**Development of a Mandarin Expressive and Receptive Vocabulary Test for Children Using Cochlear Implants**

**Xiaopan Lu, MPhil[1,2] (E-mail: joyce-xpan@hotmail.com)**

**Corresponding author: Lena L.-N. Wong, PhD[1,2]**

**(E-mail: llnwong@hku.hk; Phone number: +852-28590590)**

**Anita M.-Y. Wong, PhD[1] (E-mail: amywong@hku.hk)**

**[1]Division of Speech and Hearing Sciences, the University of Hong Kong, 5th Floor, Prince Philip Dental Hospital, 34 Hospital Road, Hong Kong, China**

**[2]Clinical Hearing Sciences (CHearS) Laboratory, the University of Hong Kong, 5th Floor, Prince Philip Dental Hospital, 34 Hospital Road, Hong Kong, China**

**Xin Xi[3] (E-mail: xxi@bbn.cn)**

**[3]Department of Otolaryngology Head and Neck Surgery, Chinese PLA General Hospital, 28 Fuxing Road, Haidian District, Beijing, China**

**Running title: Development of a Mandarin vocabulary test**

## Abstract

Cochlear implants (CIs) provide children with profound hearing loss access to sounds and speech. Research on the effects of CI on speech and language development in mainland China is scarce due to the lack of standardized tests. This study aims at developing a vocabulary measure, the Mandarin Expressive and Receptive Vocabulary Test (MERVT), for pre-school children with CIs. Using responses from 102 normal-hearing preschool children, the initial vocabulary set was subjected to analyses to identify items with appropriate levels of difficulty and discrimination. Norms on 245 normal-hearing children aged 1;6 to 3;11 were later collected based on the final set of the items. Evaluation of the test's psychometric properties revealed good internal consistency. Significant correlations between the total MERVT scores and the Gesell Developmental Scale scores, between the MERVT expressive and receptive subtest scores and the total scores, and the gradual increase in MERVT scores with age, provided evidence of construct validity. Results from 29 children with CIs were also examined for evidence of the MERVT's construct validity. There was a significant correlation between these children's MERVT scores and their scores from an intelligence test. The MERVT scores increased with an increase in the duration of CI use and in chronological age. With good reliability and strong validity, the test is recommended for use in the monitoring of language development in children with CI.

Keywords: Cochlear implant, Mandarin Chinese, reliability, validity, vocabulary

## 1. Introduction

The Second China National Sample Survey on Disability in 2006 shows a prevalence of 2.11% or 27.8 million people suffering from hearing impairment (HI) in mainland China (National Bureau of Statistics of China, 2007). About 137,000 of these are children aged 0 to 6 years who exhibit a severe to profound HI (Chen, Cui, & Zhang, 2010). Because of late diagnosis, the mean age of cochlear implantation is at 2 years and above (Huang et al., 2006). With more than 16,000 cochlear implant (CI) surgeries to be performed in the next 3 years (China Disabled Persons Federation, 2011), there is an urgent need to develop measures to evaluate candidacy and outcomes.

### 1. 1 Current vocabulary measures for Mandarin-speaking children

There are three parent-report measures of vocabulary development for Mandarin-speaking children living in mainland China: the two Chinese versions of the MacArthur Communicative Development Inventory (CCDI; Tardif et al., 2008; Hao, Shu, Xing, & Li, 2008) and the vocabulary scale of the Auditory and Language Ability Evaluation Criteria and Methods for Hearing Impaired Children (Sun, 2009). These measures either contain some items that are culturally irrelevant for Mainland Chinese children (Tardif et al., 2008), or do not report test reliability and validity (Hao et al., 2008; Sun, 2009).

Direct vocabulary measures are well-suited for children two years of age or older (Okalidou, Syrika, Beckman, & Edwards, 2011; Thal, Desjardin, & Eisenberg, 2007) and are commonly used to complement parent reports in a battery of hierarchical measures. There is however only one published receptive vocabulary test for mainland Chinese children aged between 3;6 (year; month) and 8 years, the Peabody Picture Vocabulary Test (PPVT-Chinese version; Gong & Guo, 1984). Given the drastic changes in lifestyle in mainland China in the last 30 years as a result of the transition from a closed to a globalized economy, the normative data, items and illustrations are dated (Murphy & Davidshofer, 1991). A few direct vocabulary measures with report of psychometric properties have been developed for Mandarin speaking children in Taiwan: the Peabody Picture Vocabulary Test – revised (PPVT-R: Lu & Liu, 1998) and the Receptive and Expressive Vocabulary Test (REVT; Huang, Jian, Zhu & Lu, 2009). However, subtle cultural differences make some of the test items inappropriate for mainland Chinese children. The present study was thus aimed to develop a vocabulary test for the two language modalities, the Mandarin Expressive and Receptive Vocabulary Test (MERVT), for children with vocabulary development who are equivalent to 2;0 to 3;11 years and use CIs.

### 1.2 Psychometric properties and test development

A standardized vocabulary test should possess adequate psychometric properties and thus its reliability, content and construct validity should be examined. Construct validity refers to the degree to which a test actually measures a construct of interest (Cook & Beckman, 2006). Many procedures can be used to provide evidence of construct validity. For example, convergent evidence of construct validity can be established if the scores of a new test are shown to be correlated to scores of another existing and valid test that purports to measure an identical construct or related constructs (Cohen, 2010). Given the lack of a gold standard measure in Chinese, the convergent evidence of many existing vocabulary tests is often provided by correlating children’s scores to their scores in an intelligence test, because vocabulary is one of the most important contributors to intelligence measurement (Elliott, 1983). For example, the scores on the CCDI (Tardiff et al., 2008) were correlated with scores on the language subtest of the Putonghua version of the Gesell Developmental Scale (Zhang, Li, Qin, & Zhang, 1994). Performance on the Taiwan-Mandarin version of the PPVT-R (Lu & Liu, 1998) was correlated with results from three intelligence tests.

