<table>
<thead>
<tr>
<th>Title</th>
<th>Adapting and validating a developmental assessment for Chinese infants and toddlers: The ages &amp; stages questionnaires: Inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Huichao Xie, Jantina Clifford, Jane Squires, Chieh-Yu Chen, Xiaoyan Bian and Qian Yu</td>
</tr>
<tr>
<td>Source</td>
<td>Infant Behavior and Development, 49, 281-295</td>
</tr>
<tr>
<td>Published by</td>
<td>Elsevier</td>
</tr>
</tbody>
</table>

Copyright © 2017 Elsevier

This is the author’s accepted manuscript (post-print) of a work that was accepted for publication in the following source:

http://dx.doi.org/10.1016/j.infbeh.2017.09.004

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source.
Adapting and Validating a Developmental Assessment for Chinese Infants and Toddlers: The Ages & Stages Questionnaires: Inventory

Huichao Xie¹
Jantina Clifford²
Jane Squires²
Chieh-Yu Chen³
Xiaoyan Bian⁴
Qian Yu⁵

2. Early Intervention Program, Department of Special Education and Clinical Sciences, University of Oregon.
3. Department of Special Education, National Taipei University of Education, Taiwan

* Corresponding author: Xiaoyan Bian, M.D. Shanghai Maternal and Child Health Center, 339 Luding Road, Shanghai, China 20062. Email: bxy2017@126.com
Abstract: The Chinese government has announced the 2013 Guidelines for developing a national system for early detection of disability among children under 6 years of age. However, given limited resources, challenges exist with developmental measures required in the 2013 Guidelines. In order to meet the needs for a more accurate and cost-efficient measure for developmental assessment, the Ages & Stages Questionnaires: Inventory (ASQ:I) was translated into Simplified Chinese, and validated on a regional sample of 812 Chinese children ages from 1-25 months. Results indicated that the Chinese ASQ:I domain scores increased across children’s age. When dividing the sample into seven age intervals, Cronbach’s alpha in each interval ranged from .59 to .96 across five domains. When using the whole sample for analyses, item expected a posteriori/plausible value (EAP/PV) reliability was above .99 in all domains, test-retest reliability using intra-class correlation analyses ranged from .95 to .99, and the agreement with the concurrent measure ranged from .29 to .89. Domain scores on the ASQ:I correctly predicted 92-99% of participating children’s disability status. Findings suggested that the Chinese ASQ:I has adequate psychometric properties and thus provides a promising alternative measure for screening and progress monitoring in young children in China. Implications for future research and implementation are discussed.

Keywords: Developmental screening; Progress Monitoring; Cultural adaptation; Psychometric properties; Utility
1. Introduction

According to the most recent government report, the number of young children in China from birth to six years with a documented disability was 1.68 million (China Disabled Persons’ Federation, CDPF, 2006) with an estimated annual increase rate of 0.19 million (Hu and Yang, 2013). Access to public rehabilitation and intervention services requires official documentation of disability (Law of the Protection of Persons with Disabilities, 2008), which usually entails a diagnosis from a qualified physician using standardized procedures and assessment tools. Qualified tests include the Chinese Denver Development Screening Test (DDST) for screening, and the Beijing Gesell Developmental Schedule (Beijing GDS) as a diagnostic tool for documenting intellectual disabilities for young children under the age of six years (Chinese Center for Disease Control and Prevention, CCDCP, 2013).

1.1. Challenges and needs in China

While the acceptance of disability as a natural life occurrence and the awareness of the effectiveness of early intervention services continue to grow, challenges still remain in the Chinese developmental assessment service system. These challenges include 1) insufficient resources to support personnel in the health services industry, 2) a need for high quality developmental assessment instruments that are up-to-date, culturally appropriate, and psychometrically sound, and 3) a lack of family involvement in the evaluation and identification process.

1.1.1. Insufficient resources

In China, assessment of a child’s developmental status and whether he/she is eligible for intervention services is the responsibility of pediatricians. However, a typical pediatric doctor’s visit is allotted only five minutes, due to physicians’ heavy caseloads (Zhou, Pan, & Hou, 2014;
Xu & Zhang, 2014). In addition, most clinical professionals have not received training and are not equipped with the knowledge and skills necessary for the appropriate implementation and administration of developmental screening and diagnostic measures. Training on child development and assessment for medical practitioners is a relatively recent phenomenon in China, having only been provided since 2000 in a few medical schools (Jin, 2010). Given the limited supports and a stressful workload, it is not surprising that the staff turnover rate for pediatricians is increasing in China, a trend which is likely to increase the current shortages and exacerbate the need for these important professionals (Xu & Zhang, 2014). Clearly, the administration of standardized assessments within the daily practices of pediatric professionals is not realistic, given their limited training, high clinical caseloads, and the scarcity of resources in China, especially in rural areas (Luo, Gu, Jin, & Hu, 2014).

1.1.2. Need for psychometrically sound and up-to-date assessment instruments

Currently there are only a few developmental screening and diagnostic instruments available in China, most of which are often outdated, inaccurate, or inaccessible to most practitioners. Many developmental assessment tools used to screen, diagnose, and/or monitor young Chinese children are based on normative samples collected in China 20 or 30 years ago (Lin, Li, & Zhang, 1986; Zhu, Lu, Tang, Wang, & Song, 1983; Zhu, et al., 1984), such as the two tests required in the 2013 CCDCP Guidelines--the Chinese DDST and the Beijing GDS. The Chinese DDST was translated from English and normed on a Chinese sample in the 1970’s and 80’s (Zhou et al., 2013). The Beijing GDS was translated from English and standardized on 884 children in Beijing in 1985 (Lin et al., 1986). Updated normative evidence is much needed, especially as the Chinese population and child rearing practices have undergone many changes in the last 30 years. In addition, there is currently no evidence supporting the technical adequacy of
either the Chinese DDST or the Beijing GDS.

### 1.1.3. Lack of family involvement in the assessment process

The early intervention system in China follows a traditional, professional-centered, clinical approach to the delivery of early intervention services, and a transition to a more family-centered approach is very much needed (Hu & Yang, 2013; Jin, 2008). This traditional approach also exists for developmental assessment services. Identification of developmental problems is often hurriedly conducted by medical professionals during a limited timeframe, and parents and caregivers are traditionally excluded from the decision-making process. Valid and reliable instruments are needed in China that are culturally relevant, family friendly (i.e., include parents in the process), and take minimal time to administer. Using naturalistic tests that incorporate parent observation and report in a reliable manner to collect information can be helpful for yielding accurate understanding of the development of infants and young children (Squires, 2015).

