

---

## Children's Auditory Recognition With Digital Stimuli

**Erin C. Schafer, PhD**

University of North Texas, Denton, TX

**Christine Anderson, BS**

University of North Texas, Denton, TX

**Jessica Sullivan, PhD**

University of Washington, Seattle, WA

**Jace Wolfe, PhD**

Hearts for Hearing, Oklahoma City, OK

**Mila Duke, AuD**

Hearts for Hearing, Oklahoma City, OK

**Homira Osman, BS**

University of Washington, Seattle, WA

**Suzanne Wright, BA**

University of North Texas, Denton, TX

**Jessalyn Dyson, BA**

University of North Texas, Denton, TX

**Danielle Bryant, BA**

University of North Texas, Denton, TX

**Katie Pitts, BS**

University of North Texas, Denton, TX

---

The goal of this investigation was to design a new, age-appropriate, tablet-based word-recognition test, which consists of six lists of 20 digitally-recorded words each with corresponding picture slides. Prior to the study, the suitability of the stimuli was verified (i.e., content validity) by presenting the test vocabulary and photographs to five 3- to 5-year-old children. After the stimuli were deemed appropriate, the test-retest reliability, list equivalency, and convergent validity of the test were determined with 3- to 6-year-old children with normal hearing or hearing loss. Prior to administering the new test, each participant completed a hearing screening and receptive vocabulary test to rule out hearing loss and language delay, respectively. In the children with normal hearing, all lists on the test were completed in two test sessions to assess test-retest reliability and to examine list equivalency. For all participants, average performance on one list of the new test was compared to performance on the revised Word Intelligibility by Picture Identification (WIPI). Results of the study suggested good test-retest reliability and list equivalency of the CARDS for four-, five-, and six-year-old children. List equivalency was also confirmed

for a group of 13 children with hearing loss ranging in age from three to six years. However, the three-year-old children showed an effect of test session, with better performance in Session 2, and significantly poorer performance on List 6 relative to all other lists. Convergent validity was not confirmed for the three- and four-year-olds with normal hearing or for the group of 13 children with hearing loss in this study, with significantly better performance on the CARDS than the WIPI for all groups except the five-to-six-year olds with normal hearing. Further testing with children who have normal hearing or hearing loss will need to be conducted to reexamine convergent validity, collect normative data, examine unaided versus aided performance, and evaluate differences across varying severities of hearing loss and between children using hearing aids and cochlear implants.

## Introduction

The American Academy of Audiology (AAA) Audiologic Guidelines for the Assessment of Hearing in Infants and Young Children (2012) state that the gold standard of pediatric assessment should include an evaluation of ear-specific hearing thresholds and, when age appropriate, speech recognition measures at supra-threshold levels. Prior to assessing speech recognition, the child will need to master lower-level auditory skills including detection, defined as the awareness of a sound in his or her environment, and discrimination, defined as the ability to detect the difference or similarity between two sounds (Erber, 1982). Speech recognition in a quiet sound booth provides a standardized indicator of performance in a well-controlled acoustic environment (i.e., best case scenario). In younger children, ages two to five years, closed-set picture-pointing tasks may be used to determine supra-threshold speech recognition abilities. Closed-set speech recognition tests, such as the Word Intelligibility by Picture Identification (WIPI) and Northwestern University Children's Recognition of Speech (NU-CHIPS) tests, are widely known and used in audiological assessments of young children (Elliot & Katz, 1980; Cienkowski, Ross, & Lerman, 2009; Ross & Lerman, 1970). These closed-set tests consist of four lists of 25 to 50 words each which are depicted on picture plates containing four to six picture choices. Children are asked to point to the picture corresponding to the spoken word presented via monitored live voice (MLV) or recorded presentation. Both of these tests are valuable for speech recognition testing, especially in young children with expressive language delays or articulation disorders. However, the pictures used in these tests consist of paper-based line drawings and simple representations of the test stimuli rather than more realistic depictions of the stimuli. In fact, the original WIPI was found to contain several confounding picture choices and "pictures unfamiliar to children" (Cienkowski et al., 2009; Stewart, 2003, Dengerink & Bean, 1988).

### Importance of Speech Recognition Testing

In addition to routine audiological assessments, the AAA guidelines for assessing hearing in children (2012) also state that assessment of speech recognition at supra-threshold levels is critical for formulating recommendations regarding amplification, aural habilitation, and educational strategies. Speech-recognition assessment is particularly important because, generally, the primary goal of amplification is to restore audibility of the speech signal to facilitate development of speech, language, and communication (Bagatto, Scollie, Hyde, & Seewald, 2010; Seewald, Moodie, Scollie, & Bagatto, 2005). Therefore, quantifying audibility is essential to ensure that children have sufficient access to the acoustic cues that facilitate speech and language development (McCreery & Stelmachowicz, 2011). Following a fitting of a hearing aid using objective verification measures including real ear to coupler difference and real-ear aided response (REAR) probe microphone measurements, speech recognition measures may be used to validate appropriate outcomes in children (AAA Clinical Practice Guidelines: Pediatric Amplification, 2013). When a child with hearing loss is evaluated for the potential

benefit of amplification, speech recognition may be conducted at soft (e.g., 40 dB HL), conversational (e.g., 55 dB HL), and loud intensity levels (e.g., 80 dB HL) to assess audibility and comfort across a range of loudness levels in an ideal acoustic environment (i.e., sound booth). For children who have developed some speech and language, speech-recognition scores may also be used to determine cochlear implant candidacy. Following the receipt of a hearing aid or cochlear implant, the same three input intensities may be utilized to behaviorally verify the adequacy of programming. This same speech recognition testing may be conducted at follow-up appointments to monitor progress and to plan habilitative and educational goals and objectives. Finally, speech-recognition measures and outcomes may be used as a counseling tool for parents and/or caregivers. When adequate audibility is not achieved for soft speech in an unaided condition, low percent-correct scores provide concrete evidence to parents that intervention is necessary. Similarly, when adequate audibility is not achieved for conversational speech in an aided condition, poor speech-recognition scores provide evidence that further technology, such as a cochlear implant and frequency modulation (FM)/digital remote-microphone technology, may be necessary.