A vocabulary measure with good construct validity should also reveal performance that varies with age or time lapse such as age of implantation or duration of implant use (Cohen, 2010). A moderate and significant relationship between chronological age and language performance in children with CIs is expected (Tomblin, Barker, & Hubbs, 2007). Like other studies, Houston and Miyamoto (2010) found that children implanted before 2 to 3 years of age had better language outcomes. While Niparko et al. (2010) found that implantation before 18 months of age resulted in trajectories of language development paralleled those of hearing peers, Wang, Huang, Wu, and Kirk (2007) also found that compared with children implanted after 3 years of age, early implantees in Taiwan had significantly better receptive and expressive language skills. Language ability is also expected to improve with the duration of CI use. Geers (2004) reported that 43% of children who received CI between 24 and 35 months of age could achieve age-appropriate speech after 5 to 6 years of CI use.

Content validity addresses a judgment of whether the items in a test adequately represent the domain that the test purports to measure (Cohen, 2010). For the purpose of the MERVT, content validity is ensured by creating items that are age-appropriate, familiar and adequately represent the vocabulary used by children (Cohen, 2010) and items that are free from the effects of gender and socioeconomic background (Easterbrooks & O'Rourke, 2001; Hoff & Tian, 2005). Test items generated based on the vocabulary development of normal-hearing children should be applicable to those with HI because their acquisition of words follows similar patterns (Nott, Cowan, Brown, & Wigglesworth, 2009).

## 2. Material and Methods

To construct the MERVT with items that result in high reliability and good content and construct validity, four phases were involved. In Phase 1, items were selected and rated for age-appropriateness. In Phase 2, items with appropriate difficulty and discrimination ability were identified. In Phase 3, norms were established and psychometric properties of the test were evaluated. The application of the MERVT on children with CIs was evaluated in the final Phase 4.

### 2.1 Phase 1 to select the initial item pool

#### 2.1.1 Material

The initial item pool was selected from four sources. First, items were translated from English vocabulary tests. These include the PPVT-III (Dunn & Dunn, 1997), the Expressive Vocabulary Test (EVT: Williams, 1997), the Expressive One-Word Picture Vocabulary Test (EOWPVT: Brownell, 2000a), and the Receptive One-Word Picture Vocabulary Test (ROWPVT: Brownell, 2000b). Second, items were selected from existing Mandarin vocabulary tests, including the two versions of the CCDI (Tardif et al., 2008; Hao et al., 2008), the PPVT-R (Lu & Liu, 1998) and the REVT (Huang et al., 2009). Third, four Chinese corpuses from the Child Language Data Exchange System (2010) database were examined: “Beijing” from conversations of children aged 1;9 to 2;2 years old, “Contex” from conversations of children aged 2 years old, “Zhou1” from conversations between children aged 1;2 to 2;8 years and their mothers, and “Zhou2” from conversations between children aged 3 to 6 years and their mothers (Child Language Data Exchange System, 2010). Finally, Chinese story books and TV programs for the target ages were reviewed.

### 2.1.2 Participants

Item difficulty was rated separately for the expressive and receptive subtest of the MERVT by 13 kindergarten teachers. These teachers, who had between 5 to 20 years of teaching experience, were from the city of Beijing and a city in three other provinces (Zhejiang, Gansu and Sichuan), representing different socioeconomic levels and dialectal and cultural influences (Du, 2007). Although these teachers may speak a local dialect, they all reached the secondary level in the National Proficiency Test of Putonghua.

### 2.1.3 Procedures

The initial item pool contained 573 items. A majority (85.2%) was nouns, and the rest included 6.3% verbs, 6.8% adjectives and 1.7% quantifiers, covering 15 different familiar categories (e.g., animals, food and drinks). To be appropriate for children between the ages one to five, these items can all be unambiguously illustrated. The 13 kindergarten teachers rated the difficulty of each item using a five-point Likert scale, with “1” representing “very easy”, and “5” representing “very difficult”. They also suggested additional words which they felt appropriate. The items were rated in six-month increments for the age groups from 2;0 years to 3;11 years. Items with mean ratings from 2 to 4 in each age group were considered as having moderate difficulty, and thus, were appropriate for the particular age group. Items with mean values below 2 in the 2;0 to 2;5 group were considered as being too easy and were therefore kept as practice items or items for evaluating children aged 1;6 to 1;11 years. Thirty items with similar ratings from the teachers (smaller standard deviation in each age group) were kept because they were probably more resistant to differences in socioeconomic, dialectal and cultural differences. If there were overlapped items across age groups or in the two subtests, they were placed in the group or subtest where smaller standard deviations were produced. Colored drawings were made to illustrate the items. There were no written words to provide clues. Adjectives were illustrated using contrastive pictures (e.g., “small” is illustrated by contrasting a big apple with a small one)

### 2.1.4 Results

Results of this rating exercise were a list of 102 words for the receptive and another list of 98 words for the expressive subtest for the MERVT. Among the 102 items in the receptive subtest, 75.5% were nouns, 12.7% were verbs, and 11.8% were adjectives. Among the 98 items in the expressive subtest, 78.6% were nouns, 10.2% were verbs and 11.2% were adjectives. Three items with a mean value of 1 for the 2;0 to 2;5 group in the initial pool were selected as practice items for the receptive subtest, and eight practice items were selected for the expressive subtest.

## 2.2 Phase 2 to conduct item analysis

Item analysis was used to identify items with appropriate difficulty and discrimination ability to formulate the final test.

### 2.2.1 Material

The 102 target items and their distractors, three for each item, in the receptive subtest were drawn and arranged in a two-by-two format on a page. The distracters belong to the same super-ordinate category and received a similar teacher rating as the target. For example, the target vocabulary “horse” (马 /mA̜/) was illustrated with the distractors “pig” (猪 /tʂu/), “cat” (猫 /mɑu/) and “rabbit” (兔 /t‘u/). Illustrations for the 98 expressive items were presented one on each page. To ensure that the illustrations successfully elicit the target receptive and expressive responses, five children aged 3;0-3;11 years were asked to complete the two subtests. As a result of this pilot study, 18 pictures in the expressive subtest and 21 pictures in the receptive subtest were revised or repainted.