### 1.2. Introducing parent-completed screening tools to China

In order to address some of the shortcomings in the Chinese developmental assessment system described above, one developmental screening tool, the Ages & Stages Questionnaires-Third Edition (ASQ-3; Squires & Bricker, 2009), was translated into Simplified Chinese and a national Chinese normative sample was obtained (Bian, Xie, & Squires, 2014). The ASQ-3 is a parent-completed developmental screening tool that includes 21 questionnaires for specific age intervals from two to 60 months. Thirty scored items within each interval of the ASQ-3 address six age appropriate skills in five developmental domains: Communication, Gross Motor, Fine Motor, Problem Solving, and Personal-social. The Chinese version of the ASQ-3, the ASQ-Chinese (ASQ-C), was studied and published commercially in 2014 (Bian, Chen, & Chen, 2012).
An adaptation of the ASQ-C for the Chinese population was conducted following the six steps based on the *International Test Commission (ITC) Guidelines* (ITC, 2010), including translation into the target language (i.e., Simplified Chinese), translation back to source language (i.e., English), evaluation of the equivalence of the source and target versions, identification and adaptation of incorrectly translated/adapted or culturally and linguistically problematic items, pilot testing and resultant modification, and finally establishment of a normative sample to evaluate the distribution of scores and determine potential cut-off scores for the target population. The ASQ-C was pilot tested on 8,372 subjects in the Shanghai metropolitan area between 2007 and 2008 (Bian, Xie, & Squires, 2014). Results and experiences from the pilot testing facilitated the standardization of the ASQ-C at a national level in 2011 and 2012 (Wei et al., 2015). The ASQ-C sample was stratified on the basis of age, sex, location status (rural/urban), ethnic group, parent education, and annual family income to represent the population of young children ages one to 66 months in China. Psychometric examination with a national sample yielded solid internal consistency (0.51 - 0.68) and inter-rater reliability (0.79 - 0.89), and a convergent validity study indicated 84% agreement between screening categorizations for the ASQ-C compared with the Beijing GDS. Most parents who participated in the survey reported that the questionnaires were easy to understand and helpful to their parenting practices. Within six months of publication in China, the ASQ-C was widely adopted by over 50 child health organizations in 10 provincial regions (Bian, Xie, & Squires, 2014). This popularity was likely due not only to the rigorous evidence for its technical adequacy, but also to the time-efficiency of using a parent-completed protocol.

The ASQ-C, including the high-quality translation and adaptation process, national standardization, establishment of evidence for its technical adequacy, and publication and public
availability of the Chinese version of the ASQ-3 (Bian, Yao, Squires, Hoselton, Chen, & Murphy, 2012), provided a solid foundation for the development and testing of the Chinese version of the Ages & Stages Questionnaires: Inventory (ASQ:I), a recently developed measure based on the entire set of ASQ-3 items. The ASQ:I contains all the items from the ASQ:C and is structured as a longer scale for use with children of a broader age range. The primary purpose of this study was to examine the potential use of the ASQ:I for secondary screening and progress monitoring purposes in a context where assessment resources (e.g., personnel and instruments) are limited.

1.3. The Ages & Stages Questionnaires: Inventory

1.3.1. Original English version

By integrating all ASQ-3 items (within the same developmental area) from the 21 different age intervals into one scale or inventory, a “broader” measure of early childhood skills was created, as opposed to the limited scope of the six developmentally targeted items per domain found in the ASQ-3 (Clifford, 2006). While each domain of the ASQ:I contains 63 to 68 items, only a subset of the items are actually administered by establishing a “floor” and a “ceiling” to identify a child’s developmental repertoire.

Clifford (2006) first examined the use of the ASQ:I as a developmental measure for use with toddlers ages 18 to 36 months in the U.S.. Initial findings provided compelling evidence for the inter-rater reliability between parents and a professional (i.e., a trained research assistant), concurrent validity with the Battelle Developmental Inventory, Second Edition (BDI-2), convergent validity with clinical diagnoses, and adequate item fit using item response theory methods. Results from a utility study also suggested that parents found the ASQ:I easy to complete, were more likely to report benefits than drawbacks from the assessment, and that the
expense for administering the ASQ:I was three to four times less than using a standardized, professionally administered assessment such as the BDI-2. Findings from an additional study confirmed the concurrent validity of the ASQ:I with the BDI-2 as well as the adequacy of item fit (Bae, 2007). In addition, Bae’s study (2007) provided evidence for the test-retest reliability of the ASQ:I.

1.3.2. Simplified Chinese version

The items from the previously adapted ASQ-C (Squires, Bricker, 2009; Bian, Chen & Chen, 2013) were used to develop the Chinese ASQ:I, while applying the item order and format from the English ASQ:I (Edition 2.3 2011, Clifford, Squires, & Bricker, 2011). The same scoring procedures were adopted from the English ASQ:I. Z scores (\(M = 0, SD = 1\)) and percentile scores were established in order to characterize a child’s performance relative to an external reference group.

The adaptation and validation of the ASQ:I in Chinese has the potential to benefit children, families, practitioners, and researchers in both the U.S and China by providing an economical and unique method for evaluating developmental progress and outcomes using a single tool based on parent observation and report. In the U.S., a validated Simplified Chinese version of the ASQ:I could be used for assessing and monitoring child development for children of families whose primary language is Chinese. In China, the dual purposes of the ASQ:I – screening, and progress monitoring – provide promising alternatives to the existing outdated and resource-intensive assessment options.

1.3.2.1. Screening and progress monitoring

In the disability detection system described in the 2013 CCDCP Guidelines, the DDST is required to be used as a secondary, or more in depth, screening procedure. Given the advantages
of the ASQ:I compared to the DDST, this paper suggests using the ASQ:I for secondary screening, especially if the ASQ-C is used as the primary screening instrument. Assessing developmental skills across a wide age range (i.e., from 1 to 66 months), the ASQ:I can provide more comprehensive information about a child’s development than many instruments designed for brief screening such as the ASQ-C and the Chinese DDST. Given the popularity of the ASQ-C in China, using the ASQ:I for one or multiple times of follow up assessment may be especially efficient when an initial screening was completed with the ASQ-C. Since the ASQ:I is made up of all of the items from the ASQ-3, assessment with the ASQ:I can build on information gathered from the ASQ-C completed at the initial screening, saving valuable time. When monitoring child progress, the ASQ:I also allows documenting the assessment results from multiple times of administration to describe a child’s developmental trajectory. Parents who complete the ASQ-C prior to completing the ASQ:I will also be familiar with the types of questions asked, and will likely observe their child more closely after having completed an ASQ-C, which may improve their accuracy. Percentile scores and z scores generated from the ASQ:I can then be used to either confirm or reject the need for eligibility evaluation. Using the ASQ:I as a secondary screening instrument will not only support more accurate referral for eligibility evaluations, but may also provide parents an opportunity to learn more about the current skills and next steps in their child’s development.