### Study Rationale

Given the importance of supra-threshold speech-recognition testing and limitations of existing closed-set word-recognition tests for young children, a new word-recognition test consisting of simple stimuli and digital photographs was developed for computerized administration on an electronic tablet. Although one existing test, the WIPI, was updated in 2009 to include more relevant vocabulary and new artist drawings (Cienkowski et al., 2009), the presentation of this test via paper format and use of drawings still present limitations when compared to a computerized format and digital pictures. The investigators hypothesized that the computerized format and digital photographs of people, places, objects, and actions may be more relevant, recognizable, and universal to children when compared to existing test stimuli. Many young and school-aged children have access to or own digital hand-held devices and use these devices on a daily basis, making a tablet-based test format more familiar than a paper booklet to most children (Henry J. Kaiser Family Foundation, 2010). Also, the tablet, unlike a paper booklet, is portable and allows for storage of more than one test, resource, or game for children within one device. Tablet-based presentation allows for easy navigation and does not require flipping through pages of pictures during testing. The primary goal of this study was to construct an age-appropriate, supra-threshold tablet-based word-recognition test for children ages three to six years. Secondary goals were to examine aided performance of children with varying degrees of hearing loss in sound-field to examine list equivalency and to determine whether the new test may be feasible for use as a validation measure following a hearing aid or cochlear implant fitting.

## Methods

### Participants

The methods and procedures for this study were approved by the Institutional Review Boards at the University of North Texas (UNT) and the University of Washington (UW). Participants included a total of 57 children ranging in age from three to six years, and parental consent was obtained from all parents of the children prior to their participation in the study. Forty-three of these children had normal-hearing sensitivity bilaterally, defined as threshold responses of less than or equal to 20 dB HL on a pure-tone hearing test ranging at octave frequencies from 250 to 8000 Hz. Within the group of children with normal hearing, 14 were 3-years old, 14 were 4-years old, and 15 were 5- and 6-years old.

All children with normal hearing were tested at UNT and UW. The remaining 14 children had bilateral sensorineural hearing loss with severities ranging from mild to severe and were tested at UNT, UW, or Hearts for Hearing in Oklahoma City, Oklahoma. All children with hearing impairment were fit using a standard hearing aid fitting protocol utilizing Desired Sensation Level (DSL) v5 prescriptive targets (Scollie et al., 2005; Seewald et al., 2005) and real-ear verification measures. Additional information about the children with hearing loss is provided in Table 1. Test stimuli were calibrated with a Type 1 sound-level meter (Larson-Davis, 824). An electronic version of the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2012) was administered with a laptop computer.

**Table 1.** Demographic Information for Participants with Hearing Loss

Subject	Age	Aided PTA	Unaided	Description	Device
1	6;1	B: 17 dB	R:48 dB L:18 dB	R: Atresia L: sloping HL	R: Phonak Nios S H20 III L: Cochlear BAHA BP100 (Soft band)
2	5;3	B: 20 dB	R: 43 dB L: 38 dB	R: Moderate rising to normal L: Moderate rising to normal	Bilateral Phonak Nios S H20 III
3	6;2	R: 28 dB L: n/a	R: 95+ dB L: 12 dB	R: profound L: normal hearing	R: BAHA 5 Attract L: none
4	3;11	B: 32 dB	R: 78 dB L: 82 dB	Bilateral severe sloping to profound	R: Phonak Naida Q50 SP L: Cochlear Nucleus 6
5	3;4	R: n/a L: 33 dB	R: 67 dB L: 75 dB	R: flat moderately-severe L: flat severe	Bilateral Phonak Nios S H20 III
6	5;5	n/a	R: 52 dB L: 47 dB	Bilateral moderate flat	Bilateral Phonak Nios S H20 III
7	5;5	R: 27 dB L: 32 dB	R: 60 dB L: 57 dB	Bilateral moderate sloping to severe	Bilateral Phonak Sky Q50 M13
8	6;3	B: 15 dB	R: 8 dB L: 12 dB	Bilateral precipitous, normal to severe at 4kHz	Bilateral Phonak Sky Q50 M13
9	6;6	R: 28 dB L: 23 dB	R: 70 dB L: 95+ dB	R: moderate sloping to severe L: profound	R: Phonak Naida Q90 UP L: Cochlear Nucleus 6
10	4;1	R: 18 dB L: 23 dB	R: 95+ dB L: 95+ dB	Bilateral profound	Bilateral Cochlear Nucleus 6
11	4;7	R: 27 dB L: 23 dB	R: 95+ dB L: 95+ dB	Bilateral profound	Bilateral Cochlear Nucleus 6
12	3;8	B: 27 dB	R: 52 dB L: 53 dB	R: Mild sloping to moderately-severe at 4000 Hz L: Mild sloping to moderate at 2000 Hz	Bilateral Phonak Sky Q50 M13
13	4;0	R: 52 dB L: 53 dB	B: 27 dB	Bilateral mild sloping to moderate	Bilateral Phonak Nios S H20 III
14	6;8	B: 17 dB	R: 48 dB L: 50 dB	R: Moderate rising to normal at 6000 Hz L: Moderate rising to normal at 8000 Hz	Bilateral Phonak Sky Q50 M13

Note. R= right; L=left; B=binaural; n/a=not available.

