of the Snapshot was kept for all but six language-specific items (see Table III). These items were administered at the end of the instrument and they were answered by all the participants. To avoid fatigue and frustration, parents can stop after five consecutive "not yet" answers for the first 46 questions. This ceiling rule was used following the reasoning that the items on the Snapshot were ordered according to developmental progression among English-speaking children and that these 46 items targeted general behaviours that should be acquired in a universal order.[2]

The trained paediatrician at each site completed the BSID-Chinese for children between the age of 2 and 30 months, the age range for which the Chinese BSID was normed. The BSID requires direct administration from a trained paediatrician, who used a combination of direct observation and various stimuli to elicit responses from the child, and scored the child as demonstrating or not demonstrating the required

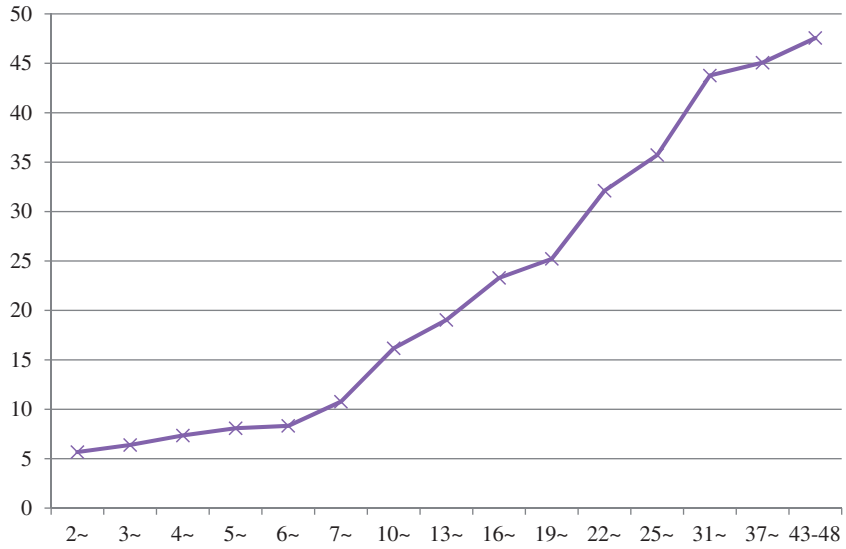

Figure 1. Mean scores on the CELCQ across the 15 age bands. Note: 7–9 months < 10–12 months, 13–15 months < 16–18 months, 19–21 months < 22–24 months < 25–30 months < 31–36 months. All significant comparisons were at the adjusted α level of 0.0036.

response or skill. To ensure procedural fidelity, personnel from all 17 test sites participated in a three-day workshop before the study commenced. The workshop covered the project objectives and administration procedures of the questionnaire, the CELCQ, and the BSID. During the workshop, each trainee was observed completing the entire test protocol and received feedback. For the BSID, the trainees were evaluated on their interaction with the infant, adherence to administration manual, and accuracy of scoring. All trainees must pass the evaluation to become a data collector. To ensure proficiency with the procedures, the first 10 participants at each site were considered practice cases and were not included in the sample. Data collection lasted for two years between February, 2012 and February, 2014. Every three months, the staff at each site was required to provide a progress report to the project Principal Investigator, during which time any procedural questions were discussed.

All data were entered using Epidata 3.0 and statistical analyses were conducted using SPSS 17.0. Original paper files were mailed from the test sites to the main site and entered by trained members of the research team. Each file was entered twice by two independent research assistants. The two entries were then compared and any inconsistencies were resolved. For the CELCQ, each "not yet" response received a code of 0 and "yes" response a code of 1. Items left blank after the ceiling was reached were coded as 0.

Result

Performance by age level

Mean scores on the CELCQ across the 15 age bands are displayed in Table II and Figure 1. A one-way ANOVA was conducted to check for significant age-related differences in scores. The age effect was significant, $F(14, 1285)=558.30$, $p<0.001$, $\eta_p^2=0.86$. To adjust for multiple comparisons, a corrected α level of 0.0036 (0.05 divided by 14) was used for follow-up pairwise comparisons. Five significant pairwise comparisons were noted: 7–9 vs. 10–12 months, 13–15 vs. 16–18 months, 19–21 vs. 22–24 months, 22–24 vs. 25–30 months, and 25–30 vs. 31–36 months.