1.3.2.2. Comparing the ASQ:I with existing Chinese assessment tools

Compared to the Beijing GDS, the ASQ:I has six advantages. First, the ASQ:I has updated items. ASQ:I items are integrated from the ASQ-C which was recently updated in 2009, while the Beijing GDS is based on the 1974 version of the American GDS. Second, the ASQ:I has evidence that suggests it yields valid and reliable scores. The Beijing GDS provides
developmental quotients calculated with developmental equivalents and a child’s chronological age (Lin et al., 1986), which could cause problems such as misinterpretations and therefore “should never be used” (Salvia et al., 2013, p. 47). The ASQ:I results are expressed in standard and percentile scores that provide more accurate and helpful information regarding a child’s developmental status compared to his or her same aged-peers. Third, the translation/adaptation process of the items in the ASQ:I was undertaken according to recommended practices in the current literature, following the six steps in the *ITC Guidelines* (ITC, 2010; Bian et al., 2014). Currently, no evidence for the translation equivalence of the Beijing GDS is reported in the literature. Fourth, the ASQ:I is an authentic assessment that is intended to be administered using naturalistic methods. The Beijing GDS requires direct testing of the child by professionals using standardized materials and instructions, while the ASQ:I allows collecting information by interviewing the parents, as well as observing and interacting with the child in natural settings. Fifth, the requirements of personnel for administering the ASQ:I are more flexible. The Beijing GDS requires qualified physicians, while the ASQ:I can be administered by a variety of practitioners including but not limited to physicians, nurses, medical assistants, social workers, and early childhood educators. Sixth, the ASQ:I takes potentially less administration time to administer. The completion of the Beijing BDS is reported to take 40 to 120 minutes (Liang & Zhu, in process), while the ASQ:I takes 20 to 60 minutes when administered by parents (Clifford, 2006). Table 1 compares features of the Chinese DDST, Beijing GDS, and the ASQ:I.

**Table 1.** *Chinese Denver Development Screening Test, Beijing Gesell Developmental Schedule, and Ages & Stages Questionnaires: Inventory*

<table>
<thead>
<tr>
<th>Features</th>
<th>Denver Development Screening Test</th>
<th>Beijing Gesell Developmental Schedule</th>
<th>Ages &amp; Stages Questionnaires: Inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of item development/ refinement</td>
<td>1967</td>
<td>1974</td>
<td>2009</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Purpose of assessment</td>
<td>screening</td>
<td>diagnostic</td>
<td>progress monitoring and determination of developmental status</td>
</tr>
<tr>
<td>Domains</td>
<td>adaptive/fine motor, gross motor, language, personal-social</td>
<td>adaptive behavior, gross motor, fine motor, language, personal-social</td>
<td>communication, gross motor, fine motor, problem solving, personal-social</td>
</tr>
<tr>
<td>Normative data</td>
<td>regional data collected from 1970’s to 1980’s</td>
<td>regional data collected in 1985</td>
<td>regional data were collected in 2015-2016</td>
</tr>
<tr>
<td>Administration requirements</td>
<td>standardized materials, instructions, and procedures</td>
<td>standardized materials, instructions, and procedures</td>
<td>interview parents and observe child in natural setting</td>
</tr>
<tr>
<td>Types of scores</td>
<td>developmental equivalents, developmental quotients</td>
<td>developmental equivalents, developmental quotients</td>
<td>standard scores, percentage scores (raw score divided by maximum possible score)</td>
</tr>
<tr>
<td>Personnel requirements</td>
<td>trained professionals</td>
<td>qualified physicians</td>
<td>trained professionals</td>
</tr>
<tr>
<td>Translation quality</td>
<td>not found in current literature</td>
<td>not found in current literature</td>
<td>Bian et al., 2014</td>
</tr>
<tr>
<td>Psychometric properties of Chinese version</td>
<td>not found in current literature</td>
<td>not found in current literature</td>
<td>Chen, 2013 (traditional Chinese version); the Simplified version is examined in this study</td>
</tr>
<tr>
<td>Time</td>
<td>10-20 minutes</td>
<td>40-120 minutes</td>
<td>15-70 minutes</td>
</tr>
</tbody>
</table>

1.4. Research purpose

As such, the adaptation of a valid and reliable instrument is critical for developmental screening and monitoring in China. The psychometric properties of the ASQ:I address needed requirements for assessing young children. The five domains assessed by the ASQ:I align with
the requirements in the Chinese legislation and the common understanding of early childhood development in the Chinese literature. Considering the limited resources available for assessment, the ASQ:I is a promising instrument for China and other countries where assessment measures may be outdated and/or resources for developmental assessment are minimal. Initial unpublished studies suggest that the ASQ:I yields results that are accurate, such as standard scores and percentile scores. Additionally, the naturalistic formats of administration of the ASQ:I (e.g., interview with parents, observing and interacting with the child with assistance from parents) allow collecting authentic information about a child as well as promoting professional-parent collaboration throughout the assessment (de Sam Lazaro, 2017). This study focused on the test adequacy of a Chinese adaptation of the ASQ:I, examining reliability, validity, and utility.

2. Method

2.1. Participants and setting

Data were collected in Kunshan, a city on the east coast of China with a population of 1.6 million.

2.1.1. Children and caregivers

Eight hundred and twelve children and their caregivers participated in the study. Two different samples were recruited for the study, including typically developing children and children with diagnosed special needs or a disability. Table 2 lists the demographic information of children and caregivers in this study. Recruitment efforts and eligibility criteria differed for each of the samples and are described below.

Table 2. *Demographic Characteristics of Participating Children and Families (N = 812)*
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Number (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Child Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>426 (52.5%)</td>
</tr>
<tr>
<td>Female</td>
<td>386 (47.5%)</td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0%)</td>
</tr>
<tr>
<td><strong>Mother’s education</strong></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>205 (25.2%)</td>
</tr>
<tr>
<td>High school</td>
<td>168 (20.7%)</td>
</tr>
<tr>
<td>AA degree</td>
<td>266 (32.8%)</td>
</tr>
<tr>
<td>Four-year college/above</td>
<td>161 (19.8%)</td>
</tr>
<tr>
<td>Don’t know</td>
<td>1 (0.1%)</td>
</tr>
<tr>
<td>Missing</td>
<td>11 (1.4%)</td>
</tr>
<tr>
<td><strong>Annual Family Income (Chinese RMB)</strong></td>
<td></td>
</tr>
<tr>
<td>0-12,000</td>
<td>44 (5.4%)</td>
</tr>
<tr>
<td>12,001-24,000</td>
<td>167 (20.6%)</td>
</tr>
<tr>
<td>24,001-40,000</td>
<td>262 (32.3%)</td>
</tr>
<tr>
<td>Over 40,000</td>
<td>326 (40.1%)</td>
</tr>
<tr>
<td>Missing</td>
<td>13 (1.6%)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Chinese Han</td>
<td>733 (90.3%)</td>
</tr>
<tr>
<td>Others</td>
<td>13 (1.6%)</td>
</tr>
<tr>
<td>Missing</td>
<td>66 (8.1%)</td>
</tr>
<tr>
<td><strong>Geographical region</strong></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>424 (52.2%)</td>
</tr>
<tr>
<td>Rural</td>
<td>173 (21.3%)</td>
</tr>
<tr>
<td>Missing</td>
<td>215 (26.5%)</td>
</tr>
<tr>
<td><strong>Special needs</strong></td>
<td></td>
</tr>
<tr>
<td>No special needs reported/documentated</td>
<td>749 (92.2%)</td>
</tr>
<tr>
<td>(included in the normative sample)</td>
<td></td>
</tr>
<tr>
<td>Parent reported special needs</td>
<td>25 (3.1%)</td>
</tr>
<tr>
<td>(included in the normative sample)</td>
<td></td>
</tr>
<tr>
<td>Medical record showed special needs</td>
<td>38 (4.7%)</td>
</tr>
<tr>
<td>(excluded from the normative sample)</td>
<td></td>
</tr>
</tbody>
</table>

**2.1.1.1. Typically developing children**

Children and their caregivers were eligible to participate in the study if the children were between the ages of one to 25 months, at least one parent was a Chinese citizen, and the family spoke Chinese. A systematic sample with a random start (Babbie, 2012) was used to recruit
participants during routine well-child checkup visits at four different health clinics. At each clinic, a random number from one to 10 was selected to indicate when to start sampling. For example, a randomly generated number of four indicated that every 4th visitor one in every four eligible visitors would receive a recruitment flyer during the check-in process. As a result, a total of 774 children and their caregivers participated from the four clinics, with sample sizes for each ASQ-C age interval ranging from 38 to 107 ($M = 67$).