### Test Stimuli

The new word-recognition test, which will be referred to as the Children’s Auditory Recognition with Digital Stimuli (CARDS) test, consisted of six lists of 20 words each. The 120 CARDS stimuli consisted of digital photographs arranged on picture plates and digitally-recorded monosyllabic words. The process of validating the stimuli is provided in the results section.

### Digital Recordings

The pictured target words were recorded by a female talker, and the acoustic editing software, Cool Edit Pro Version 2 (2003) was used to equalize the root-mean-square intensity of each word. The phonemic distribution of the words across the six lists is provided in Table 2. The procedures used to determine phonemic balance of the text lists were modeled from those used in the development of the Hearing in Noise Test (HINT; Nilsson, Soli, & Sullivan, 1994). First, the target phoneme count was calculated, which is defined as the difference between the target phoneme count across the six lists

divided by the total number of lists (6). Second, a difference score was calculated between the target phoneme count and the obtained (actual) phoneme count. The distribution of these differences is displayed in Figure 1 where a deviation of zero represents perfect phonetic balance. Finally, the percentage of the difference scores that were within + 1 phoneme were tabulated. A difference of + 1 phoneme was present in 63% of the difference scores. As a result, phonetic balance was achieved, for the most part, and was similar to the phonetic balance reported for the HINT, which had + 1 phoneme for 68% of difference scores. The equivalency of the test lists (i.e., equal difficulty) was determined in the study design. The stimuli were recorded to compact disc and were presented to the participant at 40 dB SL relative to his or her pure-tone average at 500, 1000, and 2000 Hz (aided PTA for children with hearing loss). In addition, one randomly-selected list was presented at 50 dB HL to examine performance for a fixed intensity representing conversational speech.

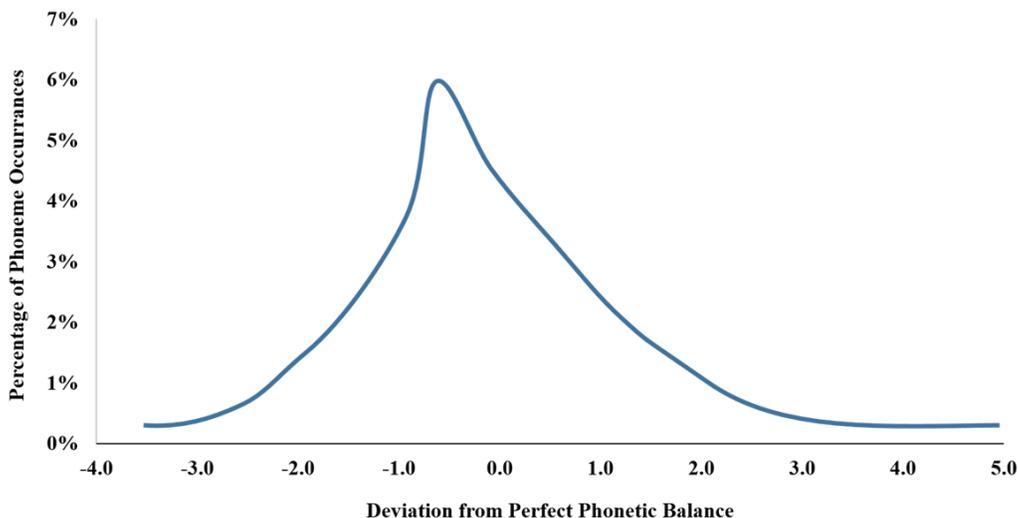


Figure 1. Deviation of difference scores for all 120 words in the six word lists.

Table 2. Phoneme Distribution for 120 Words in the Six Lists.

Consonant Distribution							
/p/	4.1%	/k/	7.6%	/s/	7.3%	/tʃ/	0.5%
/b/	4.6%	/g/	2.5%	/z/	1.0%	/dʒ/	0.8%
/t/	5.8%	/f/	2.3%	/ʃ/	1.5%	/m/	3.3%
/d/	5.1%	/θ/	1.3%	/h/	1.8%	/n/	5.1%
/ŋ/	0.5%	/l/	4.8%	/w/	2.5%	/r/	6.8%
Vowel Distribution							
/i/	2.8%	/æ/	3.8%	/ʊ/	0.5%	/əʊ/	0.3%
/ɪ/	1.8%	/ɑ/	5.1%	/u/	2.0%	/aɪ/	2.5%
/e/	1.0%	/ɔ/	1.0%	/ʌ/	3.3%	/aʊ/	1.0%
/ɛ/	2.3%	/o/	2.5%	/ɜr/	0.8%	/ɔi/	0.3%

Note. Percentages represent how often the phoneme occurred relative to the total number phonemes.

### **Other Test Measures**

In addition to the CARDS lists, children completed one list of the Word Intelligibility by Picture Identification test (WIPI; Cienkowski et al., 2009) at 40 dB SL relative to his or her pure-tone average at 500, 1000, and 2000 Hz (aided PTA for the children with hearing loss) and the PPVT-4 (Dunn & Dunn, 2012) outside of the double-walled sound booth with the examiner sitting beside the child. The PPVT-4 was used to confirm every child in the study had appropriate receptive-vocabulary levels based his or her chronological age because poor receptive vocabulary could have impacted speech recognition performance on the CARDS. According to the testing, all children had age-appropriate receptive vocabulary levels.

### **Study Design and Procedures**

The 43 children with normal hearing sensitivity were tested in two test sessions with a one- to three-week gap between test sessions. The 14 children with hearing loss were only tested in one test session. Two types of reliability were assessed in this study by calculating: (1) test-retest reliability, or the consistency of the scores from one session to another, and (2) internal consistency reliability or test list equivalency (Trochim, 2005). In addition, content validity was determined prior to the study when determining appropriate vocabulary and recognizable pictures in five, three-to six-year-old children (i.e., common vocabulary and pictures screened with pilot data). Finally, convergent validity was assessed with a comparison between scores on the CARDS test list and the WIPI test list in 41 of the 43 children with normal hearing and 13 of the 14 children with hearing loss.