Next we compare our results to Gilkerson et al. (2017, p.105, Figure 1), which showed the mean total checked items for each age group in the US Snapshot sample. The CELCQ sample started with a mean total of 5.67 items at 2 months of age, then reached 10 items at around 7 months, 20 items at about 16 months, 30 items at around 22 months, 40 items at around 31 months, and 45 items at around 36 months. On the other hand, the US Snapshot sample started with about 6 items at 2 months, reached 10 items at around 5 months, 20 items at about 12 months, 30 items at about 20 months, 40 items at around 29 months, and 45 items at around 33 months. In general, our sample trailed behind the Snapshot sample by a couple months for each milestone.

Reliability

Cronbach's alpha was calculated to examine the internal consistency of the instrument. The alpha value was 0.968. While this value indicates high interrelatedness of the test items, it is outside the acceptable range of alpha values (0.70–0.95 as recommended by Terwee et al., 2007) and suggests that there are redundancies in the instrument. Split-half reliability of the CELCQ was 0.979.

Correlations between the CELCQ and the Bayley

This analysis included 1047 children between the age of 2 and 30 months who had Bayley MDI scores. A

partial correlation was conducted to examine the degree of association between children's performance on the CELCQ and their scores for the 39 language and communication items on the Bayley while factoring out age. This correlation was significant and considered moderately strong in medical research, $r = 0.71$, $r^2 = 0.51$, $p < 0.001$ (Akoglu, 2018) and surpassed the 0.70 threshold recommended for criterion validity (Terwee et al., 2007).

Item evaluation

Items were evaluated in regards to their metric and functional properties. Metric equivalence refers to the difficulty level of the items across versions. We calculated the percent of children who demonstrated each skill/behaviour and used the percentages to re-order the items from easiest to most difficult. Because the CELCQ total scores did not show age-related changes for children younger than 7 months and older than 36 months, we limited this analysis to the 760 children between 7 and 36 months of age based on the reasoning that the refined tool should target the most appropriate age range. This resulted in a large number of items being re-ordered. For instance, item #10 on the Snapshot "Does your child shout or use vocalisations/make sounds to get your attention?" was re-ordered as #5 on the CELCQ because 97% of the sample answered "yes". Item #44 on the Snapshot "Can your child retell a story or event with a beginning, middle, and end without using pictures?" was re-ordered as #52 on the CELCQ because only 7.1% of the sample answered "yes".

Item #48 on the Snapshot asks "Does your child spontaneously produce sentences that are 10 or more words in length?" This was deemed a general behaviour question and translated literally. Based on the percent of "yes" answers (21.6%), this item was ranked #43 out of a total of 52 items. Even though the shift in item order between the original and translated question was not drastic, we caution against the use of this item in future versions. The definition of word for Chinese is controversial (Li & Thompson, 1981). The layperson may equal "word 词" to "character 字" or "syllable 音节". Therefore, the report may be inconsistent across individuals. For example, "chi-fan 吃饭"[eat-meal] may be considered either one word or two words by two different individuals.

Finally, we turn to the six language-specific items. Table III presents the item numbers (or developmental ranking) of these particular items in the original Snapshot and the CELCQ along with the percent of the sample that demonstrated those behaviours. Shift in item order was within five places for all items. Even though features such as the plural form and definite and indefinite articles are unique to English, the concepts of number and definiteness can be expressed in Mandarin in different ways. Similarly, Mandarin does not have tense but it uses aspect-marking morphemes to depict the state of actions. In other words, linguistic equivalence notwithstanding, the adaptation was able to capture concepts that are functionally equivalent across English and Mandarin.

Discussion

The objective of the study was to illustrate the process of adapting an existing parental report tool into Chinese and evaluate the reliability and validity of the adapted test. Moreover, we evaluated individual test items and made recommendations for future refinement of the instrument. As several authors have pointed out (Hambelton, 2001; van Widenfelt et al, 2005), there is a lack of consistency in how research articles provide information about test translation and adaptation. Subsequently there is a lack of documentation in published research on how test instruments are translated and adapted into other languages. At the same time, this kind of information is highly valuable to other researchers to avoid specific translation and procedural problems. The need for full documentation of the adaptation process may be greater when an English instrument is translated into an Asian language than when it is adapted for non-English European populations (Weisz et al., 1987) due to the greater differences in linguistic structure and culture. We believe the information presented in this article is helpful for other researchers who are interested in generalising English-based research to other Asian languages.