2.1.1.2. Children with an identified disability.

Children between the ages of 17 to 20 months with an identified disability were also recruited. Using child health records from the city of Kunshan, 437 potential participants were identified and contacted via phone. Thirty-eight parents (8.7%) agreed to participate in phone interviews to complete the ASQ:I.

2.1.2. Caregivers

A majority of caregivers were mothers ($n = 501, 62\%$), followed by grandparents ($n = 190, 24\%$) and fathers ($n = 75, 9\%$). In terms of mother’s education level, more than half had a college degree (i.e., associate, undergraduate, or graduate). Forty percent ($n = 326$) of families earned an annual income of more than 40,000 Chinese RMB. Most Chinese children ($n = 733, 90\%$) were reported as belonging to the Chinese Han ethnic group.

2.1.3. Pediatricians

The director of the pediatric health care system in Kunshan was consulted to identify representative sites to participate in the study. Based on the director’s recommendation, four clinics from different regions of the city were included. Sixteen pediatricians were invited by emails and phone calls to participate in the study. Participation criteria for pediatricians included regular or recent use of developmental assessments with infants and young children under the
age of three years. All 16 pediatricians who were invited agreed to participate in the study. There were 15 females and one male, between the ages of 27 to 42 years ($M = 34$, $SD = 5.4$). Fourteen of the pediatricians held a bachelor’s degree, one had an associate degree in medical science, and one did not report their educational background. On average, the pediatricians reported having eight years of practice in well-child health care, ranging from two to 24 years and fewer years of experiences in providing child developmental assessment services, ranging from six months to 10 years, with a mean of 4.5 years.

All pediatricians in this study attended two training sessions with a participant rate of 100%. In the first training session, the first author provided a 2-hour introduction to the development, structure, and scoring procedures of the ASQ:I, followed by 4-hour demonstration and in-class practice of completing the assessment. Pediatricians were given the protocols to read and practice for eight weeks. After that, the fourth author conducted an 8-hour second training session. Pediatricians worked in small groups of 4 to complete the protocols based on written scenarios. An inter-rater agreement of 85% or above of all items was required to be qualified as data collectors in the study.

2.2. Measures

2.2.1. Demographic form

Parents of participating children were asked to provide background information about their child, family, and themselves. Children’s date of birth, prematurity status (i.e., less than 37 weeks’ gestation), gender, disability status, and early intervention or rehabilitation services received (if applicable) were included. Questions pertaining to caregivers and families included their relationship to the child, education level, ethnicity, mother’s age at child’s birth, primary caregiver at home, family registration status (i.e., urban or rural), and annual family income.

2.2.2. Ages & Stages Questionnaires: Inventory, Chinese version
A Chinese version of the ASQ:I was used to collect data on children’s developmental skills. The original English ASQ:I is a developmental measure for children from ages one to 36 months (Clifford, 2006). There are 63 to 68 items in each of five developmental domains (i.e., Communication, Gross Motor, Fine Motor, Problem Solving, and Personal-Social). The items in each domain are organized in a developmental structure, ordered from “easy” to “difficult.” That means, an item that appears early on in the scale targets skills that a child usually acquires early in life. For example, the first item in Communication, “does your baby sometimes make throaty or gurgling sounds,” describes a behavior that is common for very young babies. There are three choices for rating each item, “yes” (2 points), “sometimes” (1 point), and “not yet” (0 points).

Age-appropriate start points are suggested to guide the administrator in identifying an appropriate point to begin the assessment. For example, in the Communication domain, it is suggested that administrators begin the assessment at item #11 for infants from nine up to 12 months, and at item #15 for infants from 12 up to 18 months. These start points were selected by the developers by identifying the items where at least 70% of children in a particular age interval (i.e., 12-18 months) received a score of “2” based on data from previous unpublished studies.

Basal and ceiling rules are followed to restrict the number of test items completed. A basal is established when four consecutive items are answered “yes”. Failure to establish an immediate basal with the first four items requires administering items before the start point in reverse order until a basal is established with earlier items. A ceiling is established after four consecutive items are answered as “not yet”. Unanswered items occurring before the basal are automatically assumed to be “yes” (i.e., 2 points), and items occurring after the ceiling are automatically answered as “not yet” (i.e., 0 points). Figure 1 provides an example of an established ceiling in the fine motor domain. Based on the developmental structure of the ASQ:I,
items after item #15 (the “ceiling”) are considered too advanced for the child and will be scored “not yet” (0 points).

11. Does your baby pick up a small toy with only one hand? 6-6

12. Does your baby reach for or grasp a toy using both hands at once? 2-6

13. Does your baby pick up a small toy, holding it in the center of her hand with her fingers around it? 4-6

9 up to 12 month babies start here:

14. Does your baby reach for a crumb or Cheerio and touch it with his finger or hand? (If he already picks up a small object the size of a pea, mark “yes” for this item.) 3-6

15. Does your baby try to pick up a crumb or Cheerio by using his thumb and all his fingers in a raking motion, even if he isn’t able to pick it up? (If he already picks up the crumb or Cheerio, mark “yes” for this item.) 5-6

Ceiling is established

Figure 1. An example of the establishment of ceiling when four consecutive items scored zero on the ASQ:I.

Depending on the number of items (i.e., 63-68), each domain has a maximum total score ranging from 126 to 136 as the highest possible score for each item is two points. Upon completion, scores are summed for each domain, and then total scores are divided by the maximum total score to generate a percentage score. For example, the maximum total score in the Communication domain is 130 (2 points times 65 items). A child’s Communication score of 87 at the first administration of the ASQ:I is converted to a percentage score of 67%. After 12 months, the same child received a Communication score of 112, which is 86%. An increase from 67% to 86% indicates that the child developed new communication skills in the 12 months between two administrations.

The current study used a paper-pencil format of the ASQ:I. Pediatricians administered the items by asking parents questions from the ASQ:I. Parent report was relied on for skills that
parents were sure they had seen their child perform on a regular basis. When parents were unsure if their child was able to perform a skill, the pediatricians assisted the parents in trying the item with their child.

As mentioned earlier, the Chinese ASQ:I contains the entire item pool from the ASQ-C, which has been examined for cultural equivalence and appropriateness on the Chinese population (Bian et al., 2012). In addition to the ASQ-C items, there are 65 additional items on the ASQ:I that were developed to provide more developmentally advanced skills to address the developmental repertoire of precocious two-year-olds in order to avoid a ceiling affect. These additional items were translated by the principal researcher who is proficient in both English and Chinese. A back-translation procedure as recommended in the ITC Guidelines (ITC, 2010) was used to examine the equivalence of the two language versions. The translation and back-translation procedures resulted in 29 out of the 65 new items being adapted in three ways. Table 3 addresses the three types of cultural adaptation that were made, indicates the number of items that were adapted, and provides an example as an illustration for each type of adaptation.