#### **Session 1**

After study personnel explained study procedures and obtained parental consent in Session 1, parents were asked to complete a case history form for their child. The case history was used to rule out recurrent otitis media or surgeries in the children with normal hearing and to obtain more detailed hearing history and device information from the children with hearing loss.

Following completion of the paperwork, the examiner conducted the pure-tone hearing test or previous tests were obtained from the parent for some of the children with hearing loss who received an evaluation within the past six months. After the hearing test, the PPVT-4 was administered via laptop computer. Next, each list of the CARDS, in pseudo-randomized order (i.e., no repeated lists), was presented at 40 dB SL using the iPad and compact disc player, along with an additional list presented at 50 dB HL.

For children with normal hearing, only one ear was tested with an insert earphone to avoid ear effects in this initial assessment of the stimuli. This procedure was adapted from the WIPI procedures (Cienkowski et al., 2009). The test ear was counterbalanced across participants and the stimulus intensity for the test ear was determined by calculating the child's pure tone average (PTA) at 500, 1000, and 2000 Hz. Children with hearing loss were tested in their normal aided condition (Table 1) in the sound field with the

loudspeaker at 0 degrees azimuth in order to examine the utility of CARDS for supra-threshold speech recognition assessment following a hearing aid fitting or cochlear implant activation. The stimulus intensity for the sound-field testing was determined by calculating the child's better-ear aided PTA at 500, 1000, and 2000 Hz, and testing was presented at 40 dB SL. If the child was not aided, the better unaided threshold was used (Participant #6). For both groups of children, the examiner recorded correct responses during testing and a percent-correct score was calculated for each list of 20 words. Following this testing, children completed one list of the WIPI.

Forty-one children with normal hearing and 13 children with hearing loss completed List 1 of the WIPI at 40 dB SL to examine convergent validity. Additionally, 35 children with normal hearing and 11 children with hearing loss also completed List 1 of the CARDS at 50 dB HL in order to provide normative data on expected performance at a level corresponding to normal conversational speech in an ideal acoustic environment (i.e., sound booth). The investigators expected similar performance between the 40 dB SL and 50 dB HL conditions given that both should provide adequate audibility.

#### **Session 2**

In Session 2, the children with normal hearing sensitivity completed a re-screen of hearing from 250 to 8000 Hz to verify no change in thresholds. Following the screening, all six lists of the CARDS were repeated in a pseudo-randomized order.

### **Results**

#### **Effects of Age, List, & Test Session**

Average word-recognition scores from the three groups of children with normal hearing across the six lists are shown in Figure 2, and average performance across the lists and test sessions are shown in Figure 3. Because of the performance of the children at or close to ceiling (i.e., 100% correct), percent-correct scores were converted to rationalized arcsine units (RAU) prior to statistical analysis of all data (Studebaker, McDaniel, & Sherbecoe, 1995).

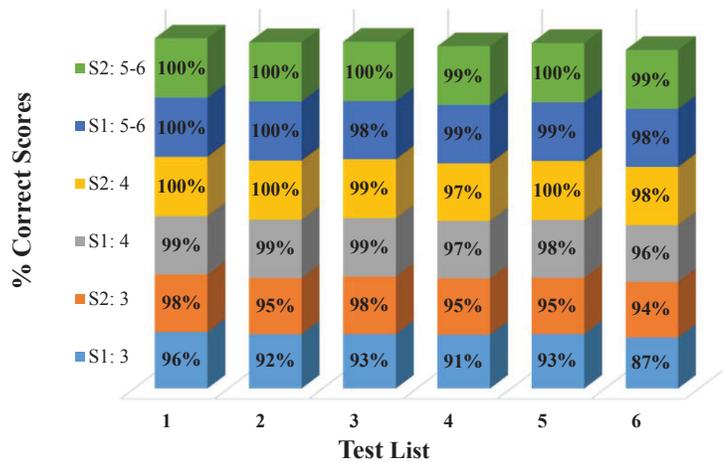
Two types of reliability, test-retest reliability and internal consistency reliability (list equivalency) were examined with parametric statistics. More specifically, a three-factor repeated measures analysis of variance (RM ANOVA) was conducted with the independent variables of age (3, 4, and 5-6), test list (1-6) and test session (1, 2). According to this analysis, there was a significant main effect of age,  $F(2,516) = 17.9, p < 0.001$ , a significant main effect of test list,  $F(5,516) = 9.9, p < 0.001$ , and a significant main effect of test session,  $F(1,516) = 19.5, p < 0.001$ . There were no significant interaction effects between age and session,  $F(2,516) = 1.5, p < 0.23$ , age and list,  $F(10,516) = 1.3, p < 0.21$ , or list and session,  $F(5,516) = 1.3, p < 0.29$ . Post-hoc analyses were conducted with the Tukey-Kramer Multiple Comparisons test to more closely examine the significant main effects. For the main effect of age, the two older age groups had significantly better average scores than the three-year olds (both  $p < .05$ ). For the main effect of test list, List 6 yielded significantly

lower average scores than Lists 1, 2, 3, and 5 (all  $p < .05$ ), and List 4 resulted in significantly lower average scores than Lists 1 and 2 (both  $p < .05$ ). No other significant differences were found. When examining the post-hoc analysis on test session, average scores in Session 2 were significantly better ( $p < .05$ ) than those in Session 1. Given the significantly poorer performance of the three-year-old children, additional analyses were conducted to more closely examine the effect of age on the results. A post-hoc analysis for age by session suggested that only the average scores of the three-year-olds differed significantly between the two test sessions ( $p < .05$ ). Similarly, an analysis of age by test list suggested that only the three-year-old children showed significant performance differences across the test lists with List 6 yielding worse scores than Lists 1, 3, and 4 (all  $p < .05$ ).