**2.2.2 Participants**

An equal number of children, aged between 1;06 and 3;11 years, (45% were female) were recruited from three different socioeconomic areas in Beijing: the Xi Cheng, Hai Dian and Da Xing district (Beijing Municipal Bureau of Statistics, 2011). The 102 children, all satisfied the following selection criteria: (1) had no history of premature birth, otitis media, or prior diagnosis of mental or intellectual disabilities, (2) used Putonghua as their first language, (3) passed the Transient Evoked Otoacoustic Emission (TEOAE) screening using the MADSEN AccuScreen to ensure bilateral normal hearing, (4) passed the developmental screening using the Chinese version of theDenver Developmental Screening Test (DDST) (Lin, Su, & Li, 1983), and (5) had intelligible speech.

Among the 102 children, 75.3% had mothers who were college graduates or above, and 14.7% of the mothers had high school education and 10.0% received lower education. The household income was more than RMB 10,000 for 19.5% of the families, RMB 8,000 to 10,000 for 26.6%, RMB 5,000 to 8,000 for 26.4%, RMB 2,000 to 5,000 for 23.4% and less than RMB 2,000 for about 4.1%.

**2.2.3 Procedures**

Each of the children completed three trial items for the receptive and five trial items for the expressive subtests. The order of administration of the two subtests was random. For the receptive subtest, pointing to the picture that matches the word given was awarded one point. The total number of points was 102. The children’s responses in the expressive subtest were scored following these criteria. A score of “1” was assigned to the correct or an alternative response that is widely acceptable. For example, the target item “tire” (轮胎 /luəntʰai/) could also be named as “轱辘 /kulu/”, “轮子 /luəntsɿ/”, “车轮 /tʂʰɤluən/” in mainland China. For the 1;6 to 1;11 group, duplicated words or one syllable words were considered as acceptable response for some items. For example, the target word “banana” (香蕉 /ɕiɑŋtɕiɑu/) may be named as “蕉 /tɕiɑu/” or “蕉蕉 /tɕiɑutɕiɑu/”. An incorrect response would receive a score of “0”. The total number of points was 98.

**2.2.4 Results**

After item analysis, 88 words in the receptive subtest and 73 words in the expressive subtest with item difficulty scores ranging from 0.15 to 0.85 and item discrimination scores greater than 0.25 were kept (Oller, 1979). In the receptive subtest, the proportion of nouns, verbs and adjectives was 73.9%, 15.9% and 10.2%, respectively. In the expressive subtest, the proportion of nouns, verbs and adjectives was 80.8%, 8.2% and 11.0% respectively.

After the selection of items with proper difficulty and discrimination ability, items that showed an improvement trend of difficulty with age were retained. In a particular age group, items with a passing rate of more than 70% were placed into the adjacent younger age group. Within each age group, all words were arranged in an ascending order of difficulty. This strategy would allow the test to proceed in a forward manner with a minimal need to go backwards to establish a basal. Because of the ascending order of difficulty in each block, the starting point for the test was set as the first item in the block that corresponded to the chronological age of children. If the child made a correct response at the starting point, the test continued until the first error was made. If 5 consecutive correct responses were made, a basal has been established. Then the examiner continued until the child made five consecutive errors. If correct responses could not be recorded for the first five items, the test should go backward to assess easier items until five consecutive correct responses were made, and thus establish the basal. The basal is then defined as the highest item number of 5 consecutive correct responses. The ceiling is defined as the lowest item number of 5 consecutive incorrect responses. The un-attempted items below the basal were coded as "correct" and those above the ceiling were coded as "incorrect". The final score for each child was then the sum of the item number that represents the basal and the number of correct responses between the basal and the ceiling.

### 2.3 Phase 3 to establish normative data and evaluate psychometric properties

To establish construct validity, the correlation between the developmental quotient (DQ) of the Language and Adaptive Behavior scales on the Gesell Developmental Schedules (GDS) (Zhang et al., 1994) and the receptive and expressive scores on the MERVT was examined on a subgroup of 48 children aged 2;6 to 3;11 years randomly selected from the entire sample and the correlation between the receptive and expressive subtest scores and the total test score was examined. The effects of age, gender, and SES on performance on each of the subtest were also explored.

#### 2.3.1 Subjects and procedures

Normative data were obtained on 245 children who were recruited from kindergartens and child healthcare units in three different socioeconomic areas in Beijing, as in Phase 2. An equal number of boys and girls in each age group were recruited to evaluate gender effects (Leaper, 2002). While 77.6% of mothers had a college education or above, 14.7% were secondary school graduates and 4.9% received lower education. Missing data was noted in 2.8% of the participants. For monthly household income, 24.5% of families had more than RMB 10,000, 22.0% earned RMB 8,000 to 10,000, 29.4% had RMB 5,000 to 8,000, 15.1% had RMB 2,000 to 5,000 and 1.2% made less than RMB 2,000. Missing data was noted in 7.8% of families.

**2.3.2 Results**

Data from five subjects were excluded because their expressive or receptive scores were more than 2 standard deviations away from the means. The mean subtest scores in each age group show that the expressive and receptive vocabulary abilities increased with age (see Table 1).

Insert Table 1 about here. The Cronbach's alpha coefficients for the two subtests ranged from .83 to .97 across the five age groups, suggesting strong internal consistency. The standard errors of measurement (SEMs) for the expressive subtest vary from 1.45 to 2.85 (median: 2.63), and the SEMs for the receptive subtest vary from 1.96 to 3.50 with a median of 2.50. The relative standard error (RSE) for each age group ranged from 2.42% to 23.65%. The confidence interval for each age group is shown in Table 2 and given an individual's MERVT score, the confidence intervals could be used to determine the likelihood that a given score is within the vocabulary level of the corresponding age norm.

Insert Table 2 about here.

With the effects of age controlled in a partial correlation analysis, the developmental quotient (DQ) of the Behavioral scale of the GDS correlated significantly with both the expressive (r =.53, $p < .01$) and receptive (r =.43, $p < .01$) subtests and the DQ of the Language scale of the GDS was significantly correlated with the performance on the two subtests as well (expressive: r =.60, $p < .01$ and receptive: r =.37, $p < .05$). Pearson Product Moment correlation showed that the total score correlated significantly with the performance on the expressive (r =.95, $p < .01$) and receptive (r =.97, $p < .01$) subtests, suggesting that items in both subtests measured the same construct (Cohen, 2010).