<table>
<thead>
<tr>
<th>Type of adaptation</th>
<th>Number of items</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptation of the measurement system</td>
<td>10</td>
<td>Item 51 in Gross Motor domain, “Does your child hop on one foot for a distance of 2 feet?” was translated as “孩子能单脚不停跳着前进至少 60 厘米的距离吗 (Does your child hop on one foot for a distance of 60 centimeters)”?</td>
</tr>
</tbody>
</table>
| Adaptation of the examples   | 13             | Item 56 in Fine Motor domain, “Does your child cut up soft food into smaller pieces using a dull knife? For example, can your child use a butter knife to cut bananas or mangos?” was translated as “孩子能用一把钝刀切东西吗? 例如, 用塑料刀切橡皮泥, 或用餐刀切香蕉或芒果 (Does your child cut up soft food into smaller pieces using a dull knife? For example, can your child use a plastic toy knife to cut playdough, or use a butter knife to cut bananas or mangos)”?
| Adaptation of the voice     | 16             | Item 47 in Communication domain, “Does your child make her voice go high at the end of a sentence that is a question?” was            |
targeted skill translated as “孩子会在问句结尾使用“吗”或“啊”表示提问吗? (Does your child use the correct particle such as “ma” or “a” at the end of a sentence that is a question)”? 

2.2.3. Beijing Gesell Developmental Schedule

In order to examine the concurrent validity of the ASQ:I, a Chinese version of the Beijing GDS was used as a criterion measure to determine a child’s ability status in the subsample of 53 children from 11 to 12 months old. As previously mentioned, the Beijing GDS is an individually administered standardized, norm-referenced assessment for children from birth through six years that is commonly used for diagnosing intellectual disabilities in China (Liu, Chen, & Zhao, 2014). The assessment includes more than 500 items in five developmental domains: adaptive, gross motor, fine motor, language, and personal-social, and requires trained professionals to use clinical observation in combination with caregiver interviews. Items in each domain are assigned an age equivalent based on normative data; a child’s performance on each item is scored as “pass” or “no pass”, and a domain age equivalent score is determined by calculating the number of “pass” items at each age interval. A developmental quotient (DQ) is derived by dividing a developmental age based on a child’s performance on the test by the child’s chronological age and then multiplying the result by 100. This DQ is then used to determine the child’s developmental status. In this study, the Beijing GDS was used as a comparison measure because it is the only officially sanctioned eligibility assessment for your young children in China, and is thus required to be used by pediatricians during eligibility determinations.

2.2.4. Utility survey

A six-question utility questionnaire was developed to survey the parents/caregivers about their experience using the ASQ:I. Parents were asked to estimate how long it took to complete the ASQ:I with their child, if the parent had experiences with other developmental assessment
tools, and if they did, to compare the experiences with the ASQ:I and the other tool(s). Questions three, four, five, and six used a five-point Likert scale in order for parents to rate if and how much the ASQ:I helped them to understand their child’s development, whether the skills and behaviors measured on the ASQ:I were critical and developmentally appropriate for their child, whether they felt respected and involved in the assessment process with the ASQ:I, and whether they would recommend the ASQ:I to other parents who have concerns about their child’s development. Twenty-one percent (169) of the parents completed the survey.

2.2.5. Pediatrician interview script

A script was developed for conducting semi-structured interviews with the pediatricians who collected data on the ASQ:I in this study. The interview examined to what extent the ASQ:I met the needs for early childhood developmental assessment in China. The script contained a demographic section about the pediatrician’s professional background, such as their academic degree and major, hospital/clinic for practice, and how many years they had been providing child health care and developmental assessment services. The second section of the interview addressed questions regarding the pediatricians’ experience with the ASQ:I. The pediatricians were asked to report the approximate time it took to administer the ASQ:I and to compare the ASQ:I with other developmental assessments they had used in their practice. In addition, four open-ended questions asked the pediatricians to evaluate whether the ASQ:I targeted the most critical skills and behaviors in child development, to reflect on the format of the protocols and the administration process, and to comment on the utility of the ASQ:I assessment results in informing screening and diagnostic decisions in their practice.

2.3. Research design and analyses

This study applied a non-experimental psychometric design to explore the technical
adequacy of the ASQ:I. SPSS software version 18.0 was used to analyze the data.

### 2.3.1. Descriptive statistics

In order to examine the developmental structure of the ASQ:I, scores in the five domains were analyzed using a quantile regression model to describe how the scores changed from younger to older children. As children’s age increases, they are expected to perform more skills therefore obtain higher scores on the ASQ:I. Corresponding percentile scores across 12 different age intervals (e.g., 1-month 0 days to 2-month 30 days, 3-month 0 days to 4-month 30 days, etc.) were graphed with trend lines. The polynomial trend lines were generated from the following regression equation: $y = b + c_1 x + c_2 x^2 + c_3 x^3 + \cdots + c_n x^n$

### 2.3.2. Reliability

Domain scores on the ASQ:I were analyzed to measure two aspects of reliability. First, internal consistency was examined with Cronbach’s Alpha on the item scores within each domain (Bailey, 2004). The total sample of 774 children was divided into seven age intervals. The Cronbach’s Alpha reliability test was performed for each age interval.

Second, item expected a posteriori/plausible value (EAP/PV) reliability was examined across five ASQ:I domains. Item response model analyses using a Rasch one-parameter partial credit model were conducted in ConQuest 4.4 (Adams, Wu, & Wilson, 2015). The EAP/PV reliability analyses indicate the ratio of modeled variance to observed variance (Scalise, 2012). In other words, EAP/PV reliability results show the replicability of person placement across a set of items measuring the same construct (Bond & Fox, 2015). Given the sample size, the EAP/PV reliability analyses were performed on the whole sample ($N = 774$) without dividing into age intervals.

Third, test-retest reliability was examined using two different analytic approaches
between the first and second administration of the ASQ:I across a one-week period. The first approach used the Pearson’s product moment correlation coefficient, which examined the linear correlation between the scores from two administrations. The second approach applied an absolute agreement intra-class correlation coefficient, which measures the strength of the agreement between scores across two administrations of the same tool. The absolute agreement coefficient is considered to be a stronger indicator when evaluating the reliability of a measure (McGraw & Wong, 1996).

2.3.3. Validity

Validity studies for the ASQ:I focused on concurrent and known groups validity.

2.3.3.1. Concurrent validity

After completing the ASQ:I, caregivers of children from the age of 11 to 12 months were given a choice to complete the Beijing GDS. Fifty-three (52%) children in this age interval participated; the agreement between the developmental quotient scores on the Beijing GDS and the z scores on the ASQ:I was examined using Pearson product-moment correlation coefficients (Bailey, 2004). Correlational analyses were conducted on corresponding domain scores of the two measures. Specifically, scores in the Language, Gross Motor, Fine Motor, and Personal-Social domains on the Beijing GDS were compared to scores in the Communication, Gross Motor, Fine Motor, and Personal-Social domains on the ASQ:I. Scores in the Adaptive domain on the Beijing GDS were compared to scores in the Problem Solving domain on the ASQ:I because of the high proportion of items targeting similar or same skills.