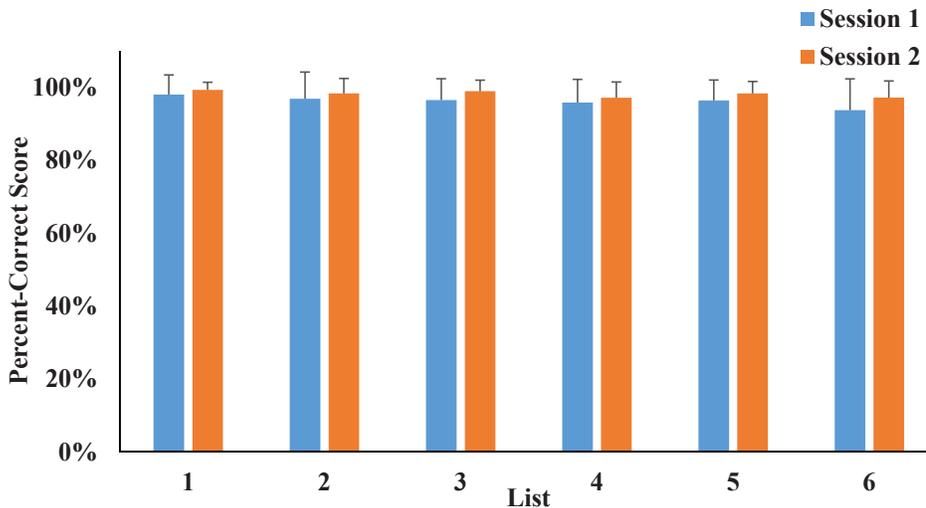
To further examine test re-test reliability for the four-, five-, and six-year olds, the average score across the six lists was determined for each participant with normal hearing in Session 1

and Session 2. A Pearson's Product Moment Correlation was then calculated using the average scores for each participant in the two separate sessions. The correlation coefficient between the scores obtained in Session 1 and Session 2 was 0.71, which confirms moderately high test-retest for this word-recognition test.

List equivalency was also examined for the 14 children with hearing loss who ranged in age from three to six years. Individual scores of the 14 children with hearing loss ranged from 80-100% across lists ( $M=97\%$ ;  $SD=5$ ). As a result, these data were also transformed to RAU to allow for statistical analysis. Given the effect of age for the children with normal hearing, a one-factor analysis of covariance (ANCOVA) was performed to control for the effect of age. According to this analysis, there was no significant main effect of test list,  $F(5,84) = 17.3$ ,  $p = 0.20$ . To account for the repeated measures aspect of the design, a RM ANOVA was also conducted and yielded the same results (i.e., no significant main effect of list;  $F[5,84] = 1.62$ ,  $p = 0.17$ ).



**Figure 2.** Average percent correct performance on the CARDS for each age group with normal hearing by session and word list.



**Figure 3.** Average speech-recognition scores from children with normal hearing across session and CARDS word list.

### *Validity of the Word-Recognition Test*

Content and convergent validity were examined in this study. Content validity was confirmed prior to the study through a series of steps to ensure appropriate test material for three-to-six-year-old children. First, four examiners determined and documented approximately 400 frequently occurring vocabulary words in children's environments through daily interactions with pre-school aged children over a period of four weeks.

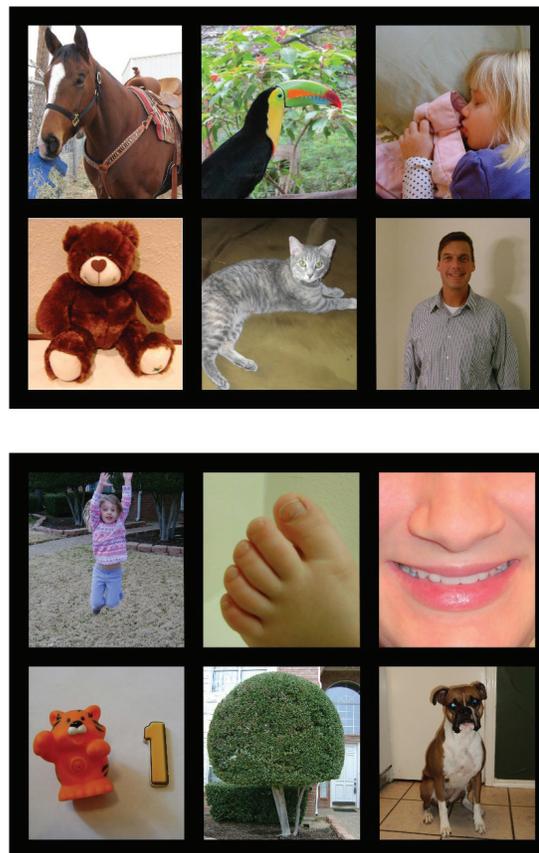
Second, these 400 words were discussed by four examiners, who had experience working with children (i.e., pediatric audiologist; speech assistant; 2 graduate assistants with pediatric experience), in order to select the most appropriate stimuli for the test. Stimuli had to meet three criteria for further consideration: (1) only nouns and verbs were considered, (2) only monosyllabic words were considered, and (3) only words that could be depicted easily in photographs were included. Using these criteria, 150 words remained in the stimulus set.

Third, the examiners digitally photographed the 150 words. For all words, multiple photographs were taken to allow for

selection of the clearest and most recognizable photograph with the best angle. The photographs were taken in everyday environments at home, at the park, and at school.