A one-way ANOVA showed a significant effect of age on the expressive, $F(4, 235) = 168.5$, $p < .001$, and receptive vocabulary scores, $F(4, 235) = 205.3$, $p < .001$. Although the growth of receptive vocabulary in the oldest two groups seemed to have plateaued, post hoc tests with Games-Howell adjustment indicated that performance of the adjacent age groups were all significantly different from each other on both the receptive and the expressive subtest ($p < .01$).

The relationship between other demographic variables (i.e., gender, maternal education level and household income) and performance on the MERVT was examined. An independent samples t-test showed no gender effect on the expressive, $t(238) = -.88$, $p > .05$, and receptive subtest scores $t(238) = -.62$, $p > .05$. Both household income and maternal education level were important components of SES, and they were not related to the MERVT scores ($r_s$ =.08-.15, $p > .05$).

### 2.4 Phase 4 to examine the application of the MERVT in children with CIs

The construct validity of the MERVT in measuring vocabulary in children with CIs was evaluated. A few relationships were hypothesized and tested. First, the convergent evidence of construct validity was provided by correlating the MERVT with the CCDI (Tardif et al., 2008), which is a parent report measure that requires parents to pick out the words their children understood or produced. Parents may overestimate their children's vocabulary repertoire, while some types of vocabulary may not be surveyed in a direct measure. Thus the MERVT was expected to have a moderate and significant relationship to the scores on the CCDI. Second, the MERVT was expected to complement the CCDI, in a hierarchical battery of tests to measure the vocabulary of children with CIs. Finally, the MERVT was expected to demonstrate better vocabulary outcomes in those implanted earlier, used CI for a longer period of time, and were older chronologically (Fagan & Pisoni, 2010; Svirsky, Robbins, Kirk, Pisoni, & Miyamoto, 2000; Thal et al., 2007).

#### 2.4.1 Participants

Children with CIs were recruited in the People's Liberation Army General Hospital (PLAGH) in Beijing. They met the following criteria: (1) exhibited pre-lingual HI, (2) had normal inner ear structure and no nerve absence or deficiency, as verified by a radiological examination, (3) exhibited normal cognitive abilities as indicated in the Chinese version of the Griffiths Mental Development Scales (Wang, Qu, Zhao, & Wei, 2007) for children younger than 3 years or the Hiskey-Nebraska Test of Learning Aptitude (H-NTLA: Yang, Qu, Sun, & Wang, 2011) for children older than 3 years, (4) were fitted with unilateral CIs; (5) had been implanted by 5 years of age, (6) exhibited no other diagnosed sensori-, oromotor, or co-morbid developmental deficits, such as autism; (7) used an oral communication mode, and (8) spoke Mandarin as the primary language in their families.

The 29 participants ranged from 21 to 78 months of age (*M* = 46.3, *SD* = 12. 9) and 11 were girls. The mean age of implantation was 27.8 months (*SD* = 11.1, range = 9 to 55 months). 27 of them (93.1%) were implanted between 1 to 5 years of age, and 6.9% received the CIs before 12 months of age. Hearing aids were tried by 48.2% of the children, for a mean length of 10.1 months (*SD* = 6.3). The mean length of implant use was 18.5 months (*SD* =10.2). About 10.3% of them had used CI for less than 12 months, 69.0% of them had CI for 12 to 24 months, and 20.7% had more than 2 years of CI use. About 27.6% of mothers went to college or above, 27.6% had secondary education and 44.8% had primary or lower education. About 3.5% of households had a monthly income of RMB8,000 to 10,000, 24.0% earned RMB5,000 to 8,000, 48.3% earned RMB2,000 to 5,000 and 20.7% earned RMB2,000. Missing data werenoted in 3.5% of the children. After implantation, 82.7% of children were still attending special schools, 13.8% were mainstreamed and 3.5% stayed home.

### 2.4.2 Procedures

Testing was conducted in a sound treated room. Parents provided demographic information. The MERVT test procedures were modified to suit children with HI. First, the starting points were determined based on the duration of CI use, because vocabulary ability is commensurate with the duration of CI use (Fagan & Pisoni, 2010). Second, because early language development is marked by the use of some nonconventional and idiosyncratic vocalizations (Fletcher & MacWhinney, 1995) and children with CIs may have misarticulation, 37.9% of the parents were asked to comment on the responses when they were not apparent. To avoid bias, the parents were seated where they could not see the picture cards. In addition to the MERVT, all the parents also completed the Words and Gestures form (WG) (8 to 16 month olds) and Words and Sentences form (WS) (16 to 30 month olds) of the Putonghua version of the CCDI (Tardif et al., 2008). Using the WG form, parents reported whether their children could produce and/or understand each listed word. On the WS form, parents identified the words their child could produce.

### 2.4.3 Results

#### 2.4.3.1 Evidence for construct validity

Although the chronological age of 96.5% of the children was outside the age range for the CCDI, they could still be measured using the CCDI, because the language development of children with CIs was delayed compared with those with normal hearing (see Table 3 for descriptive data). Performance on the receptive MERVT correlated significantly with the Words Understood section ($r_s = .63$, $p < .05$) and on the WG form of the CCDI, and the expressive subtest score on the MERVT correlated with Words Produced section of the WG form of the CCDI ($r_s = .70$, $p < .01$). Both subtests correlated with the Words Produced section of the WS form ($r_s > .64$, $p < .01$). Thus, results provided evidence for construct validity.

Insert Table 3 about here.

Table 4 shows that 17 children were categorized in the same age-equivalent categories on both the CCDI and the expressive subtest of the MERVT; 24 were placed within the adjacent age-equivalent category and participants were more likely to be classified by the MERVT to a higher age-equivalent category than the CCDI. A total of 12 children were categorized in the same age-equivalent categories on the CCDI and the receptive subtest of the MERVT; and 25 were placed in the adjacent age-equivalent categories. It is more likely for the receptive subtest of the MERVT to classify participants into a lower age-equivalent category than the CCDI. The above results show that the MERVT reveal vocabulary development that is close to the findings obtained using the CCDI.

Insert Table 4 about here.