Using a cross-sectional sample of 1300 children, we found significant age-related differences in scores on the CELCQ between certain subgroups of children aged 7 to 36 months. The total scores were rather flat among children younger than 7 months of age or older than 36 months of age, indicating floor and ceiling effects and a lack of developmental sensitivity for these age groups. The CELCQ showed very high internal consistency and split-half reliability. However, Cronbach's alpha values higher than .95 are indicative of redundancies and suggest the potential need to shorten the instrument. Furthermore, Terwee et al. (2007) suggested that for health questionnaire development, internal consistency should be calculated after factor analysis has been performed. Factor analysis can determine if the test items form one cohesive dimension or belong to more than one dimension or subscale. This is especially important if the instrument is designed to assess multiple subcomponents of a construct. Further development of the tool should consider inclusion of a factor analysis.

There was a moderately strong correlation between total score on the CELCQ and total score for the language items of the Bayley, even after participants' age was factored out. Recall that the Bayley was administered by the paediatrician and scored according to clinical observation of the child's behaviours. The significant correlation between parent

report and paediatrician observation provided evidence of the criterion validity of the CELCQ as a potential tool to measure language (Terwee et al., 2007).

This test adaptation process affords us the opportunity to learn several valuable lessons that will inform future efforts to refine this instrument. Despite the lack of linguistic equivalence, functional and conceptual equivalence can be achieved when adapting English-specific items into another language. Even though Mandarin Chinese does not have grammatical markers for tense (e.g. *-ed* in English), it uses perfective aspect markers such as *-le* to indicate completed actions and imperfective aspect markers such as *-zai* to depict ongoing actions. This provides a closely parallel form of verb morphology, which has also been found to develop more slowly in Chinese children with DLD (Fletcher, Leonard, Stokes, & Wong, 2005). In the absence of definite and indefinite articles, we targeted demonstratives (e.g. this) and "numeral-classifier-noun" phrases. In future versions, it would be prudent to separate these targets into individual items because demonstratives and classifiers are separate categories of words.

We followed the Snapshot ceiling rule when administering the CELCQ given the assumption of universal patterns in general linguistic behaviour and out of the pragmatic concern of avoiding frustration for parents of the youngest groups. In retrospect, this decision was based on the unverified assumption of a universally applicable ceiling rule and was not advisable. This stopping rule may have contributed to the overall lower scores of the Chinese sample compared to the American sample. However, the lower scores could also be a result of other factors. First, our sample had comparable parental education as other large-scale studies of the Chinese population. For instance, the modal parental education was at the upper secondary level in Tardif, Fletcher, Liang, and Kaciroti's (2009) study of 1694 Beijing-based families. Similarly, over 60% of the parents had no more than a high school education in Zhang, Jin, Shen, Zhang, and Hoff's (2008) sample of 608 Shanghai-based families. However, compared to the US sample reported in Gilkerson et al. (2017), our sample had lower maternal education level. Given the strong relationship between mother's education and children's language development (Hoff, 2014), it is not surprising that our sample had lower scores than the US sample. Second, it is not uncommon for test adaptation studies to find lower performance in cross-language and cross-cultural samples than in the original sample (Sheng, 2018). Adapting the original *MacArthur-Bates Communicative Development Inventory* (Fenson et al., 2007) and the *Language Development Survey* (Rescorla & Alley, 2001), several studies found smaller vocabularies in children sampled in the UK, Australia, Denmark, Greece, and Poland than the American sample (Bavin et al., 2008; Bleses et al., 2008; Hamilton et al., 2000; Papaeliou & Rescorla, 2011; Rescorla, Constants, Białecka-Pikul, Stępień-Nycz, & Ochał, 2017). Issues related to differences in sample size and socioeconomic composition (Bavin et al., 2008), differences in parent response styles (Rescorla et al., 2017), complexity of the language to be acquired (Bleses et al., 2008; Rescorla et al., 2017); administration format (filling it out alone in the paediatrician's waiting room vs. interviewed by staff at home, Papaeliou & Rescorla, 2011), children's access to day-care settings (Bleses et al., 2008), and differences in frequency of occurrence of normed words (Hamilton et al., 2000) could all contribute to these observed differences. Nevertheless, the premature adoption of the ceiling rule remains a major limitation. Future refinement of the tool needs to empirically determine the appropriate ceiling rule.