2.3.3.2. Known-groups validity

Known-groups validity was tested with a sub-group of children between the ages of 17 and 20 months, using a logistic regression model to examine the relationship between the ASQ:I
domain scores and the child’s disability status (i.e., disability, typically developing) documented in their medical records. In the logistic regression analysis, the child’s disability status served as a two-level dependent variable – yes (i.e., the child has been identified with a disability) and no (i.e., the child has been identified as typical). Each domain raw score on the ASQ:I served as the predictor or independent variable. At first, children’s age was also included as one of the two predictors (i.e., age and ASQ:I domain score) in the regression model. Preliminary results indicated that age did not make a statistically significant impact on children’s disability status ($p > .05$), therefore it was removed from the model, leaving ASQ:I domain score as the only predictor.

2.3.4. Utility

Quantitative responses from caregivers and pediatricians were analyzed using descriptive statistics. Qualitative comments were summarized and interpreted for major patterns, relevance, salience, threads, and emergent themes (Creswell & Clark, 2011) by the principal researcher. Results from both the parent survey and the pediatrician interview (e.g., administration time) were integrated to inform further interpretation.

3. Results

3.1. Age trends

Using a quantile regression approach, the polynomial trend lines were estimated to display trends in percentile scores (i.e., 10th-, 25th-, 50th-, 75th-, and 90th-percentile) from five domains of the ASQ:I, as shown in Figures 2 to 6.

The five trend lines in each domain represent the increasing trend of the ASQ:I scores in relationship to child age. The $r^2$ value of the trend lines ranged from .97 to .99, indicating good representation of the data.
3.2. Reliability

3.2.1. Internal consistency

Item level data from the Chinese normative sample \((n = 774)\) were divided into seven age
intervals, analyses were conducted in each age interval. Cronbach’s alpha ranged from .59 to .96, as shown in Table 4. Internal consistency results in the youngest two age intervals (i.e., 1-2 months and 3-5 months) were generally lower than in other age intervals. Across five domains, results indicated relatively higher internal consistency in Communication and Fine Motor domains.

### Table 4. Cronbach’s Alpha and sample size in each Age Interval across the ASQ:I Domains

<table>
<thead>
<tr>
<th>Age interval</th>
<th>Sample size</th>
<th>Communication</th>
<th>Gross Motor</th>
<th>Fine Motor</th>
<th>Problem Solving</th>
<th>Personal-Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2 months</td>
<td>95</td>
<td>.96</td>
<td>.69</td>
<td>.96</td>
<td>.59</td>
<td>.70</td>
</tr>
<tr>
<td>3-5 months</td>
<td>118</td>
<td>.94</td>
<td>.70</td>
<td>.77</td>
<td>.79</td>
<td>.68</td>
</tr>
<tr>
<td>6-8 months</td>
<td>112</td>
<td>.93</td>
<td>.77</td>
<td>.83</td>
<td>.84</td>
<td>.79</td>
</tr>
<tr>
<td>9-11 months</td>
<td>56</td>
<td>.92</td>
<td>.86</td>
<td>.78</td>
<td>.76</td>
<td>.80</td>
</tr>
<tr>
<td>12-17 months</td>
<td>197</td>
<td>.93</td>
<td>.91</td>
<td>.83</td>
<td>.89</td>
<td>.88</td>
</tr>
<tr>
<td>18-23 months</td>
<td>141</td>
<td>.95</td>
<td>.95</td>
<td>.94</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td>24-25 months</td>
<td>55</td>
<td>.95</td>
<td>.79</td>
<td>.73</td>
<td>.76</td>
<td>.83</td>
</tr>
</tbody>
</table>

#### 3.2.2. Expected a posteriori/plausible value reliability (EAP/PV)

Results from every domain showed EAP/PV reliability results approximate to 1. Some items were automatically excluded by the ConQuest program for lack of variability (e.g., all responses are “0”). This resulted in 43 to 55 items included in the analyses for each of the 5 domains.

#### 3.2.3. Test-retest reliability

Results from the Pearson’s product moment correlation coefficients and the absolute agreement intra-class correlation coefficients between the two administrations of the ASQ:I at one to three week intervals are listed in Table 5.

Significantly high correlations were found in all five domains, with Pearson correlation coefficients ranging from .91 to .99 ($p < .001$). Similar results were found from the intra-class
correlational analyses. Absolute agreement correlation coefficients between the two administrations ranged from .95 to .99 (p < .001), indicating significantly high agreements.

Table 5. Test-Retest Reliability: Correlations of Domain Scores on ASQ:I between Two Times of Administration (n = 23)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Pearson’s product moment correlation coefficients</th>
<th>Absolute agreement intra-class correlation coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Communication</td>
<td>42.13</td>
<td>23.32</td>
<td>44.00</td>
<td>22.80</td>
</tr>
<tr>
<td>Gross Motor</td>
<td>41.43</td>
<td>21.72</td>
<td>42.61</td>
<td>21.46</td>
</tr>
<tr>
<td>Fine Motor</td>
<td>39.09</td>
<td>18.47</td>
<td>41.78</td>
<td>18.94</td>
</tr>
<tr>
<td>Problem Solving</td>
<td>40.04</td>
<td>23.55</td>
<td>42.48</td>
<td>22.92</td>
</tr>
<tr>
<td>Personal-Social</td>
<td>40.09</td>
<td>24.75</td>
<td>42.48</td>
<td>24.66</td>
</tr>
</tbody>
</table>

*** p < .001

3.3. Validity

3.3.1. Concurrent validity

In addition to completing the ASQ:I, a subgroup of 53 children from 11 to 12 months old (52.4%) also participated in a standardized assessment procedure using the Beijing GDS as the concurrent measure. The DQ scores on the Beijing GDS were calculated by dividing the developmental age of a child by their chronological age. In order to compare DQ scores from the Beijing GDS with results from the ASQ:I, z-scores for each domain of the ASQ:I were computed based on all 101 children from 11 to 12 months old in the sample. Pearson product moment correlation coefficients indicated that four comparison pairs (i.e., domains) were highly correlated at the .01 level, while the Communication domain in the ASQ:I and the Language domain in the Beijing GDS showed a relatively low correlation of .29, yet still significant at the .05 level. Results from the correlational analyses between the two measures are presented in Table 6.

Table 6. Correlations between ASQ:I and Beijing GDS Scores (n = 53)
3.3.2. Known-groups validity

The ASQ:I domain scores were found to significantly predict children’s disability status (i.e., disability, typically developing) in all five domains. As shown in Table 7, all five odds ratios were less than one, indicated that children with higher scores on the ASQ:I were less likely to be placed in the disability group than children with lower scores. The ASQ:I scores correctly predicted 92%-99% of children about their disability status.