Prior to evaluating how performance on the MERVT related to chorological age, age at implantation and duration of CI use, data from three children were removed as outliers, leaving the number of children to 26. A significant relationship between receptive and expressive vocabulary was found (see Table 5). MERVT subtest scores were significantly related to chronological age (expressive subtest: $r_s = .48$, $p < .05$; receptive subtest: $r_s = .67$, $p < .01$) and the duration of CI use (expressive subtest: $r_s = .55$, $p < .01$; receptive subtest: $r_s = .55$, $p < .01$). Age at implantation did not relate to MERVT scores.

Insert Table 5 about here.

**2.4.3.2 The use of MERVT in a hierarchical battery for vocabulary measurement**

To examine whether the MERVT was better at evaluating more advanced vocabulary development, performance of children at the ceiling of the CCDI was examined. For the Words Understood section in the WG form of the CCDI, 17 children (mean age at implantation = 27.5 months, mean duration of CI use = 21.5 months) scored at the ceiling; but they attained a mean score of 41.8, out of a total of 88 (SD = 21.2) on the receptive subtest of the MERVT. Eight children (mean age at implantation = 29.0 months, mean duration of CI use = 24.2 months) received full marks in the Words Produced section of the WG form and a mean score of 33.2, out of a total of 73 (*SD* = 12.1) on the expressive subtest of the MERVT. On the WS form, two children (mean age at implantation = 20.5 months, mean duration of CI use = 42.5 months) scored at the ceiling, but achieved mean scores of 39.0 (*SD* = 1.4) and 51.0 (*SD* = 12.7) on the expressive and receptive subtests of the MERVT, respectively.

To the contrary, three children (mean age at implantation = 27.3 months, mean duration of CI use = 11.0 months) scored “0” on the receptive subtest and two of these three children (mean age at implantation =24.0 months, mean duration of CI use = 12.5 months) had a score of “0” on the expressive subtest of the MERVT. These children obtained mean scores of 27.3, 63.3 and 11.6 on the Words Produced and Words Understood sections of the WG form, and the WS form of the CCDI, respectively. These results indicated that the CCDI, as a parent-report measure, was more appropriate for children at the early stage of language development; the MERVT was better at measuring more advanced vocabulary skills.

**3. General discussions**

In phase 2, items with a good range of difficulty and sufficient discriminatory power were identified. While most items were nouns, there were a small percentage of verbs (8.2% and 15.9% of in the expressive and receptive subtests, respectively), partly because action verbs were difficult to illustrate using static pictures. These proportions are consistent with those in the REVT (Huang et al., 2009) and Liu (2008). Compared with the CCDI (Tardif et al., 2008), the MERVT has a higher percentage of nouns, probably because parent-report measures are better able to profile people’s names (e.g., mom), interjections, pronouns and onomatopoeia (e.g., dog’s bark) that cannot be easily illustrated with pictures.

The Cronbach’s alpha coefficients for both subtests suggest good internal consistency (Streiner & Norman, 2003). Thus, items in both subtests measured a single construct. In the current study, the RSE for each age group is less than 25%, indicating that the MERVT is a reliable measure of vocabulary skills (Australian Bureau of Statistics, 2012). As evidence for convergent validity, performance on the MERVT was found to correlate with results on the GDS (Cook & Beckman, 2006). Second, scores in each subtest related significantly to the total scores of the MERVT, again suggesting the two subtests were measuring the same construct. Third, performance improved with chronological age, as theoretically predicted (Li, 1995; Owens, 2008). Given that the MERVT items also have good discrimination ability and are of appropriate difficulty, and have been shown to discriminate the vocabulary abilities of children of different ages, the MERVT is considered to be a valid measure of vocabulary.

While the MERVT scores are expected to improve with age, the items should be free from the effects of gender and SES so that the MERVT could be used in children with different demographic characteristics. In the present study, SES was evaluated as a function of household income and maternal education because women with a higher education level tended to get a job with a higher salary, resulting in an increase in household income (Lamer, 2011). While many studies have revealed that maternal education level and household income are significantly correlated with language development (Basilio, Puccini, Silva, & Pedromonico, 2005; Hoff & Tian, 2005), the items in the MERVT have been selected so that they are free from these influences.

### 3.1 Evaluation of construct validity

The MERVT scores exhibited a significant relationship with the CCDI, thus provided evidence of convergent validity for the MERVT in measuring vocabulary in children using CI (Cohen, 2010). Age equivalents on the MERVT were compared with those on the CCDI. About half of the children were categorized into the same age-equivalent categories on the CCDI and the MERVT; and most of them were placed within the adjacent age-equivalent category. The results show that the MERVT can reveal vocabulary development of children with CIs that is congruent with that obtained using the CCDI, despite possible measurement errors.

Construct validity was also evaluated by relating the results on the MERVT to three factors shown to be significant determinants of vocabulary development in English-speaking children. In the present study, chronological age correlated significantly with expressive and receptive vocabulary growth, as predicted. Previous research also showed similar relationships (Tomblin et al., 2007). It is commonly accepted that vocabulary expanded with increased CI use (Fagan & Pisoni, 2010; Thal et al., 2007). Fagan and Pisoni (2010) also indicated that compared to chronological age, the duration of CI use was more commensurate with children's receptive vocabulary size, because longer CI use would result in more language experience and auditory language exposure for word learning than will their chronological age alone (Geers, Nicholas, & Sedey, 2003; Thal et al., 2007). Similar findings were noted in the current study. These findings provided evidence for construct validity of the MERVT.

Many studies reported that early CI would result in better language outcomes (Connor, Craig, Raudenbush, Heavner, & Zwolan, 2006; Fagan & Pisoni, 2010; Nicholas & Geers, 2007), however, the current study did not find such relationships. Dawson, Blamey, Dettman, Barker, and Clark (1995) also found that age at implantation did not significantly correlate with the vocabulary acquisition rate in children who received CI between 2;6 to 20 years of age, and used CI for 6 months to 7;8 years. Results in the present study probably have been confounded by the lack of infrastructure (e.g., good hearing aid fitting, aural rehabilitation) to support a successful CI program (Cao & Wang, 2006). In fact, 55.1% of the participants did not have a pre-implant hearing aid trial and 82.7% of them were still attending special schools. There is a lack of speech-language pathologists and there are less than 100 professionally trained audiologists. Thus, the lack of relationships between age of implantation and vocabulary development should not be viewed negatively in terms of providing evidence of validity for the MERVT.