Finally, in the pilot phase of test development, it would be important to conduct interviews with participants to have an in-depth look of their reactions and interpretations of the instructions, the items, and the response categories (van Widenfelt et al., 2005). These pilot participants should be representative of the target population in terms of age, socioeconomic background, and geographical region. These qualitative interviews are necessary to examine the cultural equivalence between the source and the adapted tool.

Questions may arise regarding the absence of items related to tone acquisition on the CELCQ because lexical tone is a salient speech characteristic of Chinese. We did not add questions related to tones in the adaptation because lexical tones are not dialect-neutral (Hou, 2002). In other words, different Chinese dialects may use the same sound sequence to represent the same meaning but these sound sequences may carry different tones. For example, the word "chi-fan 吃饭"[eat-meal] is pronounced with the first and fourth tone in standard Mandarin (chi1-fan4) but it is pronounced with the third and fourth tone in Hubei dialect (chi3-fan4). Moreover, the size of tone inventories differs across Chinese dialects (Hou, 2002). For example, standard Mandarin has five tones whereas Cantonese has nine tones. Therefore, asking parents to judge the tone accuracy of their children may become overly complicated. Remarkably, tone is reported to be mastered by 99% of children by two years of age and tone errors are extremely rare in child speech (Li & To, 2017; Zhu & Dodd, 2000).

The current version of the CELCQ can be obtained by contacting the LENA foundation and the authors of this paper. Diagnostic use of the tool is not appropriate for the entire Chinese population because we over-sampled families with higher education. Clinicians who work with Mandarin-speaking children may refer to this information for broad outlines of development. We recommend that Chinese paediatricians and early intervention personnel use the CELCQ as a screening tool for language delay and

use the preliminary results presented in Table II to guide their decision-making before the norms for this tool become available. Finally, clinicians may use the developmental information to give recommendations about intervention targets.

Refinement of the questionnaire should continue with broader population sampling and generation of new items through an expert panel of linguists, speech therapists, developmental psychologists, and early childhood educators. New items should capture Mandarin-specific linguistic structures and more advanced linguistic forms to enhance the tool's sensitivity for older age groups. Future efforts in improving the tool should incorporate the CCDI as an additional check of concurrent validity and account for the test-retest reliability and long-term predictive validity of the instrument.

Acknowledgments

This project was supported by the Chinese Ministry of Health [Grant 201002006]. We would like to thank Dr. Jill Gilkerson at the LENA foundation for sharing with us the LENA Developmental Snapshot and for her input on an earlier draft of the manuscript. We acknowledge the support of the following individuals and organisations: Ms. Cai Xia Ban at the Child Care Department of the Maternal and Child Health Hospital of the Guangxi Autonomous Region; Ms. Li Zhang at the Child Care Department of the Maternal and Child Health-Care Center of Huai'an City, Jiangsu Province; Ms. Rui Fang Zhang at the Child Care Department of the Maternal and Child Health Hospital of Inner Mongolia; Ms. Yun Liu at the Child Care Department of the Children's Hospital of Kunming City, Yunnan Province; Ms. Li Fang Kong at the Maternal and Child Health Hospital of Liuyang City, Hunan province; Ms. Jie Li at the Child Care Department of the Rizhao People's Hospital of Shandong Province; Ms. Sheng Xia Zhao at the Child Care Department of the Maternal and Child Health Hospital of Shijiazhuang City, Hebei Province; and Ms. Ai Ping Wang at the Child Care Department of the Maternal and Child Health Hospital of Yiwu City, Zhejiang Province. We also thank all the parents and children who participated in this research. Yiwen Zhang and Xiaojuan Xu contributed equally to this work.

Disclosure statement

The authors do not have any conflict of interest to disclose.

Notes

1. A permission letter from the LENA foundation was obtained to adapt the Snapshot and present some of the questions in journal publication.
2. This decision was made to shorten the visit considering the time it took to administer the consent, the demographic questionnaire, the CELCQ, and the BSID. It was also intended to reduce frustration by parents of the youngest children. In retrospect, we recognise that the ceiling rule developed for US children may not apply to a different context. We discussed this at length in the limitation section.