Table 7. Adjusted odds ratio and 95% confidence interval for using each ASQ:I domain score to predict disability status (n = 82)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Adjusted odds ratio</th>
<th>95% Confidence interval</th>
<th>Percent of correct predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>0.73***</td>
<td>0.63-0.84</td>
<td>92%</td>
</tr>
<tr>
<td>Gross motor</td>
<td>0.69**</td>
<td>0.56-0.85</td>
<td>99%</td>
</tr>
<tr>
<td>Fine motor</td>
<td>0.55**</td>
<td>0.39-0.78</td>
<td>95%</td>
</tr>
<tr>
<td>Problem solving</td>
<td>0.66**</td>
<td>0.52-0.84</td>
<td>95%</td>
</tr>
<tr>
<td>Personal-social</td>
<td>0.73***</td>
<td>0.62-0.85</td>
<td>93%</td>
</tr>
</tbody>
</table>

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

3.3.3. Utility

3.3.3.1. Caregiver survey

One hundred and sixty-nine out of the 774 caregiver participants completed the six-
question utility survey. Caregivers reported spending 10 to 70 minutes to complete all five domains of the ASQ:1 on their child, with a mean of 33 minutes and a standard deviation of 13. Administration time tended to increase in direct correlation with child age ($r = .91, p < .01$).

A majority of caregivers (93%) reported that they had no experience with child developmental assessment other than using the ASQ:1 in this study. The majority of caregivers (97%) indicated that they “agreed” or “strongly agreed” with the response to the statement “I feel that I understand my child’s development better after participating in the ASQ:1 assessment”. Most caregivers (96%) also agreed with the statement, “I think the skills described in the ASQ:1 items are critical and representative of my child’s age”. When asked whether they felt that their opinions about their child’s development were given serious consideration during the assessment, a majority (98%) agreed or strongly agreed. Most caregivers (96%) also indicated that they would recommend the ASQ:1 to other caregivers who have concerns about their child’s development.

3.3.3.2. Pediatrician interview

Fifteen of the 16 pediatricians who used the ASQ:1 participated in one-on-one interviews with the lead researcher. Pediatricians were asked to report the advantages, disadvantages, and general usefulness of the ASQ:1. Qualitative responses from all 15 pediatricians were summarized and grouped into four themes related to advantages, and four to disadvantages.

For advantages, 11 pediatricians commented that the ASQ:1 collected comprehensive information about a child’s development which is helpful for identifying children at risk in their daily clinical practices. Eight pediatricians reported that the tool was parent friendly because the test items were easy to understand, involving, and educative for parents. Four pediatricians reported convenience as an advantage because the ASQ:1 took less time to administer than other
diagnostic assessment tools (such as the Beijing GDS). Three pediatricians also reported that the children seemed to enjoy the activities that they were encouraged to try during administration of the ASQ:I.

For disadvantages, 6 pediatricians reported that the time needed to administer the ASQ:I would be a challenge if they were to implement it in their daily practices, especially when using the tool with children older than 24 months. All 15 pediatricians reported that the older or the more developmentally advanced the child was, the longer it took to complete the ASQ:I. On average, pediatricians reported that it took about 16 to 28 minutes to assess babies 12 months old, and 33 to 63 minutes for children 14 months and older. Two pediatricians reported minor differences in the wording between the ASQ:I and ASQ-C, which was due to different structures and scoring procedures of the ASQ:I. Two pediatricians reported concerns on collecting information from parent report, suggesting that parents might rate their children based on their expectations rather than on observations. In addition, two other pediatricians questioned the accuracy of the information when relying solely on parent report, and three pediatricians also expressed concern about collaborating with caregivers who have low education levels to complete the assessment. As commented by one of the pediatricians, “Some parents just say ‘yes’ to every question, but their child did not demonstrate even the emerging behaviors of the targeted skill during the assessment.”

In terms of the usefulness of the ASQ:I, 14 of the 15 pediatricians responded positively (e.g., “definitely”, “I think so”, “overall, yes”) to the question, “Do the ASQ:I items represent the critical skills in each stage of development?” One pediatrician elected not to respond to this question as she felt she did not have enough experience with developmental assessment to make a determination regarding critical skills. All 15 pediatricians reported that the ASQ:I could be
helpful in their well-child health care practice as a secondary screening tool and for diagnosing developmental delays. Seven of them further commented on the need for innovations in the current infrastructure of the child health care system in order to provide the resources (e.g. time, personnel) for administering diagnostic tools.

4. Discussion

4.1. Interpretation of findings

4.1.1. Chinese regional normative sample

The regional sample of 774 children collected in Kunshan city is relatively small in size, especially in some age intervals (e.g., \( n = 56 \) in 9-11-months). The gender ratio of the normative sample was similar to that in the local population (i.e., 51% male and 49% female), based on the most recent Kunshan census data (City Bureau of Statistics of Kunshan, 2014). However, the education level of the mothers in the sample was higher than in the general population, as 52.7% of the mothers reported attaining the equivalent of an AA degree or above, compared to 15% in the local population (City Bureau of Statistics of Kunshan, 2014). While the reason for this finding is unclear, one possible explanation might be that mothers with higher educational attainment may be more likely to take their children to well-child check-ups at the community child health clinics, and would therefore be more likely to be recruited. Additionally, the format of recruitment (i.e., distribution of a one-page recruitment flyer to caregivers) might also have been a barrier to some mothers with lower levels of education. Finally, it is possible that well-educated mothers were more likely to sign up to participate in the study because they were more comfortable with or cognizant of the importance of developmental assessment.

4.1.2. Development hierarchy

The percentile score trend lines from the quantile regression analysis showed that ASQ:I
domain scores increased as the children’s ages increased (see Figures 2 to 6). These results provided evidence for the developmental structure of the ASQ:I. Specifically, older children seemed to have developed more skills than younger children did. In addition, none of the 90th percentile trend lines reached the maximum possible scores in each domain, indicating that the ASQ:I items covered an adequate range of development, even for the most developmentally advanced children in the sample.

4.1.3. Reliability

The Cronbach’s alpha coefficients in each of the seven age intervals ranged from .59 to .96, generally higher than reported on the ASQ-C (Bian et al., 2012). In each of the 35 age interval by domain groups, 32 (91%) showed an acceptable internal consistency above .70. The relatively fewer (e.g., less than 10) items administered before reaching the ceiling point in the 1-2 months and 3-5 months intervals might have resulted in comparatively poorer internal consistency (.59, .68 and .69).

Results from EAP/PV reliability analyses using a one-parameter partial credit Rasch model indicated high level of consistency among the ASQ:I items, which was not surprising given the large number ($M = 50.2$) of questions included in the analyses within each domain. Future research is needed to collect a larger sample size to allow breaking down of age intervals as well as test items into subscales to examine EAP/PV in each age interval.

Test-retest reliability was examined for a subsample of 23 children by administering the ASQ:I twice within a brief timeframe of one to three weeks. Results from the Pearson and Absolute agreement intra-class correlational analyses indicated significant agreement between scores from both administrations, demonstrating that the ASQ:I results were consistent across time.
4.1.4. Concurrent validity

Concurrent validity was examined by administering the Beijing GDS along with the ASQ:I for 53 one-year-old children (i.e., 11 to 12 months). Strong and statistically significant correlations ranging from .74 to .89 were obtained in the Gross Motor, Fine Motor, Problem-Solving/Adaptive, and Personal-Social domains, indicating that the items measured very similar skills on the two tests. A relatively lower correlation, yet still significant, was found in the Communication/Language domains. A possible explanation was that items in the Communication domain of ASQ:I were measuring somewhat different behaviors than measured on the Language domain in Beijing GDS. For example, the behavior described in Communication item #19 in the ASQ:I, “Does your child point to, pat, or try to pick up pictures in a book?” was not included in the Beijing GDS. In order to further examine the construct validity of the ASQ:I, more research is needed to investigate specific targeted skills on the two measures. Findings from this study indicated good concurrent validity for the ASQ:I with the Beijing GDS in four domains (i.e., Gross Motor, Fine Motor, Problem-Solving, and Personal-Social), while scores in the Communication domain were not as consistent with resulting scores from the concurrent measure.