### 3.2 The MERVT in a hierarchical battery for vocabulary measurement

A hierarchical battery of measures is required to evaluate language development in children with CIs. Thus, how well the MERVT and CCDI complemented each other was evaluated. In this study, three children could not be tested using the receptive subtest of the MERVT and two could not be tested using the expressive subtest. In contrast, the CCDI successfully recorded the vocabulary development of these children. The early stage of language development is characterized by the use of onomatopoeias, words that express social relationships (e.g., yes, no, want) and names of their family members. These items could not be illustrated in picture and thus were not included in the MERVT.

A significant proportion of the participants were performing at the ceiling on the CCDI. Similar issues were reported by Thal et al. (2007) who used the long form of the CDI on 24 children implanted at a mean age of 28.6 months and had used CI for 3 to 60 months. The researchers also found that four, and two, out of the 24 children in their study scored at the ceiling on the Words Understood, and Words Produced sections of the WG form of the CDI, respectively. Thus, the WG form was considered inappropriate when the vocabulary development exceeded the age equivalent of 16 months in the WG form of the CCDI. Only two children scored at the ceiling of the WS form in the current study, thus, the WS form could be used for measuring children with CIs whose receptive language ability had developed beyond the level that could be measured on the WG form.

Therefore, for children whose language ability have not achieved the age equivalent of 16 months on the WG form of the CCDI or have not attained the age equivalent of 18 to 23 months on the MERVT, the CCDI is preferred for measuring their vocabulary ability. In particular, the WG form is useful for children who have very low levels of language skills (Thal et al., 2007). The MERVT should be used when results on the CCDI indicate more than 97 words acquired in the Words Understood section and/or more than 24 words in the Words Produced section of the WG form (or an age equivalent of 16 months is exceeded on the WG form of CCDI). Age equivalents, instead of other objective rules (e.g., age or duration of implantation) should be used to determine which measure should be adopted in a hierarchical battery of language outcomes. In addition, if some children with CIs could finish all the practice items on the MERVT successfully, the MERVT could be tried earlier on these children; because words selected in both the practice plates and plates for age group 1;6 to 1;11 years in Phase 1 have the same teachers' rating of difficulty.

## 4. Conclusions

The MERVT could be used for evaluating the expressive and receptive vocabulary of children with CIs and with those who exhibited vocabulary development equivalent to normal hearing listeners aged 1;6 to 3;11 years. The items were selected based on the difficulty rating of kindergarten teachers from geographical regions that represent different levels of SES and culture in mainland China. There was also a lack of gender and SES effects. Thus, all the selected items were expected to be appropriate for the general preschool population in mainland China. However, these findings do not preclude the need for validation of normative data in other parts of the country where the average household income and education level are much lower than Beijing.

While item analysis showed that the words in the final version of the MERVT had a good range of difficulty and sufficient discriminatory ability, the MERVT also has good internal consistency and strong content and construct validity. Significant relationships between the scores of MERVT and those for the CCDI, chronological age and the duration of CI use provided evidence of construct validity for measuring vocabulary in children with CIs (Cohen, 2010). In addition, The MERVT can be combined with the CCDI and used in a hierarchical set of battery tests for assessing language outcomes of children with CIs.

## 5. References

Australian Bureau of Statistics (2012). Survey Methodology. Retrieved December 21, 2012 from http://abs.gov.au/ausstats/abs@.nsf/Previousproducts/4430.0.30.002Main%20Features202003%20(Reissue)?opendocument&tabname=Summary&prodno=4430.0.30.002&issue=2003%20(Reissue)&num=&view=

Basilio, C. S., Puccini, R. F., Silva, E. M., & Pedromonico, M. R. (2005). Living conditions and receptive vocabulary of children aged two to five years. *Revista Saúde Pública*. *39*(5), 725-730.

Beijing Municipal Bureau of Statistics (2011). Beijing Statistical Yearbook in 2011. Retrieved March 18, 2012 from http://www.bjstats.gov.cn/nj/main/2011-tjnj/index.htm.

Brownell, R. (2000a). *Expressive one-word picture vocabulary test* (3rd ed.). Novato, CA: Academic Therapy Publications.

Brownell, R. (2000b). *Receptive one-word picture vocabulary test* (3rd ed.). Novato, CA: Academic Therapy Publications.

Cao, K. L., & Wang, L. E. (2006). Current status and correlated issues on cochlear implantation. *Chinese Medical Journal, 119*(12), 971-973.

Chen, G., Cui, B., & Zhang, L. (2010). Study on the prevention and strategy of disability in China. *Procedia - Social and Behavioral Sciences, 2*(5), 6906-6913.

Child Language Data Exchange System (2010). Browsable Database. Retrieved July 15, 2012 from http://childes.psy.cmu.edu/browser/index.php?url=EastAsian/Chinese.

China Disabled Persons Federation (2011). Poor deaf children (cochlear implants) rehabilitation programme Retrieved January 18, 2012 from http://www.cdpf.org.cn

/kangf/kf.htm

Cohen, R. J. (2010). *Psychological testing and assessment : An introduction to tests and measurement* (7th ed.). Boston, MA: McGraw-Hill Higher Education.

Connor, C. M., Craig, H. K., Raudenbush, S. W., Heavner, K., & Zwolan, T. A. (2006). The age at which young deaf children receive cochlear implants and their vocabulary and speech-production growth: Is there an added value for early implantation?. *Ear and Hearing, 27*(6), 628-644.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *Amerian Journal of Medicine 119*(2), 167-116.

Corona, R., Cowgill, B. O., Bogart, L. M., Parra, M. T., Ryan, G., Elliott, M. N., . . . Schuster, M. A. (2009). Brief report: A qualitative analysis of discussions about HIV in families of parents with HIV. *Journal Pediatric Psychology, 34*(6), 677-680.

Dawson, P. W., Blamey, P. J., Dettman, S. J., Barker, E. J., & Clark, G. M. (1995). A clinical report on receptive vocabulary skills in cochlear implant users. *Ear and Hearing, 16*(3), 287-294.