Fourth, the same four examiners reviewed all photographs taken for the 150 words and agreed collectively on which photograph best depicted the word, keeping in mind the age group for which the test was designed. At this stage, the examiners reduced the number of stimuli to include only the most clearly depicted 120 words.

Fifth, using Microsoft PowerPoint, 120 digital picture plates were created. On each plate, a photograph for the target word was shown along with five photographs for non-target words, which were randomly selected from the pool of remaining photographs. The 120 picture plates were divided equally into six separate digital folders containing 20 picture plates each (i.e., 6 lists of 20 words each), which were then uploaded to an Apple iPad. Two sample picture plates are shown in Figure 4 and the six word lists are provided in Table 3.



**Figure 4.** Two picture plates from the new word-recognition test.

**Table 3.** Word Lists 1 Through 6.

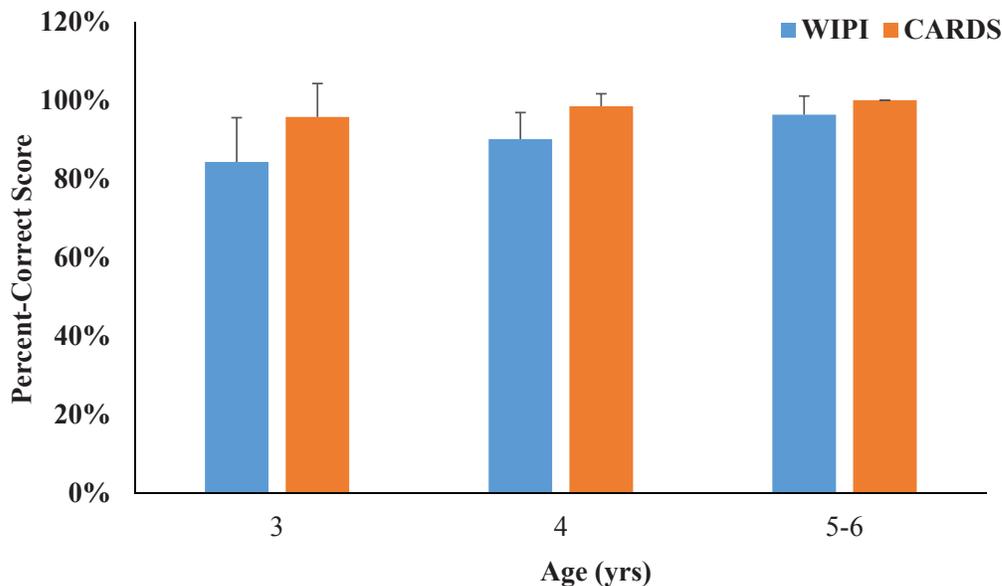
List 1	List 2	List 3	List 4	List 5	List 6
Bird	Three	Rocks	Mouth	Dress	Orange
Boat	Fan	Fish	Bed	Stairs	Sleep
Book	Dad	Cat	Pool	Clock	Sheep
Car	Girl	Blocks	Duck	Pie	Tie
Food	Boy	Black	Mouse	Cold	Walk
Hat	Hair	Two	Bat	Heart	Pants
Juice	Corn	Nap	Plate	Knife	Sad
Nose	Key	Brush	Bowl	Truck	Red
Doll	Paint	Teeth	Run	Snow	Pig
Mom	Eggs	Foot	Bear	Desk	Smile
Sand	Sky	Milk	Watch	Kick	Ball
Sun	Door	Star	Stop	Swing	Soap
Bug	Horse	Eat	Box	Wheel	Fork
Dog	Plane	Toes	Can	Goats	One
Phone	Green	Cup	Arm	Draw	Eye
Fire	Bath	Man	Hug	Drink	Cow
Sock	Hand	Light	Blue	Grass	Cake
Swim	Mop	Boot	Tree	Slide	Spoon
Drum	Frog	Sick	Shirt	Bike	Shoe
House	Thumb	Chair	Jump	Wash	Ear

Sixth, pilot data were then collected from five, three- to five-year old children (two, 3-yr olds; one, 4-yr old; two, 5-yr olds) with normal hearing sensitivity (< 20 dB HL from 250-8000 Hz), as determined by a hearing screening and no history of otitis media, ear surgeries, or speech-language disorders, as reported by parents on a case history form. During testing, an examiner was seated next to a child in a quiet room; the examiner presented each word via live voice with no visual cues. Following the auditory stimulus, the child was asked point to the photograph on an Apple iPad that best depicted the word that was heard. This process was repeated for all 120 picture plates. Given that all five children identified the 120 stimuli and picture plates with 100% accuracy, the stimuli were deemed appropriate and valid for use in the present study.

To evaluate convergent validity, or similarity of the CARDS to an existing word-recognition test, 41 of the children with normal hearing completed List 1 of the WIPI. These data are shown in Figure 5. The children's WIPI scores were compared to scores obtained with a list on the CARDS by calculating a correlation coefficient. The correlation coefficient was .33, which suggests a weak to moderate relationship between the two tests. Additionally, a two-factor RM ANOVA was conducted with the independent variables of age and test. The analysis yielded a significant main effect of age,  $F(2,82) = 10.6$ ,  $p = 0.0002$ , and a significant main effect of test,  $F(1,82) = 48.6$ ,  $p < 0.0001$ . Post-hoc analysis with the suggested that average performance between the CARDS and WIPI was similar for the five-to-six-year-olds, but significantly different for the three- and four-year olds. Although this analysis

did not confirm a strong correlation (i.e., high convergent validity) and the same performance between the two tests, it did confirm that the CARDS will produce scores that are equal to or higher than scores obtained on the WIPI. Similar to the children with normal hearing, correlation coefficient was calculated and a two-factor RM ANOVA was conducted for 13 children with hearing loss. A weak to negligible correlation coefficient of -.09 was calculated, which may be related to the small sample size. Also, the RM ANOVA for the children with hearing loss showed no significant main effect of age,  $F(2,26) = 4.0$ ,  $p = 0.052$ , and a significant main effect of test,  $F(1,26) = 35.9$ ,  $p = 0.0001$ . A post-hoc analysis suggested better performance on the CARDS ( $p < .05$ ) than the WIPI. Potential reasons for the performance discrepancy will be outlined in the discussion section.