Funding

This study was supported by the Chinese Ministry of Health under Grant 201002006.

References

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*, 91–93. doi:10.1016/j.tjem.2018.08.001

Armon-Lotem, S., de Jong, J., & Meir, N. (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol, UK: Multilingual Matters. ISBN: 978-1-78309-311-3

Bavin, E.L., Prior, M., Reilly, S., Bretherton, L., Williams, J., Eadie, P., … Ukoumunne, O.C. (2008). The early language in Victoria study: Predicting vocabulary at age one and two years from gesture and object use. *Journal of Child Language*, *35*, 687–701. doi:10.1017/S0305000908008726

Bayley, N. (1969). *The Bayley scales of infant development*. New York, NY: Psychological Corporation.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T.O., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language*, *35*, 619–650. doi:10.1017/S0305000908008714

Chan, K.W., & Zhang, L. (1999). The Hukou system and rual-urban migration in China: Processes and changes. *The China Quarterly*, *160*, 818–855. doi:10.1017/S0305741000001351

Conti-Ramsden, G., & Botting, N. (2008). Emotional health in adolescents with and without a history of specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *49*, 516–525. doi:10.1111/j.1469-7610.2007.01858.x

Conti-Ramsden, G., Durkin, K., Toseeb, U., Botting, N., & Pickles, A. (2018). Education and employment outcomes of young adults with a history of developmental language disorder. *International Journal of Language & Communication Disorders*, *53*, 237–255. doi:10.1111/1460-6984.12338

Fenson, L., Marchman, V.A., Thal, D.J., Dale, P.S., Reznick, J.S., & Bates, E. (2007). *MacArthur-Bates communicative development inventories*. Baltimore, MD: Brookes Publishing.

Fletcher, P., Leonard, L.B., Stokes, S.F., & Wong, A.M.-Y. (2005). The expression of aspect in Cantonese-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *48*, 621–634. doi:10.1044/1092-4388(2005/043)

Gesell, A. (1949). *A. Gesell development schedules*. New York, NY: Psychological Corporation.

Gilkerson, J., Richards, J.A., Greenwood, C.R., & Montgomery, J.K. (2017). Language assessment in a snap: Monitoring progress up to 36 months. *Child Language Teaching and Therapy*, *33*, 99–115. doi:10.1177/0265659016660599

Hambelton, R. (2001). The next generation of the ITC Test translation and adaptation guidelines. *European Journal of Psychological Assessment*, *17*, 164–172. doi:10.1027//1015-5759.17.3.164

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, *27*, 689–705. doi:10.1017/s0305000900004414

Hilton, A., & Skrutkowski, M. (2002). Translating instruments into other languages: Development and testing processes. *Cancer Nursing*, *25*, 1–7. doi:10.1097/00002820-200202000-00001

Hoff, E. (2014). *Language development*. 5th ed. Belmont: Wadsworth, Cengage Learning. ISBN: 978-1133939092

Hou, J. (2002). *Xiandan Hanyu Fangyan Gailun* [Modern Chinese Dialects]. Shanghai: Shanghai Education Press.

Johnson, C.J., Beitchman, J.H., Young, A., Escobar, M., Atkinson, L., Wilson, B., … Wang, M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research*, *42*, 744–760. doi:10.1044/jslhr.4203.744

Li, C., & Thompson, S.A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press. ISBN: 978-0520066106

Li, X., & To, C.K.S. (2017). A review of phonological development of Mandarin-speaking children. *American Journal of Speech-Language Pathology*, *26*, 1262–1278. doi:10.1044/2017_AJSLP-16-0061

McKean, C., Reilly, S., Bavin, E.L., Bretherton, L., Cini, E., Conway, L., … Mensah, F. (2017). Language outcomes at 7 years: Early predictors and co-occurring difficulties. *Pediatrics*, *139*, e20161684. doi:10.1542/peds.2016-1684

National Statistics Bureau of China. (2010). *The 2010 Chinese National Census*. http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm

Ning, C.Y., Liu, X.L., & de Villiers, J.G. (2014). *The diagnostic receptive and expressive assessment of Mandarin*. Dallas, TX: Bethel Hearing and Speaking Training Center.