Du, J. (2007). 我国各省市社会经济综合发展水平评价分析[The evaluating of the integrative development levels of social economy in different provinces in China]. *Shandong Textile Economy, 5*, 28-29.

Dunn, L. M., & Dunn, D. M. (1997). *PPVT-III Peabody picture vocabulary test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Easterbrooks, S. R., & O'Rourke, C. M. (2001). Gender differences in response to auditory-verbal intervention in children who are deaf or hard of hearing. *American Annals of the Deaf, 146*(4), 309-319.

Elliott, C. (1983). Development and features of the British ability scales. *British Journal of Educational Psychology, 53*(1), 129-129.

Fagan, M. K., & Pisoni, D. B. (2010). Hearing experience and receptive vocabulary development in deaf children with cochlear implants. *Journal of Deaf Studies & Deaf Education, 15*(2), 149-161.

Fletcher, P., & MacWhinney, B. (1995). *The handbook of child language*. Oxford, United Kingdom: Blackwell.

Geers, A. E. (2004). Speech, language, and reading skills after early cochlear implantation. *Archives of Otolaryngol— Head and Neck Surgery, 130*(5), 634-638.

Geers, A. E., Nicholas, J. G., & Sedey, A. L. (2003). Language skills of children with early cochlear implantation. *Ear and Hearing, 24*(1 Suppl), 46S-58S.

Gong, Z. X., & Guo, D. (1984). 学前和初小儿童智能筛查——图片词汇测试法 [An intelligence screening test for preschool and primary school children- Picture vocabulary test]. *Acta Psychologica Sinica 4*: 392-401.

Hao, M., Shu, H., Xing, A., & Li, P. (2008). Early vocabulary inventory for Mandarin Chinese. *Behavior Research Methods, 40*(3), 728-733.

Hoff, E., & Tian, C. (2005). Socioeconomic status and cultural influences on language. *Journal of Communication Disorders, 38*(4), 271-278.

Houston, D. M., & Miyamoto, R. T. (2010). Effects of early auditory experience on word learning and speech perception in deaf children with cochlear implants: Implications for sensitive periods of language development. *Otology & Neurotology, 31*(8), 1248-1253.

Huang, L. H., Han, D. M., Zhang, L., Guo, L. S., Liu, S., Tang, X. Q., . . . Qi, B. E. (2006). [Analysis of the found age and way for children age 0 to 6 with hearing loss]. *Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi [Chinese Journal of Otorhinolaryngology Head & Neck Surgery], 41*(5), 331-334.

Huang, R. Z., Jian, X. Y., Zhu, L. X., Lu, L. (2009). *華語兒童理解與表達詞彙測驗* [*Receptive and expressive vocabulary test*]. Taibei, Taiwan: Psychological Press.

Lamer, A. (2011). *Vocabulary development in young children: The influence of maternal factors.* Senior Honors Thesis Proposal, Ohio State University. Retrieved June 25, 2012 from http://kb.osu.edu/dspace/handle/1811/51951

Leaper, C. (2002). Parenting girls and boys. In M. H. Bornstein (Ed.), *Handbook of Parenting, vol. 1: Children and parenting* (pp. 189–225). Mahwah, NJ: Lawrence Erlbaum Associates.

Li, Y. (1995). *儿童语言发展 [Language Development of Children]*. Wuchang, China: Central China Normal University Press.

Lin, C. J., Su, Z. Y., & Li, J. P. (1983). 丹佛智能发育筛查法在我国六市正常儿童中再标准化 [Collection of data about restandardization of DDST from six cities in north China] . *Chinese Journal of Pediatrics, 21*(4), 206-209.

Liu, S. (2008). *Early vocabulary development in English, Mandarin, and Cantonese: A cross-linguistic study based on CHILDES* (Master’s thesis). Retreived May 20, 2011 from https://dspace.lasrworks.org/bitstream/handle/10349/711/07PSY-LiuShuxia.pdf?

sequ..

Lu, L., & Liu H. X., (1998). *修訂畢保德圖畫詞彙測驗* [*Peabody Picture Vocabulary Test-Revised*]. Taibei, Taiwan: Psychological Press.

Murphy, K. R., & Davidshofer, C. O. (1991). *Psychological testing: Principles and applications* (2nd ed.). Englewood Cliffs, N J: Prentice Hall.

National Bureau of Statistics of China (2007, May). The second China national sample survey on disability major data communique (no. 2). Retrieved January 18, 2012 from http://www.stats.gov.cn/tjsj/ndsj/shehui/2006/html/ fu3.htm

Nicholas, J. G., & Geers, A. E. (2007). Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to profound hearing loss. *Journal of Speech Language & Hearing Research, 50*(4), 1048-1062.

Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N. Y., Quittner, A. L., . . . Team, C. D. I. (2010). Spoken language development in children following cochlear implantation. *Journal of the American Medical Association, 303*(15), 1498-1506.

Nott, P., Cowan, R., Brown, P. M., & Wigglesworth, G. (2009). Early language development in children with profound hearing loss fitted with a device at a young age: Part II-content of the first lexicon. *Ear and Hearing, 30*(5), 541-551.

Okalidou, A., Syrika, A., Beckman, M. E., & Edwards, J. R. (2011). Adapting a receptive vocabulary test for preschool-aged Greek-speaking children. *International Journal of Language & Communication Disorders, 46*(1), 95-107.

Oller, J. W. (1979). *Language tests at school : A pragmatic approach*. London: Longman.

Owens, R. E. (2008). *Language development : An introduction* (7th ed.). Boston, MA:

Pearson Allyn and Bacon.

Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales : A practical guide to their development and use* (3rd ed.). Oxford, United Kingdom: Oxford University Press.

Sun, X. (2009). *听力障碍儿童听觉、语言能力评估标准及方法 [Auditory and language ability evaluation criteria and methods for hearing impaired children]*. Beijing, China: San Chen Ying Ku Press.

Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., & Miyamoto, R. T. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science, 11*(2), 153-158.

Tardif, T., Fletcher, P., Zhang, Z. X., Liang, W. L., Zuo, Q. H., & Chen, P. (2008). *User's guide and manual for the Chinese communicative development inventories (Putonghua and Cantonese)*. Beijing, China: Peking University Medical Press.