When examining the results in the 50 dB HL CARDS test condition, average performance was 98% (SD=2.8; Range=90-100%) for the 35 children with normal hearing and 98% (SD=2.5; Range=95-100%) for the 11 children with hearing loss. The implications of these results will be explored in the discussion section.



**Figure 5.** Average percent correct performance for one list of the WIPI and List 1 of the CARDS for 41 children with normal hearing. Note. Vertical lines represent one standard deviation

### Discussion

On average, test-retest reliability and list equivalency of the CARDS were confirmed for the four-, five-, and six-year-old children. However, the three-year-old children showed an effect of test session, with better performance in Session 2, and significant worse performance on List 6 relative to all other lists. Because the average performance of the three-year-olds in Session 2 was slightly higher than what was obtained in Session 1, there was a likely a learning effect present, which may be related to the brief, one to three-week period between test sessions. As a result, if this test were used with three-year-olds in clinical practice, the authors of this study recommend completing a practice list prior to the test list at each appointment. Additionally, given the poorer performance of the three-year-olds on List 6, that list should not be used with this age group. List equivalency was confirmed for a group of 13 children with hearing loss ranging in age from three to six years.

Necessary steps were taken prior to the study to document content, and a comparison of performance on the CARDS and the WIPI was conducted to examine convergent validity. However, convergent validity was not confirmed for the three- and four-year-olds with normal hearing or for the group of 13 children with hearing loss in this study because performance on the CARDS was significantly better than performance on the WIPI. However, average performance of the five-to-six-year olds with normal hearing was similar on the CARDS and WIPI. Again, it is possible that learning effects were involved in the differences for

the younger age groups. The order of the test lists on the CARDS was pseudo-randomized and, as a result, many children completed one or more test lists of the CARDS before List 1, which was used for the comparison to the WIPI. Conversely, there was no practice list(s) for the WIPI.

It is important to note that a 40 dB SL presentation level is not sensitive to performance differences in children with normal hearing or the sample of children with hearing in the present study who had, for the most part, good aided thresholds (Table 1). The present investigation is only the first step in the development of this test and was necessary to examine reliability and validity when stimuli were equally audible to all participants. The next step in the development of the CARDS would be to collect normative data on Lists 1-4 from a large sample of children with normal hearing and also with hearing impairment. Future research should examine the concurrent validity of the CARDS (i.e., the measure should distinguish between groups that should be different) by comparing performance of a group of children with normal hearing to a group of children with more severe degrees of hearing loss than those in the present study (Trochim, 2005). Additional research may also assess two-year old children in the closed-set format or evaluate performance in an open-set condition in young and older children with and without hearing loss. Although concurrent validity could have been shown in the 50 dB HL condition, there was highly similar performance between the normal hearing and hearing loss groups in the present study. First, for many of the children with normal hearing, 40 dB SL was 50 dB HL; therefore, limited

additional information was gained from the extra condition at 50 dB HL for this group. Second, while the children with hearing loss all had aided thresholds (PTA) higher than 10 dB HL, they still achieved excellent performance in the 50 dB HL condition likely due to the fairly good aided thresholds (i.e., normal to mild hearing loss range). Different results might have been obtained if the children with hearing loss were tested in a unilateral condition or in the unaided condition. At the same time, the use of an insert earphone in one ear of normal hearing participants and use of sound field speakers for the aided/implanted children with hearing loss confirms that both presentation modes are feasible. Additional demographic information about the children with hearing loss, such as length of amplification/cochlear implant use and the quality of the fitting, would have been helpful for examining results from the children in this study. However, given their excellent performance at 40 dB SL relative to their aided PTA, these children appeared to have adequate recognition abilities with their hearing aids or cochlear implants. Future research should examine performance of children with hearing loss at soft (e.g., 40 dB HL), conversational (e.g., 55 dB HL), and loud intensity levels (e.g., 80 dB HL) because audibility and comfort across a range of loudness levels is important for optimal hearing aid fittings. Additionally, future investigations may determine the reliability and validity of the CARDS when presented in background noise.

When conducting this test with children who have hearing loss, it is important to consider whether the 40 dB SL presentation level is appropriate (Hornsby & Mueller, 2013). In cases where the child has a precipitously sloping hearing loss, it may be more appropriate to present the stimuli at 40 dB SL relative to the child's high-frequency PTA (average of 1000, 2000, and 4000 Hz). However, if 40 dB SL relative to high-frequency PTA exceeds the child's uncomfortable loudness level, the audiologist may consider using a fixed 80 dB HL presentation level. Of course, use of an 80 dB HL presentation level represents louder speech and will not simulate soft or conversational speech levels. It is also important to note that a 40 dB SL presentation level may be uncomfortably loud for some children with hearing loss, particularly relative to an unaided PTA. If this occurs, the examiner will need to present the CARDS stimuli at the child's most comfortable listening level.