4.1.5. Known-groups validity

For this analysis, ASQ:I scores from a subsample of 82 children between the ages of 17 to 20 months were used. Thirty-eight of these children had an identified disability and 44 were considered to be typically developing. Results from logistic regression analyses indicated that the ASQ:I domain scores correctly identified over 90% of children in the sample regarding their disability status. It is important to note that the different methods of administration (i.e., the ASQ:I was administered in an interview format over the phone for the children in the disability
group) might have impacted the results. For example, for typically developing children, when caregivers were not sure about their child’s performance, the pediatrician could observe or interact with the child during a visit. This did not happen for the children with known disabilities in the sample because data were collected by phone conversations with the caregivers. In addition, the subsample of children with identified disabilities might be under-representative of children with mild disabilities, since this population is often under-identified in areas with limited diagnostic resources. Future research is required to investigate the sensitivity of the ASQ:I in detecting children with mild disabilities.

4.1.6. Utility

Based on the results from the utility survey and interviews, most caregivers and pediatricians reported positive experiences with the ASQ:I. They rated the ASQ:I items as highly representative of critical skills and the information collected as useful, as well as providing feedback on the advantages and disadvantages of the tool. Results from the utility study will be helpful for future adaptation of the ASQ:I.

4.2. Limitations

As a preliminary study, the normative sample was collected in only one city in Kunshan, China. Generalization of findings to other regions in China should be done with caution. Future research is needed to establish a national sample that includes populations from different regions. In addition, the subsamples for the studies on test-retest reliability, concurrent validity, and known-groups validity were limited to specific age intervals. The sample of children with identified disabilities might also have under-represented children with mild disabilities. Thus, it is questionable to assume that the ASQ:I will have the same level of psychometric rigor when used with children of different ages and/or with mild disabilities.
Each of the 16 pediatricians administered the ASQ:I to 22 to 144 ($M = 51$) children. Since administration methods for the ASQ:I are not strictly standardized, it was possible that different pediatricians administered it in slightly different ways. For example, some pediatricians might prefer to observe the child before interviewing the caregivers on each item, others might collect more information from caregiver report. This dependence of observations might have resulted in a nested effect, which can inflate error rate and effect size (Wampold & Serlin, 2000). However, one-on-one data collection on children would not be feasible given the lack of personnel in China as well as study resources for recruiting and training hundreds of pediatricians. Given that the sample of children was recruited using randomly generated numbers, and that the inter-rater reliability presented a high agreement, the normative sample can still be considered as an adequate representation of the region’s population.

4.3. Implications

Findings from this study contribute to the body of knowledge regarding the adaptation of early childhood developmental measures for use in developing countries. Evidence from this preliminary study demonstrated that the Chinese version of the ASQ:I generates consistent and valid assessment results. The translated ASQ:I items were perceived by caregivers and pediatricians as representing skills that are critical to Chinese children’s development. It took less time than administering the Beijing GDS or other diagnostic assessment tools. Active involvement of caregivers in the assessment process was perceived as a positive feature by Chinese caregivers, but more controversial among pediatricians. Some pediatricians recognized that caregivers learned about child development and got new ideas regarding how to interact with their child by participating in the assessment administration, while others expressed concern about the accuracy and efficiency of relying on caregiver report and observation in the
This study provides initial evidence for validating the ASQ:I as a new alternative in measuring child development for Chinese practitioners, such as pediatricians, early childhood educators, and other professionals (e.g., researchers) interested in evaluating child outcomes for infants and toddlers. Both caregivers and pediatricians in this study reported that they gained valuable knowledge about child development from administering this tool. This educational feature of the ASQ:I might also be utilized by primary child health care providers in the community for training pediatricians as well as parents. When compared with the tools that are currently used for measuring child development in China, the ASQ:I had stronger evidence of technical adequacy and took less time to administer. In addition, the ASQ:I may be more parent/child friendly as administration methods are more flexible than other comprehensive developmental measures regarding materials, instructions, and scoring. The convenience and low-cost of the ASQ:I suggest a promising instrument for use in the diagnostic process when a standardized assessment is not feasible, as is often the case in present day China. The overlap of the ASQ:I with the ASQ-C has the potential to support more efficient assessment practices, as existing scores from the ASQ-C screening tool can be directly transferred to the ASQ:I. As discussed by Bricker and her colleagues (2015), information collected during screening is often not utilized in the eligibility determination and education/intervention processes, resulting in a waste of valuable information about the child’s development. However, existing information from the ASQ-C can be used to complete the ASQ:I if the two assessments are completed within a short time frame (e.g., one to two weeks). Results from the ASQ-C can be used to identify the most efficient start point of the ASQ:I, as well as to provide responses for six of the items in each domain, likely decreasing the amount of time needed to complete the ASQ:I.
At the systems level, the ASQ:I presents a promising option to meet the urgent needs for a cost-effective, efficient, and accurate early childhood assessment instrument in China’s early detection system, as described in the 2013 Guidelines (CCDCP, 2013). Given the large variability in resources across different regions in China, the ASQ:I could be useful at different stages in the nation’s early detection system. In communities where diagnostic services are available but limited, the ASQ:I could be used at the district/county level as a secondary screening measure, to increase the sensitivity and specificity rates in children being referred for diagnostic assessment, thereby limiting diagnostic services to those who are really in need. In communities where diagnostic services are too distant to access, the ASQ:I could be used by local district and county child health care providers to obtain a more accurate and comprehensive understanding of a child’s developmental achievements. A goal for future research activities is to establish a national normative sample and to examine the use of the ASQ:I in determination of potential delays in rural areas of China, where professionals typically have only minimal training, and thus have neither expertise nor time to administer traditional diagnostic assessment tools such as the Beijing GDS. In this way families with children whose screening results indicated concerns would not need to make a long trip to capital cities in order to obtain a diagnosis and would be able to access and benefit from intervention services earlier.

5. Conclusion

This preliminary study examined the Chinese ASQ:I with a regional sample of Chinese children and caregivers, and established initial evidence for its reliability, validity, cultural appropriateness, and utility. For future use of the ASQ:I in China it will be necessary to collect a national sample and examine its psychometric properties nationwide. In addition, research is needed to examine the use of the ASQ:I by professionals other than pediatricians, such as child
care providers and educators. Future research on the broad-based dissemination of the ASQ:I is needed to address possible challenges and needs in the implementation process. Initial findings suggest that the ASQ:I presents a promising tool to assist in the identification process for young children with disabilities in China, especially in areas with limited resources for conducting standardized testing.
References