Thal, D., Desjardin, J. L., & Eisenberg, L. S. (2007). Validity of the MacArthur-Bates Communicative Development Inventories for measuring language abilities in children with cochlear implants. *American Journal of Speech-Language Pathology, 16*(1), 54-64.

Tomblin, J. B., Barker, B. A., & Hubbs, S. (2007). Developmental constraints on language development in children with cochlear implants. *International Journal of Audiology 46*(9), 512-523.

Wang, H. Q., Qu, C. Y., Zhao, S. P., & Wei, F. (2007). Griffith 精神发育量表在山西0-7岁儿童中的修订 [Standardization of the Griffith Mental Development Scales for children aged 0-7 years in the cities of Shanxi province]. *Chinese Mental Health Journal*, *21*(10), 700-703.

Wang, N. M., Huang, T. S., Wu, C.M., & Kirk, K. I. (2007). Pediatric cochlear implantation in Taiwan: Long-term communication outcomes. *International Journal of Pediatric Otorhinolaryngology, 71*(11), 1775-1782.

Williams, K. T. (1997). *AGS expressive vocabulary test*. Circle Pines, MA: American Guidance Service.

Yang, X., Qu, C., Sun, X., Zhu, F., & Wang, J. (2011). 希-内学习能力测验中国 3-7 岁儿童常模修订 [Norm revision of H-NTLA for children from 3 to 7 years old in China]. *Chinese Journal of Clinical Psychology, 19*(2), 195-197.

Zhang, X. L., Li J. P., Qin, M. J., & Zhang, C. Y. (1994). Gesell 发展诊断量表3.5-6岁北京修订本的制定 [The revise of Gesell Developmental Scale on 3.5-6 years of age in Beijing]. *Chinese Journal of Clinical Psychology, 2*(3), 148-150.

**Table 1.** *Mean score (M) and standard deviation (SD) for each age group (N = 245)*

| **Age group** | **N** | **Expressive test** | | **Receptive test** | |
|---|---|---|---|---|---|
| **(years)** | | ***M(SD)*** | **Range** | ***M(SD)*** | **Range** |
| 1;6-1;11 | 47 | 6.13 (5.47) | 0-18 | 18.55 (12.90) | 1-50 |
| 2;0-2;5 | 51 | 19.98 (9.74) | 1-53 | 40.04 (17.70) | 11-72 |
| 2;6-2;11 | 45 | 31.71 (12.73) | 9-64 | 58.11 (14.29) | 25-80 |
| 3;0-3;5 | 48 | 40.81 (10.49) | 21-64 | 75.13 (7.92) | 52-85 |
| 3;6-3;11 | 50 | 52.96 (9.95) | 26-68 | 81.16 (4.76) | 69-87 |

**Table 2.**

*Confidence interval (CI) of MERVT scores for each age group*

| Age (years) | Expressive subtest | | | Receptive subtest | | |
|---|---|---|---|---|---|---|
| | 68% CI | 90% CI | 95% CI | 68% CI | 90% CI | 95% CI |
| 1;6-1;11 | 4.68-7.58 | 3.75-8.51 | 3.29-8.97 | 16.32-20.78 | 14.89-22.21 | 14.18-22.92 |
| 2;0-2;5 | 17.80-22.16 | 16.40-23.56 | 15.71-24.25 | 36.97-43.11 | 35.01-45.07 | 34.02-46.06 |
| 2;6-2;11 | 28.86-34.56 | 27.04-37.38 | 26.12-37.30 | 54.61-61.61 | 52.37-63.85 | 51.25-64.97 |
| 3;0-3;5 | 38.03-43.59 | 35.25-46.37 | 35.36-46.26 | 72.63-77.63 | 71.03-79.23 | 70.23-80.03 |
| 3;6-3;11 | 50.33-55.59 | 48.65-57.27 | 47.81-58.11 | 79.20-83.12 | 77.95-84.37 | 77.31-85.00 |

**Table 3.** *Performance of children using CI, as measured using the MERVT and CCDI (N=29)*

| **Measures** | ***M*** | ***SD*** | **Range** |
|---|---|---|---|
| **MERVT** | | | |
| Receptive subtest | 30.76 | 23.74 | 0-70 |
| Expressive subtest | 22.38 | 15.79 | 0-65 |
| **CDI: WG form** | | | |
| Words Understood | 96.45 | 17.78 | 50-106 |
| Words Produced | 84.69 | 27.93 | 12-106 |
| **CDI: WS form** | | | |
| Words Produced | 61.07 | 32.22 | 5-113 |

**Table 4.** *The number of participants in each age equivalent categories (total N = 29; numbers with an underline represent the number of children categorized into the same age-equivalent categories on the CCDI and the MERVT)*

| **CCDI (months)** | **MERVT (months)** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Expressive subtest** | | | | **Receptive subtest** | | | |
| | **<18** | **18-23** | **24-29** | **>30** | **<18** | **18-23** | **24-29** | **>30** |
| **<18 (n=5)** | <u>4</u> | 1 | 0 | 0 | <u>4</u> | 1 | 0 | 0 |
| **18-23 (n=13)** | 0 | <u>7</u> | 4 | 2 | 6 | <u>4</u> | 2 | 1 |
| **24-29 (n=7)** | 1 | 0 | <u>3</u> | 3 | 2 | 1 | <u>3</u> | 1 |
| **>30 (n=4)** | 0 | 1 | 0 | <u>3</u> | 0 | 1 | 2 | <u>1</u> |

**Table 5.** *Spearman rho correlations between chorological age, age at implantation, duration of CI use and scores on the MERVT*

| Variables | Receptive score | Expressive score | Chorological age | Age at CI | Duration of CI use |
|---|---|---|---|---|---|
| Receptive score | 1.00 | .84** | .67** | .31 | .55** |
| Expressive score | - | 1.00 | .48* | .17 | .55** |
| Chorological age | - | - | 1.00 | .77** | .38 |
| Age at CI | - | - | - | 1.00 | -.23 |
| Duration of CI | - | - | - | | 1.00 |

***p* < .01, * *p* < .05, 2-tailed