One unexpected finding in this investigation was the significantly higher average scores on the tested list of the CARDS relative to the average scores on List 1 of the WIPI. As stated above, this could be due to a learning effect, or it is possible that the higher CARDS test scores may be related to the use of digital photographs instead of line drawings. As mentioned in the introduction section, closed-set speech recognition tests, such as the WIPI and the NU-CHIPS, are an important part of the audiological test battery. However, if a child does not recognize the picture as matching the verbal stimulus, the validity of the response may be questionable (Dengerink & Bean, 1988). For example, in a previous study of the NU-CHIPS, the picture tongue elicited labels of hat, God, and other body parts (Dengerink & Bean, 1988). This specific item was missed because the picture did not represent a recognizable item to the children. By using

digital photographs on the CARDS, a child may be more likely to recognize the picture. As a result, the CARDS was developed to eliminate unfamiliarity with line drawings, as well as provide auditory and digital stimuli more relevant to modern children's lexicon. Research by Dengerink and Bean (1988) also show that test subjects found the common pictures of the WIPI to be more identifiable than the common pictures on the NU-CHIPS. More specifically, colored pictures were more readily identifiable than the black and white line drawings. It may be inferred from the results in the present study that photographs, which are more realistic than colored sketches or black and white line drawings, would also be more recognizable for children. At the same time, these differences could also be due to different vocabulary used for the CARDS and WIPI. The exact reason for the discrepancy between tests cannot be confirmed at this time.

Additionally, the use of a tablet-designed test allows the examiner to use the CARDS test program on multiple platforms, including but not limited to computers, Apple iPads, and various other tablets. The examiner also has the ability to upload various other programs onto a specific device, such as games to reengage the child during testing, counseling tools (e.g., digital pictures of ear anatomy), or auditory training applications. Furthermore, as technology advances, it is possible that various other audiological test materials will be provided in this format, condensing the testing material needed into one portable device. From the anecdotal experience of the examiners, the tablet-based platform for testing is also more engaging for children who are familiar with tablets and often possess positive associations with tablet-like applications.

In summary, the results of this investigation suggest that four lists of the CARDS may be used as a reliable measure (good test-retest and list equivalency) of closed-set word recognition in children 4 to 6 years of age with normal hearing. A group of children with hearing loss also showed list equivalency. The authors of this study believe that learning effects impacted performance for three-year-olds with normal hearing on the CARDS, but Lists 1-5 may be used clinically after a practice list is utilized at each appointment. Content validity was documented; however convergent validity (i.e., CARDS vs. WIPI) could not be confirmed for most children, with the exception of the five-year-olds with normal hearing. Additional research is necessary with a larger sample size and a range of hearing losses. At this time, the test may be used to examine supra-threshold or multi-level word-recognition performance during a general audiological assessment, to examine audibility of speech following a hearing aid fitting or cochlear implant programming, and to plan habilitative intervention.

## References

- American Academy of Audiology Clinical Practice Guidelines. (2012). *Audiologic guidelines for the assessment of hearing in young infants and children* [Guidelines].
- American Academy of Audiology Clinical Practice Guidelines. (2013). *Pediatric amplification* [Guidelines].
- Bagatto, M., Scollie, S., Hyde, M., & Seewald, R. (2010). Protocol for the provision of amplification within the Ontario infant hearing program. *International Journal of Audiology, 49*(Suppl. 1), S70-S79.
- Cool Edit Pro. (Version 2)(2003). Phoenix, AZ: Syntrillium Software Corporation.
- Cienkowski, K., Ross, M., & Lerman, J. (2009). The Word Intelligibility by Picture Identification (WIPI) test revisited. *Journal of Educational Audiology, 15*, 39-43.
- Dengerink, J. & Bean, R. (1988). Spontaneous labeling of pictures on the WIPI and NU-CHIPS by 5-year-olds. *Language, Speech, and Hearing Services in Schools, 19*, 144-152.
- Dunn, L. M. & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. Minneapolis, MN: Pearson Assessments.
- Elliott, L.L. & Katz, D. (1980). Development of a new children's test of speech discrimination (Technical Manual). St. Louis, MO: Auditec.
- Erber, N. (1982) Auditory Training. Washington DC: Alexander Graham Bell Association.
- Henry J. Kaiser Family Foundation Generation M2: Media in the Lives of 8- to 18-Year Olds. (2010). Menlo Park, CA.
- Hornsby, B., & Mueller, G. (2013). Monosyllabic word testing: Five simple steps to improve accuracy and efficiency. Retrieved from Audiology Online: <http://www.audiologyonline.com/articles/word-recognition-testing-puzzling-disconnect-11978>.
- McCreery, R. & Stelmachowicz, P. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *Journal of the Acoustical Society of America, 130*, 4070-4081.
- Nilsson, M., Soli, S., & Sullivan, J. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95*(2), 1085-1099.
- Ross, M., Lerman, J. (1970). A picture identification test for hearing-impaired children. *Journal of Speech, Language, and Hearing Research, 13*, 44-53.
- Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., ... & Pumford, J. (2005). The Desired Sensation Level multistage input/output algorithm. *Trends in Amplification, 9*(4), 159-197.
- Seewald, R., Moodie, S., Scollie, S., & Bagatto, M. (2005). The Desired Sensation Level method for pediatric hearing instrument fitting: Historical perspective and current issues. *Trends in Amplification, 9*, 145-157.
- Stewart, B. (2003). The Word Intelligibility by Picture Identification Test: A two-part study of familiarity and use. *Journal of Educational Audiology, 11*, 39-48.
- Studebaker, G. A., McDaniel, D. M., & Sherbecoe, R. L. (1995). Evaluating relative speech recognition performance using the proficiency factor and rationalized arcsine differences. *Journal of the American Academy of Audiology, 6*(2), 173-182.
- Trochim, W.M. (2005). *Research Methods: The Concise Knowledge Base*. Cincinnati, OH: Atomic Dog Publishers.