Papaeliou, C.F., & Rescorla, L.A. (2011). Vocabulary development in Greek children: A cross-linguistic comparison using the Language Development Survey. *Journal of Child Language*, *38*, 861–887. doi:10.1017/S030500091000053X

Peña, E.D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, *78*, 1255–1264. doi:10.1111/j.1467-8624.2007.01064.x

Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language, and Hearing Research*, *52*, 16–30. doi:10.1044/1092-4388(2008/07-0171)

Rescorla, L., & Alley, A. (2001). Validation of the language development survey (LDS): A parent report tool for identifying language delay in toddlers. *Journal of Speech, Language, and Hearing Research*, *44*, 434–445. doi:10.1044/1092-4388(2001/035)

Rescorla, L., Constants, H., Białecka-Pikul, M., Stępień-Nycz, M., & Ochał, A. (2017). Polish vocabulary development in 2-yer-olds: Comparisons with English using the Language Development Survey. *Journal of Speech, Language, and Hearing Research*, *60*, 1029–1035. doi:10.1044/2016_JSLHR-L-15-0385

Robertson, S.B., & Ellis Weismer, S. (1999). Effects of treatment on linguistic and social skills in toddlers with delayed language development. *Journal of Speech, Language, and Hearing Research*, *42*, 1234–1248. doi:10.1044/jslhr.4205.1234

Rossetti, L.M. (1990). *The Rossetti infant-toddler language scale: A measure of communication and interaction*. Austin, TX: LinguiSystems.

Salas-Provance, M.B. (2011). The China gap for speech-language services. *The ASHA Leader*, *16*(13), 10–11. doi:10.1044/leader.WB1.16132011.np

Sheng, L., (2018). Typical and atypical lexical development. In A. Bar-On & D. Ravid (Ed.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistic perspectives* (pp.101–116). Berlin, Boston: De Gruyter Mouton.

Sheng, L., Shi, H., Wang, D., Hao, Y., & Zheng, L. (2020). Narrative production in Mandarin-speaking children: Effects of language ability and elicitation method. *Journal of Speech, Language, and Hearing Research*, 63, 774–792. doi:10.1044/2019_JSLHR-19-00087

Soli, S.D., Zheng, Y., Meng, Z., & Gang, L. (2012). Clinical assessment of early language development: A simplified short form of the Mandarin communicative development inventory. *International Journal of Pediatric Otorhinolaryngology*, *76*, 1255–1264. doi:10.1016/j.ijporl.2012.05.015

Sun, X., Ren, Y., & Su, Z. (1996). Bayley scales of infant development: A validation study. *Maternal & Child Health Care of China*, *11*, 51–53.

Tardif, T., & Fletcher, P. (2008). *The Chinese communicative development inventory (Putonghua and Cantonese versions): Manual, forms, and norms*. Beijing: Beijing University Medical Press.

Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (2009). Early vocabulary development in Mandarin (Putonghua) and Cantonese. *Journal of Child Language*, *36*, 1115–1144. doi:10.1017/S0305000908009185

Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt, D.A.W.M., Knol, D.L., Dekker, J., … de Vet, H.C.W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*, 34–42. doi:10.1016/j.jclinepi.2006.03.012

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. doi:10.1044/jslhr.4006.1245

van Widenfelt, B.M., Treffers, P.D.A., de Beurs, E., Siebelink, B.M., & Koudijs, E. (2005). Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clinical Child and Family Psychology Review*, *8*, 135–147. doi:10.1007/s10567-005-4752-1

Weisz, J.R., Suwanlert, S., Chaiyasit, W., Weiss, B., Achenbach, T.M., & Walter, B.R. (1987). Epidemiology of behavioral and emotional problems among Thai and American children: parent reports for ages 6 to 11. *Journal of the American Academy of Child and Adolescent Psychiatry*, *26*, 890–897. doi:10.1097/00004583-198726060-00014

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

Xu, X. (2014). *The establishment of infant and toddler language evaluation tools adapted to family application* (Unpublished master's thesis). Shanghai Jiaotong University, Shanghai, China.

Yi, S. (1995). *Bayley scales of infant development handbook*. Chinese City Revised Ed. Chang Sha: Hunan Medical University Press.

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, *27*, 3–42. doi:10.1017/s030500099900402x

Zimmerman, I.R., Steiner, V.G., & Pond, R.E. (2011). *Preschool language scales*. 5th ed. San Antonio: Pearson Education.